STUDYING SEX AND GENDER

Key Concepts

Learning Objectives

Students who read this chapter should be able to do the following:

2.1 Evaluate the meaning of sex differences.

2.2 Explain the scientific method and specific quantitative and qualitative methods used in the study of sex and gender.

2.3 Describe meta-analyses and explain how to interpret effect sizes of different magnitudes.

2.4 Analyze methodological challenges and biases in sex and gender research.

2.5 Explain the principles of gender-fair and inclusive research, and describe issues of diversity in sex and gender research.

STUDYING SEX AND GENDER

To understand the research in this book, it is essential to understand the methods that sex and gender researchers use. You may be thinking that this is one chapter that you would like to skip. We get it. Research methods can seem dry, and statistics can be intimidating. But learning research methods need not be painful. Think of the study of gender as a mystery and the researcher as the detective trying to crack the case. Think of research methods and statistics as powerful tools that allow you to understand the complexities of human behavior. Armed with a good understanding of research methods, we hope that you will also appreciate the importance (and dare we say, fun) of learning research methods.

Rather than presenting a thorough, technical review of research methods, this chapter instead focuses on common methodological approaches and challenges in gender research. Why is it important to study questions of gender systematically? Without systematic research, people would likely rely too heavily on stereotypes and intuitions to understand questions of gender, making them prone to misconceptions. They might overlook many of the complex and counterintuitive findings that emerge through a careful study of gender phenomena. For example, consider the following beliefs that do not stand up to empirical scrutiny:

- **Many believe that men inherently possess greater math ability than women**. However, large-scale reviews show no overall sex differences in math performance (Lindberg, Hyde, Petersen, & Linn, 2010). In fact, math performance is predicted less by sex than it is by other factors, such as socioeconomic status, primary school effectiveness, home learning environment, and mother's education level (Melhuish et al., 2008). Nonetheless, girls do tend to have higher levels of math anxiety than boys (Stoet, Bailey, Moore, & Geary, 2016), and this anxiety can sometimes interfere with their performance on math tests. (For more on this topic, see Chapter 7, "Cognitive Abilities and Aptitudes.")

- **Many people regard women as more talkative than men**. But when Matthias Mehl and colleagues recorded women and men in their daily lives, they found no sex differences in numbers of words spoken per day (Mehl, Vazire, Ramírez-Esparza, Slatcher, & Pennebaker, 2007). Other studies do reveal sex differences, but they are small and depend on age. For example, young girls (under the age of 3) tend to be slightly more talkative than young boys, whereas men tend to be slightly more talkative than women (Leaper, 2014).

- **Common Western views dating back to the Victorian era hold that women—in comparison with men—are less interested in sex**. For example, when approached by an attractive stranger who offered casual

sex, women declined the offer much more frequently than men did (R. D. Clark & Hatfield, 1989). More recently, however, Terri Conley (2011) found that there are several reasons for American college women's reluctance to accept casual sex offers: (1) Unfamiliar men may pose a danger, (2) women expect to be stigmatized for having casual sex, and (3) women do not expect sex with a stranger to be pleasurable. When Conley controls for these factors, American women show just as much interest in casual sex as men do.

Throughout this book, we describe the results from hundreds of studies. Some will confirm your prior beliefs about sex and gender, but others will fly in the face of conventional wisdom and debunk common gender myths. Developing an understanding of sound research methodology should help you learn how to distinguish between accurate and inaccurate claims about gender. Again, as you read this book, we challenge you to examine your existing beliefs about gender and think critically about the research findings that we present before drawing conclusions.

Of course, we do not suggest that there are perfect research methods or studies. Every methodology has flaws and can be legitimately criticized. However, the accumulation of multiple, well-designed studies on a given topic increases confidence in the conclusions. In this chapter, we will explain what makes a *well-designed study* and how even the best studies have limitations. Before doing this, we will analyze what researchers mean when they refer to *sex differences* because this serves as a good starting point for understanding and evaluating gender-specific research. Although the study of sex and gender is much more than the study of differences between women and men (and a goal of our text is to go beyond binary thinking about sex and gender), most of the existing gender research compares these two groups. This limitation led to calls to expand the focus of the field beyond female–male comparisons, or to abandon the use of binary comparisons altogether (Hyde, Bigler, Joel, Tate, & van Anders, 2019; Schellenberg & Kaiser, 2018). By maintaining a binary focus, researchers overlook individuals who identify as neither female nor male. Moreover, focusing on the binary may inadvertently imply that all women (and all men) share similar experiences regardless of their race, class, age, ability, sexual orientation, religion, and culture. You will read more about these issues later in the chapter when we discuss how to address specific methodological challenges and biases common in gender research.

Heterosexual women's reluctance to accept casual sex offers from unfamiliar men stems, at least in part, from their expectations that sex with a stranger is unlikely to be very pleasurable.

Source: © iStockphoto.com/hoozone

SIDEBAR 2.1

A Brief History of Sex Difference Research

In the 19th century, medical researchers interested in sex differences focused mainly on identifying structural brain differences that could explain women's intellectual inferiority to men (Shields, 1975). Around the turn of the 20th century, Helen Thompson-Woolley (1903), one of the first women to receive a doctorate in experimental psychology, criticized the biases in this earlier work and sought to improve the quality of sex difference research in her psychology dissertation. Her carefully designed experiment revealed only negligible sex differences in motor and intellectual abilities. Sex difference research in the 1920s and 1930s focused mostly on emotional and social tendencies and culminated in Terman and Miles's (1936) personality measure of masculinity and femininity. Still, research on sex differences was slow to take root in psychology, likely because most early psychologists were men who generally did not find the research interesting or valuable. This began to change in the United States in the 1970s, when the second wave of the women's movement brought greater attention to women's issues. In 1974, Eleanor Maccoby and Carol Nagy Jacklin published a landmark book titled *The Psychology of Sex Differences*, which reviewed over 1,400 published studies of sex differences. Though Maccoby and Jacklin's (1974) results showed overall evidence of sex similarity, differences emerged in the areas of verbal ability (favoring girls) and math ability, visuospatial ability, and aggression (favoring boys).

WHAT IS THE MEANING OF DIFFERENCE?

Gender researchers often talk about differences between women and men, or between girls and boys. For example, let's return to the finding that girls have higher levels of math anxiety than boys (Stoet et al., 2016). There are three things we cannot determine from this statement alone. First, this only tells us about the average boy and girl, not any individual boy or girl. Obviously, not all girls have higher math anxiety than all boys, and so the statement is true in general, but also not true for everyone. Second, knowing that girls have higher math anxiety than boys reveals nothing about how much boys and girls typically vary from each other *within* their own sex groups. **Variance** is a measure of how far the scores in a distribution vary, on average, from the mean of the distribution, and it plays a big role in how we interpret sex differences. In Chapter 7, you will read about gender and cognitive abilities. Sometimes researchers find no average differences between girls and boys in test scores, but they do find differences in variance. For instance, boys often display more variance than girls do, with more very low-achieving and very high-achieving group members. If we simply look at mean scores, we might miss this critical difference.

Variance
A measure of how far the scores in a distribution vary, on average, from the mean of the distribution.

Third, finding that girls have higher math anxiety than boys does not tell us about the size of the difference. Do girls and boys have very different levels of math anxiety, or do they only differ slightly? Researchers can use statistical methods to quantify the size of sex differences, and we will explain how to do this later in the chapter. For now, the point is that many people, including researchers, tend to give certain meanings to sex differences. Whereas those who take a **maximalist approach** emphasize differences between sex groups, those who take a **minimalist approach** emphasize similarity between sexes (Del Giudice, 2019). As you can see in Figure 2.1, someone with a maximalist approach might envision that girls and boys have completely non-overlapping distributions of math anxiety scores. In contrast, someone with a minimalist approach might envision that the distributions of math anxiety scores of girls and boys are largely similar and overlapping, with girls scoring just slightly higher than boys on average.

As another example, consider a study by Basow and Rubenfeld (2003) that examined sex differences in providing support. In this study, male and female participants imagined that a friend confided a problem to them and then rated their likelihood of each of these possible responses to their friend: (a) giving advice, (b) offering sympathy, (c) changing the subject, (d) sharing a similar problem, (e) joking about it, and (f) telling the friend not to worry. Overall, women rated themselves as more likely than men to give sympathy, and men rated themselves as more likely than women to change the subject. However, on the other four types of support, there were no sex differences. Despite finding sex differences on only two of six total responses, Basow and Rubenfeld interpreted their findings to suggest that men and women communicate

Maximalist approach A tendency to emphasize differences between members of different sex groups and view them as qualitatively different.

Minimalist approach A tendency to emphasize similarities between members of different sex groups.

Figure 2.1   Maximalist and Minimalist Approaches

These theoretical distributions represent (a) the maximalist approach and (b) the minimalist approach. The maximalist approach assumes little or no overlap between the distributions of women's and men's scores. In contrast, the minimalist approach assumes that although women's and men's mean scores differ, there is a lot of overlap in their distributions.

(a)   (b)

——— Men   - - - - - Women
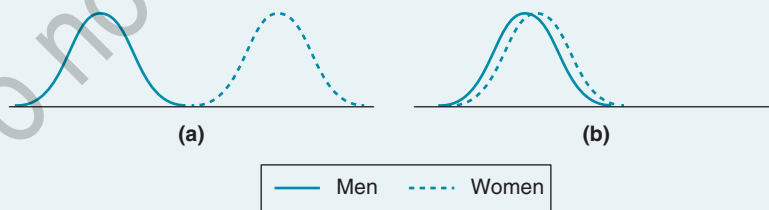
Source: Adapted from Unger and Crawford (1996).

differently, with women prioritizing interpersonal connection and men prioritizing autonomy. Thus, the researchers emphasized difference rather than similarity, which might lead readers to develop a view of women and men as fundamentally different in their communication styles. However, Erina MacGeorge and her colleagues reanalyzed Basow and Rubenfeld's data and found more evidence of similarity than difference in how women and men claimed to offer support (MacGeorge, Graves, Feng, Gillihan, & Burleson, 2004). This becomes apparent when you visualize the data across all six types of support (see Figure 2.2). The sex similarities in communication seem more pronounced than the differences.

Figure 2.2   MacGeorge and Colleagues' (2004) Reanalysis of Basow and Rubenfeld's (2003) Data

When MacGeorge et al. (2004) graphed Basow and Rubenfeld's (2003) data, it made a powerful visual point: The two statistically significant sex differences (in offering sympathy and changing the subject) appear quite small in comparison with the overall pattern of sex similarity that characterizes the data.
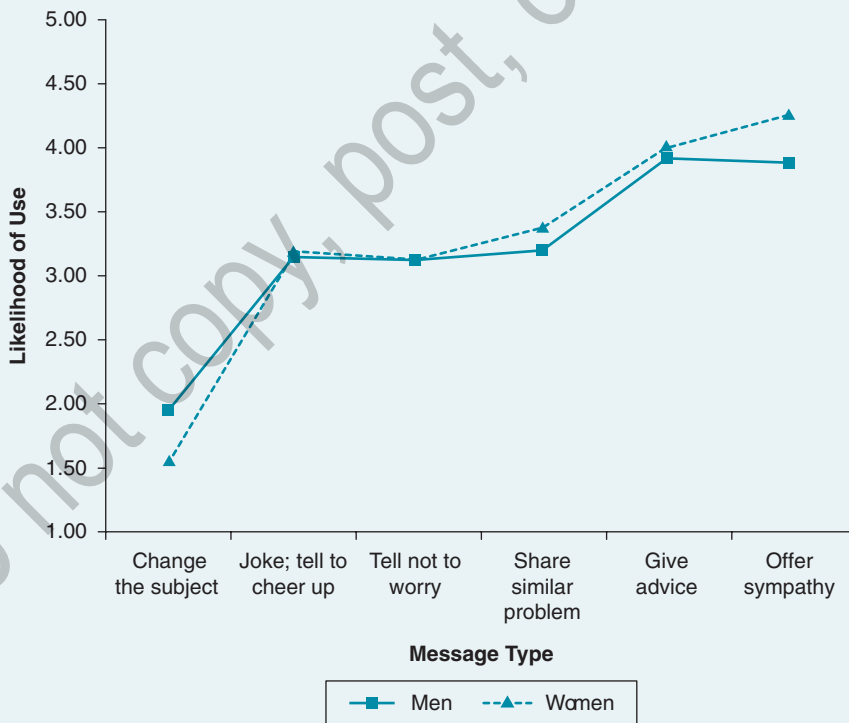

Source: MacGeorge, Graves, Feng, Gillihan, and Burleson (2004).

The maximalist approach has a potential danger in that it encourages people to ignore the overlap that often characterizes people of different sexes. Given this, some minimalist theorists argue that the study of sex differences promotes gender stereotypes and is therefore irresponsible. (For more on this issue, see the "Debate: Should Psychologists Study Sex Differences?") That is, by focusing on differences and ignoring similarities, researchers may perpetuate overgeneralized and exaggerated beliefs about the sexes. Of course, one could also argue that the minimalist approach ignores potentially important sex differences. To conclude that people of different sexes are "mostly alike" may be technically accurate in some cases, but it also fails to acknowledge the differences that do exist. Perhaps the more important point is that if researchers approach the study of sex differences with either a maximalist or a minimalist bias, this bias may influence how they and others portray and interpret their findings. For example, when graphing data, how the researcher sets up the y-axis (by either truncating it or presenting the full range of possible values) can visually exaggerate differences (see Figure 2.3). Try to keep these biases in mind as you read the results of sex difference research described in this book. For our part, we will try to minimize bias in how we visually present sex differences in this book by (when possible) showing the full range of values on the y-axis. Finally, when you encounter sex differences—in this book and in your life—we encourage you to ask yourself, *What does this difference mean*?

Figure 2.3 Sex Differences in Physician Burnout

These graphs show the mean levels of exhaustion and disengagement (two types of burnout) for a sample of female and male Norwegian physicians. Note that the data in both graphs are the same, but the graphs are in different formats. The graph on the left shows the full range of possible burnout levels on the y-axis. In contrast, the graph on the right truncates the y-axis (only reflecting a subset of the possible burnout levels). What differences do you see in the two graphs? Do you think these different presentations might lead readers to interpret the findings differently? Which graph does a better job of depicting the data? Why?
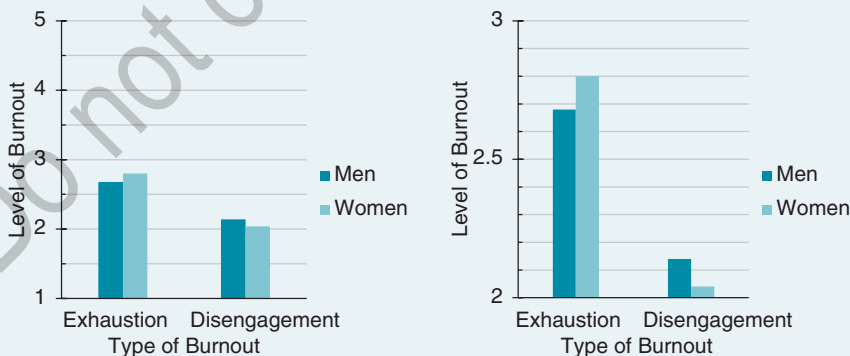
Source: Langballe, Innstrand, Aasland, and Falkum (2011).

DEBATE

Should Psychologists Study Sex Differences?

Should psychologists study sex differences? What important and useful information can be gained by studying sex differences? A debate about these questions emerged in psychology several decades ago and continues to this day (Eagly, 2013; Kitzinger, 1994). Let's examine both sides.

Yes, Psychologists Should Study Sex Differences

Science is the best tool we have to develop accurate understandings of sex differences (Halpern et al., 2007). While it is not perfect, science is a systematic method with built-in checks and balances that decrease error and increase valid findings over time. The knowledge gained from sex difference research, moreover, can help counter gender bias and misconceptions. For example, because "experts" in the 19th century believed women to be intellectually inferior to men, women were often denied access to higher education (Shields, 1975). Research on sex differences in cognitive abilities debunked the myth of women's intellectual inferiority (Halpern, 2012), paving the way for increases in women's access to education. Recent research also suggests that studying sex differences is essential in more effectively diagnosing and treating disorders such as Alzheimer's disease (Nebel et al., 2018). Thus, many believe that the benefits of studying sex differences outweigh the costs, especially in the long run.

Studying sex differences also allows psychologists to identify the contexts in which such differences do or do not emerge, which can assist theory development. For example, classic research examining sex differences in helping behavior found that men, relative to women, showed greater helpfulness toward strangers (Eagly & Crowley, 1986). The researchers hypothesized that this difference reflected social norms of male chivalry, which led them to predict—and find—a larger sex difference when an audience was present to observe the helping behavior. In communicating such findings, researchers demonstrate the complexity of sex differences.

No, Psychologists Should Not Study Sex Differences

Research on sex differences can have the unfortunate consequence of reinforcing gender stereotypes (Baumeister, 1988). This occurs, in part, because of *publication bias* (also known as the *file drawer problem*), which refers to the tendency in the field of psychology to publish studies that find significant group differences more often than studies that do not. Since studies that fail to find differences are less frequently accepted for publication, we may not know as much about sex similarities as we do about sex differences (Hanel, Maio, & Manstead, 2019). Put another way, if what we know about sex differences is based

mostly on published studies, we likely have an exaggerated understanding of these differences.

Furthermore, since the popular media tend toward exaggerating sex differences (because difference is attention-grabbing), researchers should avoid contributing to this by providing media outlets with material on sex differences. By studying and reporting sex differences, psychologists communicate that such differences merit attention. If we want people to focus less on sex and gender, then researchers should lead the way by discontinuing research that calls attention to sex differences (Baumeister, 1988).

Finally, some argue that the current methods used to compare women and men stem from a faulty and limited conceptualization of gender (Schellenberg & Kaiser, 2018). Making broad male–female comparisons excludes intersex individuals and individuals with nonbinary identities and fails to consider how people differ based on other social categories such as race, class, and sexual orientation. Comparing men and women may also imply that gender is a static, measurable quality that resides within individuals. In contrast, some view gender as a dynamic system of behaviors, shaped by societal institutions and practices, that emerges through social interaction (Deaux & Major, 1987; West & Zimmerman, 1987). From this perspective, sex difference research reinforces an overly simplistic and inaccurate view of gender.

Now that you have read the arguments on both sides, what do you think? Should psychologists study sex differences or focus their attention elsewhere? Which perspective makes the most sense? Which evidence do you find most and least convincing? Why?

WHAT IS SCIENCE?

This textbook emphasizes the scientific study of sex and gender. But what exactly makes research on sex and gender scientific? When you think of the word *science*, what images come to mind? People in lab coats holding test tubes? People peering through microscopes at microbial life? When conceptualizing science, most people think of the "hard" or natural sciences (e.g., physics, chemistry, and biology) more readily than the "soft" or social sciences (e.g., psychology, sociology, and political science). Some consider psychology and other social sciences to be less scientific than the hard sciences because human behavior does not follow precise physical and mathematical laws the way that planets or atoms do. Gravity is gravity—its rules do not change—but people are unique and can be difficult to predict.

SIDEBAR 2.2

The Unfalsifiable Freud

The perception of psychology as unscientific may have been bolstered by the work of Sigmund Freud, perhaps the most well-known psychologist in history. Though Freud's psychoanalytic theories of development and personality have been quite influential, many of

(*Continued*)

his ideas are unfalsifiable. A *falsifiable theory* can be disproved with evidence. This means that the researcher must specify a set of conditions that, if they occurred, would clearly invalidate the theory. In other words, the researcher must state, "If my theory is correct, then *x* will happen. If, instead, *y* happens, then my theory is not correct." If there is no set of conditions (*y*) that will be taken as invalidation of a theory, then a theory is not truly scientific. Rather than carefully specifying the patterns of data that could invalidate his theories, Freud tended to focus on evidence that he viewed as consistent with his theories. Nonetheless, his ideas did help to shape early understandings of gender.

So, what is the definition of science? In *Broca's Brain: Reflections on the Romance of Science*, astrophysicist Carl Sagan (1980) writes,

> Science is a way of thinking much more than it is a body of knowledge. Its goal is to find out how the world works, to seek what regularities there may be, to penetrate to the connections of things—from subnuclear particles, which may be the constituents of all matter, to living organisms, the human social community, and thence to the cosmos as a whole. (p. 15)

As Sagan indicates, science is as an ongoing process of discovery, defined more by its methods than by its contents (the specific topics under investigation). Science is a systematic, empirical way of investigating the world in order to identify rules and patterns in the way it works.

The Scientific Method

Although the range of topics studied by scientists is vast, from astrophysics to psychology, the common thread through all science is the use of the scientific method. Thus, to determine which fields do and do not qualify as science, we must examine the type and rigor of the methods used. Let's further examine what this means.

Scientific method A process by which researchers conduct systematic studies in order to test hypotheses derived from theory.

Hypothesis A testable prediction about the outcome of a study, stated in terms of the variables tested.

When using the **scientific method**, a researcher conducts systematic studies to test theory-driven **hypotheses**, or testable predictions about the outcome of a study. The scientific method unfolds in a series of steps, including hypothesis generation, study design, data collection and analysis, results dissemination, and *replication* (repeating a study to determine whether the results will recur). Figure 2.4 provides a general model of the scientific method. A researcher begins by making an observation about the world (e.g., women and men seem to exhibit different body language in public) and then develops a hypothesis (e.g., in public, women generally sit in ways that take up less space compared with men). To test the hypothesis, the researcher designs a study, which may take any number of different forms (more on specific methods in a bit). After gathering and analyzing data, the researcher decides whether the hypothesis has been supported or refuted and then may alter or refine the hypothesis for future

testing. After interpreting the results, the researcher can also develop or refine general theories about the phenomenon. For instance, in finding that men typically (but not always) sit in ways that take up more space than women, the researcher may develop a theory about how people use personal space to signal dominance. To the extent that men generally care more than women about issues of social dominance, they may use more personal space to demonstrate their dominance. This theory can then be used to generate new hypotheses. For example, perhaps individuals who are made to feel relatively powerless (whether male or female) will sit in ways that occupy less space than individuals who are made to feel powerful. Going forward, the researcher could do the following: (a) conduct an experiment in which she manipulates participants' feelings of power to be either high or low and then measures how this impacts sitting behavior, (b) analyze the results, (c) further refine the theory, and (d) generate new hypotheses to test. As this example and Figure 2.4 illustrate, the steps of the scientific method—rather than being linear—occur in a loop. There is no final endpoint in the scientific process, as each new discovery leads to the refinement of theories and hypotheses, which starts the process anew. For an example of how researchers continue to build and refine theories, turn to "Journey of Research: Conceptualizing and Measuring Masculinity and Femininity."

Figure 2.4 The Steps in the Scientific Method

Rather than following a strict linear path, the steps in the scientific method occur in a loop, with no final endpoint.

The Scientific Method
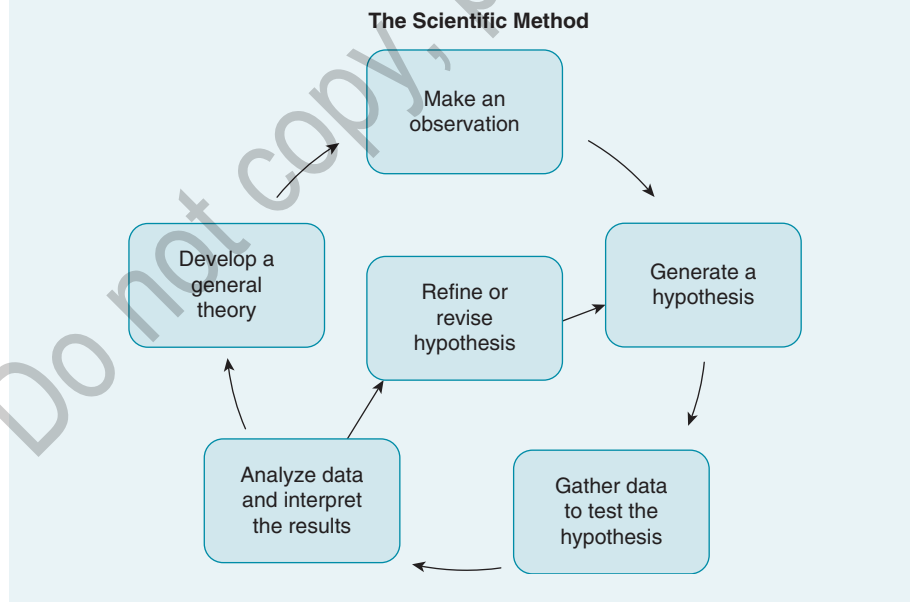
JOURNEY OF RESEARCH

Conceptualizing and Measuring Masculinity and Femininity

Before devising methods to measure any construct of interest, researchers must first clarify its meaning. Consider masculinity and femininity. What is the best way to conceptualize and measure these constructs? As you may recall from Chapter 1, gender psychologists have struggled with this question for quite some time. Here, we summarize three significant shifts in psychological thinking about masculinity and femininity.

Unidimensional Conceptualization

In 1936, Lewis Terman and Catharine Cox Miles published the M–F Test, the first measure of psychological masculinity and femininity. This test had over 400 items that were included based on their ability to distinguish between men and women. For example, on a word association subscale, respondents were presented a target word like *TENDER* and asked to select which of four words (*kind*, *loving*, *meat*, or *sore*) they most associated with the target word. Response options for each item were coded as masculine (meat), feminine (kind, loving), or neutral (sore), and an overall M–F score was calculated for each respondent.

Decades later, Anne Constantinople (1973) published an influential article that identified problems with measures of masculinity and femininity, including the M–F Test. Among other things, she criticized these tests for conceptualizing masculinity–femininity as polar opposites on a single dimension. In other words, Constantinople challenged the assumption that masculinity necessarily meant an absence of femininity (and vice versa), arguing instead that individuals can simultaneously possess elements of both masculinity and femininity.

Two-Dimensional Conceptualization

Building on Constantinople's insights, researchers developed two new scales in the early 1970s: the Personal Attributes Questionnaire (Spence, Helmreich, & Stapp, 1974) and the Bem Sex-Role Inventory (Bem, 1974). These scales included

The best indicators of people's masculinity and femininity include gender-related hobbies and everyday activities.

Source: © iStockphoto.com/michal-rojek; iStockphoto.com/AleksandarNakic

separate subscales of male-typed (M) or agentic traits (e.g., independent and competitive) and female-typed (F) or communal traits (e.g., gentle and kind). Importantly, the M and F trait scales were found to be uncorrelated, meaning that individuals could score either high or low on each set of traits (Bem, 1974; Helmreich, Spence, & Wilhelm, 1981). This led researchers to develop the construct of psychological androgyny, which refers to possessing high levels of both M and F traits.

Multidimensional Conceptualizations

Despite the progress made in the 1970s, researchers soon began questioning whether gender traits alone could capture masculinity and femininity. Janet Spence (1993) proposed that a person's sense of gender reflects not only traits but also attitudes, roles, and interests that may or may not relate to one another. Inspired by Spence's approach, newer theories better capture the multidimensionality of masculinity and femininity. For example, Richard Lippa (2005) assesses masculinity and femininity with what he calls a gender diagnosticity (GD) score. This score refers to the estimated probability that an individual is male or female given the individual's gender-related interests. Lippa (2005) finds that GD scores do a better job—compared to M and F trait scales—at predicting gender-related outcomes, and he argues that the core of masculinity and femininity consists of occupational interests, hobbies, everyday activities, nonverbal behavior, and sexual orientation.

As you can see, overly simplistic conceptualizations of masculinity and femininity from earlier eras gave way to more multifaceted conceptualizations as researchers steadily refined these constructs. This process of revision continues today as researchers question the validity and usefulness of measuring masculinity and femininity as static qualities that reside within individuals (Schellenberg & Kaiser, 2018). It will be interesting to see what comes next.

WHAT ARE THE PRIMARY METHODS USED IN SEX AND GENDER RESEARCH?

Once a researcher identifies a topic of interest and a research question, the next steps involve designing a study to examine the research question, collecting and analyzing data, and communicating the results. When designing a study, the researcher may choose from many different methods, often categorized broadly as quantitative or qualitative. In this section, we examine some common quantitative and qualitative methods used in gender psychology, along with some gender-specific methodological challenges in the research process.

Quantitative Research Methods

Quantitative methods allow researchers to turn variables of interest into numbers that can be submitted to statistical analyses. All of the methods reviewed in this section share the property of relying on numerical data. These methods are summarized in Table 2.1.

Gender diagnosticity (GD) score The estimated probability that an individual is male or female given the individual's gender-related interests. A GD score of .85 means that the individual has an 85% chance of being male and a 15% chance of being female.

Table 2.1 Quantitative Research Methods

Design	Main Features	Description
Experimental	Independent variables, dependent variables, researcher has good experimental control, uses full random assignment	A researcher manipulates an independent variable, randomly assigns participants to different conditions, and measures their standing on a dependent variable. Because there is random assignment and good experimental control, cause-and-effect relationships can be determined.
Quasi-experimental	Independent variables, participant variables, dependent variables, researcher has low to moderate experimental control, uses partial or no random assignment	A researcher identifies pre-existing participant groups (e.g., men and women; two different preschool classrooms) and assigns them to different levels of an independent variable (either with or without random assignment). Examples include *person-by-treatment* and *ex post facto designs*. Because of the lack of full random assignment, these designs have reduced experimental control. Cause-and-effect conclusions are limited.
Person-by-treatment	Independent variables, participant variables, dependent variables, researcher has moderate experimental control, uses partial random assignment	A researcher takes pre-existing participant groups (e.g., women and men), randomly assigns them to different levels of at least one independent variable (e.g., status level), and compares them on one or more dependent variables. Due to incomplete random assignment, these designs have moderate experimental control. Cause-and-effect conclusions are possible, but limited.
Ex post facto	Participant variables, dependent variables, no independent variables, no random assignment, low experimental control	A researcher compares pre-existing participant groups (e.g., women and men; people from different cultures) on one or more dependent variables. Cause-and-effect relationships cannot be determined due to lack of random assignment and low experimental control, resulting in the reverse causation and third variable problems.
Correlational	Continuous (not categorical) participant variables, no random assignment, low experimental control	A researcher measures people's standing on continuous participant variables (e.g., stress and optimism) and examines the strength and direction of the association between them. Cause-and-effect relationships cannot be determined due to lack of random assignment and low experimental control, resulting in the reverse causation and third variable problems. In a cross-sectional design, all variables are measured at one time; in a longitudinal design, variables are measured at multiple points across time.

True Experimental Designs. Well-conducted experiments allow researchers to establish cause-and-effect relationships between variables. Thus, to know whether smiling *causes* happiness or whether effective studying *reduces* test anxiety, you have to conduct an **experiment**. In experiments, the researcher manipulates variables of interest (called **independent variables**) to observe whether this causes changes in outcome variables (called **dependent variables**). For instance, to determine whether dominance causes people to use more physical space, a researcher might ask some people to write about a time when they were socially dominant, while other people write about what they ate yesterday. The researcher might then ask people to sit in a waiting room and unobtrusively observe how much space they take up. In this experiment, the independent variable is dominance and it is manipulated by having people write about either their own dominance or a neutral topic. The dependent variable is people's physical use of space while seated.

Researchers use **random assignment** in true experiments, meaning that each participant has an equal chance of being assigned to each of the different experimental conditions in the study. Why is this important? Participants naturally vary in many ways (e.g., age, race, cultural background, socioeconomic status), and random assignment increases the likelihood that these pre-existing differences are spread out evenly across conditions at the outset of a study, before the manipulation of the independent variable. Thus, by manipulating an independent variable, using random assignment, and holding all other variables constant (to the extent possible), researchers can establish whether an independent variable causes changes in a dependent variable. If a study has good *experimental control*—meaning that no variables other than the independent variable differ systematically across the conditions—then the researcher can confidently conclude that any observed differences in the dependent variable were caused by the manipulation of the independent variable. This ability to determine cause-and-effect relationships leads many to view experiments as the gold standard of the scientific method.

Gender research faces a special challenge when it comes to experimental methods because, strictly speaking, sex cannot be treated as an independent variable. If you cannot easily or ethically assign people into the different conditions or levels of a variable, then it is not a true independent variable. Despite this, many gender researchers conduct true experiments. How is this possible? Although researchers cannot manipulate *actual* sex, they can manipulate the *perceived* sex of a target and measure people's reaction to that target.

Imagine, for instance, a psychologist interested in whether people treat babies differently based on their sex. Using a nonexperimental method, she could observe people as they interact with babies of different sexes, but if she found differences, it would be difficult to know whether those differences were due to the sex of the babies or to some other associated variable (e.g., differences in the babies' clothing or temperament). A better method would be to show adults a baby who is dressed in gender-neutral clothing (e.g., a diaper and a white shirt) after telling half of them that the baby is a girl and half that the baby is a boy. A classic study testing perceptions of

Quantitative methods Methods in which researchers convert variables of interest into numbers and use statistical analyses to test hypotheses. Examples include experimental, ex post facto, quasi-experimental, and correlational designs.

Experiment A type of research design in which a researcher systematically manipulates one or more independent variables to observe whether this causes changes in one or more dependent variables.

Independent variable A variable that is assumed to cause changes in a dependent variable; in an experiment, the independent variable is systematically manipulated by the researcher.

Dependent variable An outcome variable; in an experiment, the dependent variable is the one hypothesized to change as a result of manipulation of an independent variable.

emotionality in infants did just this (Condry & Condry, 1976). Participants watched a video of an infant reacting to a jack-in-the box. Those who thought the infant was male were more likely to describe "his" reaction as anger, whereas those who thought the infant was female were more likely to describe "her" reaction as fear. In this way, the baby's perceived sex is a true independent variable. Consider another example. To determine whether employers have a sex bias in hiring, a researcher could conduct an experiment by randomly assigning real employers to receive nearly identical résumés that differ only in the name and sex of the applicant (e.g., Ana Garcia versus Antonio Garcia). The researcher could then ask employers to rate how competent, hirable, and likable the applicant is and then compare across the résumés. These examples allow for tests of cause-and-effect relationships because the researcher manipulates perceived sex in an experimental setting while attempting to hold all other variables constant.

Gender researchers could also conduct experiments by manipulating a variable that is related to sex and gender. For instance, suppose a researcher hypothesizes that men exhibit more dominance than women in the workplace because men more frequently hold high status positions in these settings. While the researcher cannot manipulate participants' sex, she can systematically manipulate status, for example, by randomly assigning both women and men to play either a supervisory role (higher status) or a subordinate role (lower status) in some work-related task. If status causes differences in dominance behavior, then both women and men should exhibit more dominance when playing a supervisor role as compared to a subordinate role. However, what if displaying dominance in the workplace causes men to attain high status positions in the workplace, instead of the other way around? If this were really the case, then the study just described would yield *null results*, or results that do not support the hypothesis. At this point, the researcher might design another study to examine the alternative cause-and-effect relationship between dominance and workplace status. The point is that a well-designed experiment allows researchers to determine which cause-and-effect relationships have merit and which do not. Of course, experiments are not always a viable option. In many cases, gender researchers turn to pseudo-experimental and nonexperimental designs, which we examine next.

In contrast to true experiments, pseudo-experiments are research designs that appear experimental on the surface but lack one or more of the features of true experiments. Examples of pseudo-experimental designs include quasi-experiments and ex post facto designs (see Table 2.1).

What sex is this baby? We have no idea, and neither would participants in an experiment. But consider this important question: Would people treat this baby differently depending on whether they believed it to be female, male, or intersex?

Source: © iStockphoto.com/DONOT6

Quasi-Experiments. **Quasi-experiments** come in many forms, though the unifying thread is that the researcher lacks full control over at least one independent variable manipulation, usually due to a lack of random assignment (Pelham & Blanton, 2018). In one common type of quasi-experiment, the researcher selects pre-existing groups of participants and exposes them to different levels of an independent variable without assigning individual participants to conditions randomly. For example, a researcher might select two different but comparable groups of students (e.g., two different preschool classrooms), and expose them to different experiences. In one classroom, the teacher emphasizes gender by having the children line up and do activities by sex, and in the other classroom, the teacher does not mention gender. After two weeks, the researcher measures the children's levels of gender stereotyping (e.g., Hilliard & Liben, 2010). Note that individual children in this study are not randomly assigned to a classroom. Now suppose that children in the first classroom, which emphasized gender, have higher gender stereotyping scores than children in the second classroom at the end of the two-week period. Can we conclude that emphasizing gender caused increases in gender stereotyping in the preschoolers? What other factors might have caused this outcome? In this case, without using random assignment to the different levels of the independent variable, it is difficult for the researcher to identify the true cause of any observed outcomes.

Another type of quasi-experiment, called a **person-by-treatment design**, offers researchers more control than the study just described. In a person-by-treatment design, a researcher selects people who differ on some **participant variable**—a naturally occurring feature of participants such as sex, gender identity, or cultural background—and then randomly assigns them to different conditions of an independent variable. Person-by-treatment designs are common in gender research because they allow researchers to compare, for example, how women and men react differently to some manipulated variable. The study described earlier in this chapter, in which a researcher randomly assigns women and men to play either a supervisory or subordinate role, uses a person-by-treatment design. These designs allow some degree of cause-and-effect conclusion. For instance, imagine that men who got assigned to a supervisory role displayed the same level of dominance as men assigned to a subordinate role, but that women in the supervisory role displayed much more dominance than women in the subordinate role. Given this pattern of results, we could conclude that the status of assigned roles has a larger effect on women's—as compared to men's—dominance behavior. Of course, we still cannot conclude that people's sex causes them to react differently to the status manipulation, given that sex is not a true independent variable. There may be other, unmeasured variables (e.g., self-confidence) that cause women and men to respond differently based on levels of status.

Ex Post Facto Designs. In **ex post facto designs**, researchers compare groups of people (e.g., smokers/nonsmokers, Southerners/Northerners, women/men) to see whether they differ on some dependent variable of interest. For example, a researcher would use this type of design to test the hypothesis that women smile more than men. Because they

Quasi-experiment A design that mimics the appearance of a true experiment, but in which the researcher lacks control over one or more manipulations.

Person-by-treatment design A quasi-experimental design involving at least one participant variable and at least one true independent variable with random assignment.

Participant variable A naturally occurring feature of research participants (e.g., sex, personality, cultural background) that is measured in a study rather than manipulated.

Ex post facto design A nonexperimental design in which groups of people who differ on a participant variable (e.g., sex) are compared on some dependent variable.

lack random assignment to the levels of a true independent variable, ex post facto designs do not allow for cause-and-effect conclusions. Despite this limitation, ex post facto studies can lay the foundation for future research that clarifies or explains the results. For instance, we know that sex differences in smiling do not emerge until approximately age 13 (Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006). This suggests that sex differences in smiling may have less to do with people's sex and more to do with other factors that covary with sex, such as gender socialization processes that encourage people to adopt gender role norms.

Note that in either type of experimental design (true or pseudo) with more than one independent or participant variable, researchers can examine **interaction effects**, which occur when the strength or direction of the association between one independent (or participant) variable and a dependent variable differs as a function of another independent (or participant) variable. Let's return to the example of sex differences in smiling. If adherence to gender role norms plays a role in smiling, then sex differences in smiling should be larger when people think they are being observed by others. To test this, LaFrance, Hecht, and Paluck (2003) compared the smiling behavior (dependent variable) of men and women (a participant variable) after randomly assigning them to contexts in which they were either observed or not observed (an independent variable). The results of this person-by-treatment study followed an interaction pattern: sex differences in smiling were greatest when people believed that they were being observed. In other words, the relationship between sex and smiling behavior differed as a function of observation condition. See Figure 2.5 for another example of an interaction effect.

Correlational Designs. In correlational designs, researchers test hypotheses about the strength and direction of relationships between pairs of variables. In contrast to ex post facto designs, which compare two or more groups on some variable of interest (e.g., whether women tend to have more body image problems compared with men), the most prevalent correlational design examines the relationships between continuous variables (e.g., feminine personality traits and body image problems). While correlations do not allow conclusions regarding cause-and-effect relationships, they are useful because they allow researchers to make predictions. If two variables (*x* and *y*) are correlated, then you can predict a person's score on *y* given that person's score on variable *x*. However, the accuracy of predictions based on correlations differs as a function of the strength of the correlation. The stronger the correlation between two variables, the more accurate the prediction.

When they believe that they are being observed—as people posing for a picture clearly do—women tend to smile more than men.

Source: © iStockphoto.com/guvendemir

Figure 2.5 An Interaction Effect

In an interaction effect, the strength of the association between an independent or participant variable and a dependent variable differs as a function of another independent or participant variable. Here, the researchers randomly assigned women and men to receive feedback that either threatened or affirmed their gender status and then measured their aggressive cognitions. As you can see, the effects of the feedback manipulation (independent variable) on aggressive cognitions (dependent variable) were stronger for men than they were for women (participant variable).
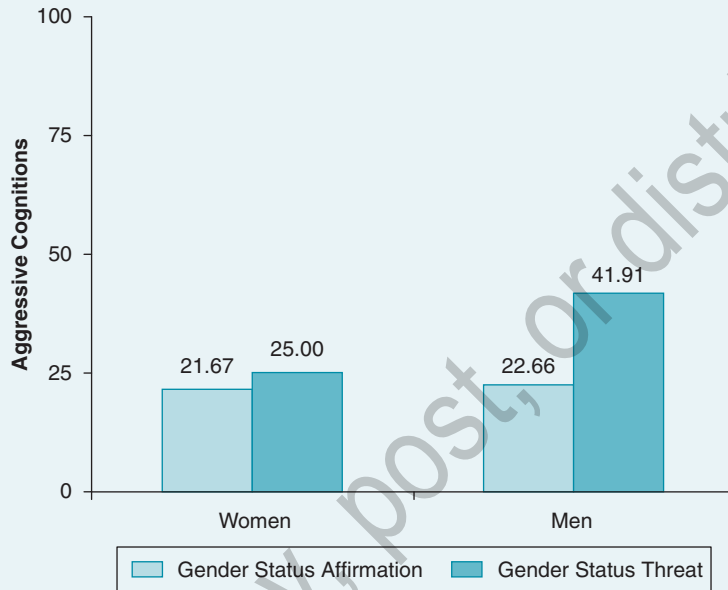
SIDEBAR 2.3

A Correlation Refresher

Pearson correlation coefficients, or *r* values, range from –1.0 to +1.0, and the farther the r value is from 0 in either direction, the stronger the relationship. In terms of direction, *r* values can be positive or negative. *Positive correlations* indicate that the variables change in the same direction (i.e., as one increases, the other increases, and vice versa). For example, as agentic traits increase, self-esteem increases, and as agentic traits decrease, self-esteem decreases. *Negative correlations* indicate that the variables change in opposite directions (i.e., as one increases, the other decreases, and vice versa). For example, as stress increases, body esteem decreases, and as stress decreases, body esteem increases.

Let's consider an example of a correlational design. Kevin Swartout (2013) surveyed college men about their attitudes toward women and sexual aggression, as well as their perceptions of their peers' attitudes toward women and sexual aggression. He found that perceived peer rape-supportive attitudes were positively correlated with men's own hostility toward women. It might be tempting to assume a causal relationship here—for instance, that hanging out with peers who are believed to have hostile attitudes toward women causes men to develop hostile attitudes themselves. However, we cannot draw this conclusion from a correlational design. As shown in Figure 2.6, there are at least two alternative possibilities. First, the possibility of **reverse causation** means that the causal relationship might be the reverse of what is initially assumed. That is, rather than perceived peer attitudes (*x*) causing individual men's attitudes (*y*), men's hostile attitudes toward women (*y*) might lead them to associate with others whom they perceive as like-minded (*x*). Second, the **third variable problem** means that some unmeasured third variable (*z*) could be responsible for the association between two correlated variables (*x* and *y*). For instance, perhaps men's adherence to male gender role norms (*z*) shapes men's attitudes toward women (*x*) and leads them to associate with peers whom they perceive to hold similar attitudes (*y*). So, it may appear that peer attitudes influence individuals' attitudes toward women, but really the unmeasured variable (*z*, adherence to male gender role norms) may be causing both individual attitudes (*x*) and peer attitudes (*y*).

Figure 2.6 Problems With Determining Cause-and-Effect Relationships in Correlational Designs

The reverse causation problem (does *x* cause *y*, or does *y* cause *x*?) and the third variable problem (does *z* cause both *x* and *y*?) make it impossible to draw cause-and-effect conclusions from correlational findings.
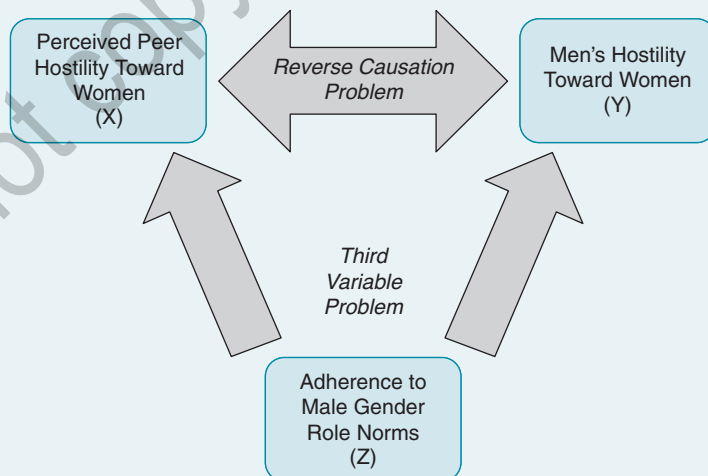
Although uncertainty regarding cause-and-effect relationships is an inherent problem of correlational research, longitudinal designs can help address the problem of reverse causation. In *longitudinal designs*, researchers follow people over time and measure variables at multiple points, whereas in *cross-sectional designs*, researchers measure variables at one point in time. For example, suppose a researcher predicts that the amount of contact between heterosexual and gay people correlates with heterosexual people's attitudes toward gay people. To test this, the researcher conducts a cross-sectional study and finds a positive correlation between amount of intergroup contact and heterosexual people's attitudes toward gay people (the greater the contact, the more positive the attitudes). Due to the possibility of reverse causation and the third variable problem, the researcher cannot conclude that the amount of contact itself causes changes in heterosexual people's attitudes. However, a longitudinal design can help reduce this ambiguity. To this end, Gregory Herek and John Capitanio (1996) measured both contact with and attitudes toward gay people over a 2-year period in a sample of heterosexual U.S. adults. They found that more contact with gay people at Time 1 predicted increases in favorable attitudes toward gay people 2 years later at Time 2. We know that Time 2 attitudes could not have caused the Time 1 contact that occurred 2 years earlier, so the much more likely conclusion is that Time 1 contact caused the changes in Time 2 attitudes. However, Herek and Capitanio (1996) also found that more favorable Time 1 attitudes predicted increases in contact with gay people at Time 2. Thus, the causal relationship between these variables likely goes both ways: More contact leads to more favorable attitudes, and more favorable attitudes also lead to more contact.

Although longitudinal designs do not fully safeguard against reverse causation problems, they can increase confidence in particular directions of causality. They cannot, however, address the third variable problem because some unmeasured variable could always account for an observed association in correlational designs. For instance, perhaps having a more agreeable personality causes heterosexual people to have more daily contact with, and hold more favorable attitudes about, gay people.

STOP AND THINK

In the media, it is easy to find examples of correlational research findings mistakenly described in cause-and-effect terms. "Dividing housework equally in marriage prevents divorce!" "Cuddling after sex increases relationship satisfaction!"

"Playing video games makes teenagers smarter!" Why is this mistake so frequently made? Why is it problematic? If it were your job to decrease this mistake in the popular media, what strategies would you use?

Qualitative Research Methods

Although quantitative methods allow for a great deal of precision, they are not without weaknesses. For example, there may be times when the richness and complexity of human behavior cannot be reduced to numbers. In such cases, researchers may use

Qualitative methods
Methods in which researchers collect in-depth, non-numerical information in order to understand participants' subjective experiences within a specific context. Examples include case studies, interviews, and focus groups.

qualitative methods (see Table 2.2 for a summary). Rather than relying on numerical data and statistical analyses, **qualitative methods** allow in-depth interpretations of situations, with an emphasis on how the individuals who are being studied make sense of their own experiences in context. Though qualitative methods are not one unitary approach, there are some unifying themes. In qualitative studies, researchers tend to emphasize depth over breadth, subjective interpretations over objective reality, and contextualized understandings over universal truths (Gergen, 2010).

Qualitative methods are well established in disciplines outside of psychology, such as sociology and education, but what about within psychology? Although qualitative methods have been used by many prominent psychologists (e.g., Sigmund Freud, William James, Jean Piaget) and they are widely used in some areas of psychology (e.g., feminist psychology), quantitative methods remain the dominant research paradigm in mainstream psychology (Wertz, 2014). This has been shifting, however, with qualitative methods gaining some ground. In 2011, the American Psychological Association (APA) established the Society for Qualitative Inquiry in Psychology (http://qualpsy.org) as a section of Division 5. Three years later, the APA published *Qualitative Psychology*, its first journal dedicated solely to psychological research that uses qualitative methods such as case studies, interviews, and focus groups.

Case Studies. In a case study, the researcher conducts an in-depth investigation of a single entity, usually a person, although case studies are sometimes conducted on a group or event (such as a natural disaster). As an example, Roe-Sepowitz, Gallagher, Risinger, and Hickle (2014) did a case study of female pimps charged with child sex-trafficking crimes

---

Table 2.2   Qualitative Research Methods

Design	Main Features	Description
Case study	In-depth, individualized investigation	A researcher conducts an in-depth investigation of a single entity such as a person, group, or event. Data provide rich detail but may lack generalizability, and interpretations may differ based on researchers' perspectives.
Interview	Open-ended questions	A researcher asks participants a series of open-ended questions in an unstructured, semi-structured, or structured format.
Focus group	Open-ended questions, group format, moderator	A moderator conducts an interview in a semi-structured group format.
Narrative research	Written documents, texts	A researcher analyzes the stories that people tell (e.g., in journals, autobiographies, and letters).
Ethnography	Immersion within a culture or subculture	A researcher attempts to understand the meanings and practices of a particular culture or subculture by living with the group for an extended period of time.

---

in the United States. By examining arrest and court records, personal histories, and media releases for each pimp, the researchers were able to describe different types of female pimps and develop a better understanding of their role in this male-dominated industry. Although case studies provide rich detail about the cases under study, their results tend to lack **generalizability**, which means that it is difficult to generalize the findings to the larger population. Moreover, the interpretation of the results can vary widely based on the perspectives of the researchers. As you will see, these strengths and weaknesses are associated with the interview and focus group methods as well.

Generalizability
The extent to which the findings of a study would apply beyond the sample in the original study to the larger population.

Interviews. Interviews typically involve asking participants (either individuals or groups) to answer open-ended questions that vary in how structured versus unstructured they are. As one example, Chen and colleagues interviewed 201 transgender military veterans about both the challenges and the strengths associated with their unique identity (Chen, Granato, Shipherd, Simpson, & Lehavot, 2017). To identify themes, the researchers asked each person three open-ended questions and then analyzed their responses. While negative themes included the discrimination and stigma that the veterans faced from both the outside world and within the military, positive themes included personal resilience, authenticity, and pride in both their gender expression and their ability to serve their country. Note that a couple of years after this study was published, a Trump administration policy went into effect that banned transgender individuals from enlisting in the U.S. military and banned currently enlisted transgender service members from undergoing surgery or taking hormones to transition (Sonne & Marimow, 2019). And yet, transgender adults are disproportionately more likely than members of the general population to serve in the U.S. military, and transgender veterans are twice as likely to die from suicide compared to cisgender veterans (Tucker, 2019). Interview research with transgender service members will be important in the coming years, in order to examine how the 2019 ban is affecting this already vulnerable group.

Focus Groups. Focus groups, or interviews conducted in a group format, are often guided by a moderator. On principle, qualitative researchers seek to represent marginalized groups in their research, and the focus group format serves this purpose well by convening people to describe their experiences in their own voices (Gergen, 2010). For example, Parra-Cardona, Córdova, Holtrop, Villarruel, and Wieling (2008) used a focus group format to study the parenting experiences of foreign-born and U.S.-born Latinx parents in the United States. By conducting group sessions in the preferred language of the participants at familiar community sites, the researchers aimed to make the participants feel comfortable describing their experiences. Their findings revealed that the two groups were similar in many ways, although the foreign-born parents felt greater language barriers and isolation.

Mixed Methods

You may wonder which method—qualitative or quantitative—is more effective or more "scientific." Taking a historical look, people's thinking on this question has evolved over

In a focus group, a moderator guides a group interview with people who share similar experiences.

Source: © iStockphoto.com/asiseeit

time. As mentioned, quantitative methods have dominated the field of psychology, and researchers did not begin to advocate for the use of qualitative methods in psychology until the 1960s (Wertz, 2014). This push gained momentum in the 1970s as feminist psychologists criticized quantitative research approaches on several grounds (Eagly & Riger, 2014). For example, critics suggested that quantitative methods were *androcentric*, meaning that they treated men and men's experiences as universal while viewing women and women's experiences as deviations from the male norm. This was reflected in the fact that quantitative research was conducted primarily by men, used only male participants, and assumed a male standard for all humans. Moreover, the emphasis in quantitative methods on experimental control and numerical reductionism implies an objectivity and value neutrality that some view as misleading. In fact, no type of research can be completely objective and value neutral because all research is conducted by humans who inevitably bring their biases to their work. In pushing for the use of qualitative methods, feminist psychologists emphasize the idea of *reflexivity*, which means recognizing and acknowledging that the values of the researcher play an active role in shaping the design, findings, and interpretations in any study (Gergen, 2010). The push toward qualitative methods also emphasized the need to include the voices and perspectives of marginalized groups, not just those of people in positions of privilege. Ultimately, this debate led many to recognize both qualitative and quantitative methods as equally effective and scientific (Sale & Thielke, 2018). The primary difference between them is thus not *whether* they produce valid knowledge but *how* they produce this knowledge.

Mixed-methods approach A research approach that combines both qualitative and quantitative methods within the same study or same program of research.

Today, many psychology researchers capitalize on the strengths of both approaches by using **mixed methods** (R. B. Johnson, Onwuegbuzie, & Turner, 2007). In mixed-methods approaches, researchers may use qualitative methods as a first step to develop ideas or hypotheses that they later test with quantitative methods, or they may integrate qualitative and quantitative methods within the same study. For example, Mary Crawford and Michelle Kaufman (2008) did a case study analysis of 20 girls rescued from the sex-trafficking industry in Nepal and then statistically summarized the main themes in the cases. By examining the girls' case files, the researchers developed a detailed picture of the typical trafficking experiences of Nepali girls as well as the common behavioral and physical symptoms experienced by survivors. Then, by quantitatively analyzing data regarding survivors' experiences, the researchers provided empirical evidence that survivors of sex trafficking can be successfully reintegrated into their communities.

As social psychologists, we (the authors of this book) are trained in quantitative methods, and you may notice an emphasis on quantitative over qualitative methods in the research that we cite. Despite this, we are proponents of mixed-methods research because single-method approaches of any kind tend to yield an incomplete picture.

When trying to understand a phenomenon as complex as gender, approaching it from multiple perspectives and methods can lead to a more complete understanding than would be afforded by a single-method approach.

HOW DO WE DRAW CONCLUSIONS FROM MULTIPLE STUDIES?

Ideally, if a topic is important, many researchers will investigate it. What happens when different studies do not come to the same conclusion? Researchers can use methods that combine the results of many individual studies to look at overall trends in the results. This allows researchers to draw broad conclusions and to identify why inconsistencies sometimes emerge across individual studies. A **meta-analysis** is a quantitative technique for analyzing the results across a set of individual studies. In a meta-analysis, researchers compute an **effect size**, which quantifies the magnitude and direction of a difference between groups or the strength of a relationship between variables. Let's look more closely at effect sizes and how to interpret them.

Meta-analysis
A quantitative technique that allows researchers to integrate research findings across a large collection of individual studies.

SIDEBAR 2.4

Getting Very Meta

Just as a meta-analysis summarizes the results of a set of individual studies, a *second-order meta-analysis* (or *meta-synthesis*) summarizes the results of a set of meta-analyses (Zell, Krizan, & Teeter, 2015). Of course, second-order meta-analysis is only possible when a given topic is studied enough that there exist multiple meta-analyses of the research. Why would a researcher want to conduct a second-order meta-analysis? Researchers who conduct individual meta-analyses may have different decision criteria for which individual studies to include, or they may focus on some subsets of studies while ignoring others (given their research questions and interests). These practices can introduce error. A second-order meta-analysis can reduce some sources of error and allow an even broader, more comprehensive view of an entire body of research findings.

Effect Sizes

Many gender studies compare the responses of women or girls to those of men or boys. For example, a study might compare the reading comprehension scores of girls and boys on a standardized test. If the scores show a statistically significant difference (that is, a difference that is very unlikely to result from chance), the researcher will conclude that there is a sex difference in reading comprehension. But finding a statistically significant sex difference still leaves questions unanswered. Is the difference large or small, meaningful or trivial? To address these questions, researchers calculate effect sizes, which quantify the magnitude of research findings. Whereas statistical significance tells us how likely it is that an observed difference occurred by chance, the effect size tells us how large or small the effect is.

The ***d* statistic**, one common measure of effect size, quantifies the difference between two group means (averages) in standardized units. This statistic can be calculated in a single study, or in a meta-analysis across a set of studies. For example, the *d* statistic in a single study of sex differences in reading comprehension would quantify the difference between the average male and female reading scores. In a meta-analysis, the *d* statistic would express the average reading comprehension sex difference across all studies included in the meta-analysis.

To calculate *d* (see Figure 2.7), you would subtract the average reading comprehension score for girls from the average reading comprehension score for boys, and then divide this difference by the pooled standard deviation (a weighted average of each group's standard deviation). Similar to variance, the **standard deviation** is a measure of how far the scores in the set differ, on average, from the mean value. In fact, the standard deviation of a given set of scores is calculated by taking the square root of the scores' variance.

Table 2.3 shows how to interpret *d* values, in terms of whether a given effect size is close-to-zero, small, medium, large, or very large. Reilly, Neumann, and Andrews (2019) performed a meta-analysis of sex differences in reading comprehension in over 3 million students in Grades 4, 8, and 12 in the United States from 1988 to 2015. They found *d*

Figure 2.7   Cohen's *d*

This is the formula for computing Cohen's *d* statistic, formatted for sex difference research. The *d* value is an estimate of the size of the effect. See Table 2.3 for information about how to interpret the *d* value.

$$d = \frac{M_m - M_f}{SD_{pooled}}$$

Source: J. Cohen (1988).

Note: M_m = mean of male scores; M_f = mean of female scores; SD_{pooled} = pooled standard deviation (weighted average of each group's standard deviation).

Table 2.3 Interpreting Effect Sizes
The column on the right provides descriptive labels to accompany various ranges of *d* values (left). This helps researchers interpret the magnitude of effects. The *d* values in this table are expressed in absolute value terms. The sign (+ or –) of the *d* statistic is irrelevant in evaluating its size.

d Value	Description of Size
0.00–0.10	Close-to-zero
0.11–0.35	Small
0.36–0.65	Medium
0.66–1.00	Large
> 1.00	Very large

Source: Adapted from Hyde (2005).

values ranging from -0.19 to -0.32. This means that, averaged across all of the assessments, there was a small sex difference favoring girls in students' reading comprehension scores. It also means that boys, on average, were slightly less than one-third of a standard deviation lower than girls in reading comprehension. You will read more about sex differences in verbal abilities in Chapter 7 ("Cognitive Abilities and Aptitudes").

We will report *d* statistics many times throughout this book, so take a moment to familiarize yourself with Table 2.3. Note that effect size descriptions range from *close-to-zero* to *very large* (Hyde, 2005). However, the size of an effect does not always correspond with its importance: A small effect can have important consequences, and a large effect may be unimportant in the grand scheme of things (Funder & Ozer, 2019). For example, a small effect size associated with sex differences in reading scores might have important real-world consequences if teachers use these scores to determine student placements into advanced or remedial classes.

SIDEBAR 2.5

An Androcentric Convention

Note that throughout this book, negative *d* values mean that girls or women score higher than boys or men on the variable of interest, and positive *d* values mean that girls or women score lower than boys or men. This convention stems from an androcentric formula that treats men's scores as the standard and subtracts women's scores from that standard.
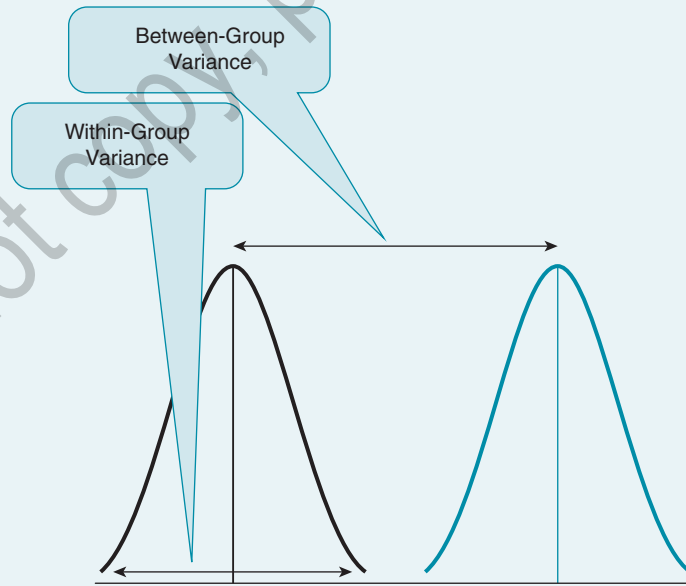
Overlap and Variance

Another way of thinking about effect sizes is in terms of how much overlap exists in the two distributions being compared. More overlap (i.e., more similarity) between two distributions yields a smaller effect size, whereas less overlap (i.e., less similarity) yields a larger effect size. Understanding two types of variance helps clarify this (see Figure 2.8). **Within-group variance** reflects how spread out the values are among people within the same group. For instance, although the average life expectancy of women is 74.2 years, this life expectancy varies widely across countries, from a high of 87.1 years in Japan to a low of 53.8 years in Sierra Leone (World Health Organization, 2019). This represents a fairly large amount of within-group variance. On the other hand, **between-group variance** reflects the difference between the average values of different groups (e.g., men, with an average life expectancy of 69.8 years, die about 4.4 years earlier than women, on average). Compared to the amount of within-group variance in women's life expectancy, the between-group variance is relatively small.

When distributions have relatively large between-group variance and small within-group variance, there is little overlap between the distributions, and the effect

Figure 2.8   Within-Group and Between-Group Variance

Within-group variance refers to how spread out the values are within a given group. Between-group variance refers to the difference between the average values of each group.

size is large. Conversely, with small between-group variance and large within-group variance, there is a lot of overlap, and the effect size is small. As shown in Figure 2.9, a small effect size ($d = 0.20$) means that there is a lot of overlap (85%) between the two distributions, a medium effect size ($d = 0.50$) means that there is a moderate amount of overlap (67%), and a large effect size ($d = 0.80$) means that there is relatively less overlap (52%) between the distributions.

To provide some concrete examples of sex differences with different levels of overlap, interrupting behavior has a small effect size ($d = 0.15$), risk-taking has a medium effect size ($d = 0.49$), and empathy has a large effect size ($d = -0.91$; Archer, 2019).

Figure 2.9 Overlap for Distributions Given Different Effect Sizes

These figures show how much overlap occurs between two distributions when the effect size for the difference between the groups is small (top), medium (middle), and large (bottom).
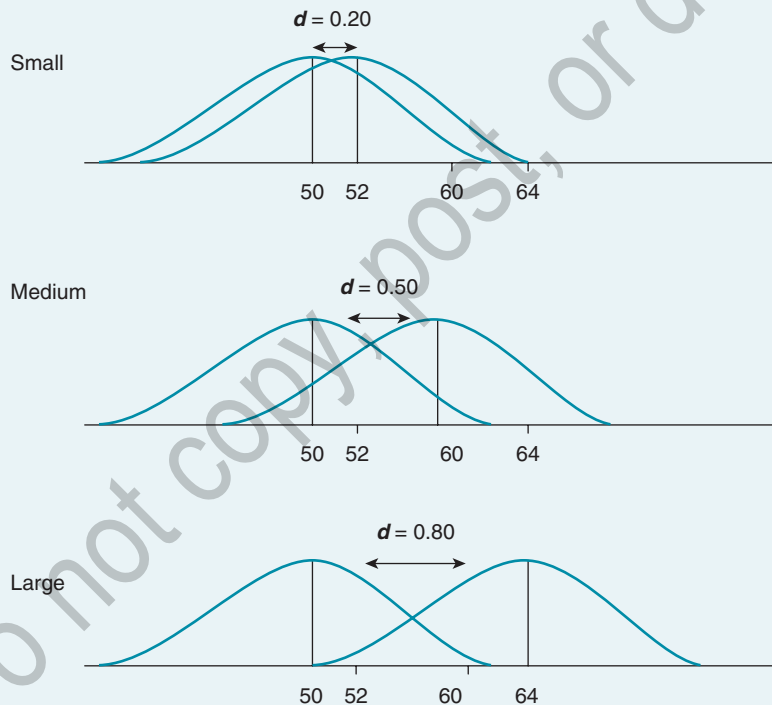
Source: J. Cohen (1988).

Note:

A. *Small Effect Size (d = 0.20):* 85% overlap in two distributions.

B. *Medium Effect Size (d = 0.50):* 67% overlap in two distributions.

C. *Large Effect Size (d = 0.80):* 52% overlap in two distributions.

Let's consider what overlap in distributions means, taking the medium effect size in risk-taking as an example. The average man is more prone to risk-taking than the average woman; however, with about 67% overlap between the male and female risk-taking distributions, it would not be unusual for a given woman (selected at random) to be more risk-prone than a given man. In fact, for medium effect sizes, sex accounts for only about 6% of the total variance in the variable of interest, which means that 94% of the population variation in risk-taking is accounted for by something other than people's sex.

A few sex differences fall in the very large range. Examples include physical traits and abilities such as throwing velocity (d = 1.92; Lorson, Stodden, Langendorfer, & Goodway, 2013), occupational preferences (d = 1.40; Lippa, 2010), and tendency to commit extreme violence such as homicide (d = 2.54) or rape (d = 2.32; Archer, 2019). Many psychological sex differences, however, are in the small or close-to-zero ranges (Hyde, 2005), a point we will revisit shortly. For most psychological variables, then, female and male score distributions overlap substantially, even when statistically significant differences emerge between the average scores of women and men. Why does this matter? When talking about sex and gender, people often focus on the between-group variability and ignore the within-group variability. This leads to the maximalist bias noted earlier, in which people overemphasize differences between people of different sex groups.

STOP AND THINK

Suppose you were a gender researcher and you conducted three different meta-analyses. Consider the following research outcomes: (a) a d statistic of –0.43 in spelling ability, (b) a d statistic of +0.01 in life satisfaction, and (c) a d statistic of +0.28 in physical aggression. What do these three effects tell you about the size and direction of the difference between men and women for spelling ability, life satisfaction, and physical aggression?

Beyond Overall Effect Sizes

Compared to single studies, meta-analyses allow researchers to uncover patterns across multiple studies in order to provide meaning and coherence, which can lead to theory generation and refinement. This is especially useful when individual research findings are inconsistent. For example, consider gender and leadership. For years, researchers have been interested in the qualities of effective leaders, a particularly important topic given women's underrepresentation in high-level leadership positions in the United States and around the world (Noland, Moran, & Kotschwar, 2016). Research on gender and leadership, however, often produces inconsistent findings that differ from one study to the next. To make sense of these inconsistent findings, Alice Eagly and her collaborators conducted a meta-analysis of 76 leadership studies (Eagly, Karau, & Makhijani, 1995). Overall, they found that men and women were equally effective as leaders, with a close-to-zero effect size (d = –0.02). But when they examined contextual factors, an interesting pattern emerged. Men were more effective leaders in leadership roles in

highly male-dominated contexts, such as the military (d = 0.42). In contrast, women were more effective leaders in leadership roles in less male-dominated contexts, such as education (d = –0.11). In other words, Eagly and her collaborators identified an interaction effect, such that the association between participant sex and leadership effectiveness differed as a function of context. Had they focused only on the overall finding of no sex difference in leadership effectiveness, they would not have detected the more nuanced interaction pattern.

WHAT ARE SOME BIASES COMMON IN SEX AND GENDER RESEARCH?

No research is completely free of bias and error, and sex and gender research comes with some of its own unique challenges and biases. With *researcher bias*, researchers behave in subtle ways that influence the outcome of a study. For example, in a meta-analysis of sex differences in intrusive interruptions, K. J. Anderson and Leaper (1998) reported that studies with female first authors found larger sex differences (favoring men) in interruptions than did studies with male first authors. This raises the possibility that female and male researchers may have subtly—and perhaps without awareness—acted in ways that confirmed what they expected to find (for example, in choosing how to measure interruptions or in choosing how to analyze or interpret data). In another form of error known as *participant bias*, participants' responses are influenced by what they think the researcher expects. For example, in a classic study of menstrual symptoms, women directly informed about the researchers' interest reported significantly more menstrual symptoms than women not informed about the researchers' specific interest (AuBuchon & Calhoun, 1985). In this section, we show how bias can enter the research process at any step: in identifying the research question, designing the study and collecting data, and interpreting and communicating the results.

Identifying the Research Question

Try as they might to be objective, researchers have values and beliefs that can introduce bias into the kinds of research questions that they ask. For example, some gender researchers frame their research questions from the perspective of a *female deficit model*, which is the tendency to view sex differences as arising from something that women lack (Hyde, 1994). The female deficit model is rooted in *androcentrism*, which is the tendency to view men as the universal or default for the species and women as exceptions in need of explanation (Bem, 1993). To illustrate, the question of whether "girls lack math abilities compared with boys" is framed within the female deficit model. To move away from that model, researchers might ask, Under what conditions do girls and boys perform differently on math exams? As you read in the chapter opener, girls tend to have higher levels of math anxiety than boys (Stoet et al., 2016), which can impair their performance. Similarly, negative stereotypes about women's math abilities can lower their performance on math tests, while positive stereotypes about men's math competence can inflate their math performance (Danaher & Crandall, 2008). A researcher who operates

Until the 1960s, it was standard in psychology for researchers to use only men as participants and then generalize their findings to "all people." What do you think? Are these people the default for the human species?

Source: Getty Images / Heritage Images / Contributor

within the female deficit model, however, may fail to ask questions about contextual or social factors that can drive sex differences in math performance.

At times, the questions that researchers ask reflect gender differences in power, status, and social roles. For instance, researchers have long examined how working parents balance the demands of their work and home lives and the stressors that result from competing pressures and time constraints. But which parents do you think receive the bulk of research attention on this topic? Far more of this research focuses on mothers, which reflects long-standing labor divisions that more often cast women in the role of primary caregivers at home and men in the role of paid workers. Women's increasing entrance into the workforce in the United States and other industrialized nations over the past 50 years led researchers to ask questions about how mothers balance work and home roles, while they paid relatively less attention to how fathers balance these roles. Researchers pay even less attention to how single parents, nonbinary parents, same-sex parents, and parents in nonindustrialized cultures balance work and home roles (Chang, McDonald, & Burton, 2010). Consider what these trends in research questions suggest about the values, expectations, and assumptions of researchers.

Designing the Study and Collecting Data

After identifying a research question, the researcher determines specific methods to use for sampling participants, measuring variables, and collecting data. Poor sampling procedures can compromise the generalizability of research findings. As standard practice until the 1960s, psychologists used all-male samples to represent all people. During the second wave of the women's movement, this biased sampling method came under fire from feminist psychologists and subsequently began to decline steadily (Gannon, Luchetta, Rhodes, Pardie, & Segrist, 1992).

Another form of sampling bias occurs when researchers sample solely to make male–female comparisons and ignore other relevant demographic variables such as race, class, age, ability, sexual orientation, religion, and culture. Most psychological studies are conducted using largely White, Western, middle-class samples (Cundiff, 2012; Henrich, Heine, & Norenzayan, 2010), and gender researchers tend to study sex and gender by comparing the attributes of men as a group against those of women as a group. This practice necessarily ignores both the individual differences that exist within these two sex groups and the experiences of people who identify as neither men nor women.

Our discussion of intersectionality from Chapter 1 is relevant here. Intersectionality refers to the idea that people's experiences are shaped by multiple, interconnected identities, as well as by the power and privilege associated with these identities

(Collins, 2015). Those who adopt intersectional approaches argue that examining single identities in isolation (e.g., comparing women and men, without taking variables such as race, class, or sexual orientation into consideration) lacks meaning because it is the *intersection* of multiple identities that shapes a person (Parent, DeBlaere, & Moradi, 2013). In other words, it is not possible to understand the experiences of Black, lesbian women by simply adding up the separate experiences of being Black, being lesbian, and being female. Truly intersectional research takes power and privilege into account and goes beyond simply sampling and comparing participants across demographic categories such as race, socioeconomic status, and sexual orientation (Bowleg, 2008).

Beyond sampling, bias can enter a study through the measures and procedures used to collect data. For example, which sex do you think is more helpful? It turns out that the answer depends on the specific measures and methods used to test helpfulness. In a meta-analysis of sex differences in helping behavior, Eagly and Crowley (1986) distinguished between heroic or chivalrous helping (part of the male gender role) and nurturant or caring helping (part of the female gender role). They found an overall tendency for men to be more likely than women to help, but this was because the studies included in their meta-analysis disproportionately measured heroic helping. This suggests that researchers had a bias toward conceptualizing and measuring helping behavior in a very specific, male-typical manner. Studies that define helping in a more female-typical way reveal that women are more likely than men to donate to charities (Mesch, Brown, Moore, & Hayat, 2011) and to pursue people-oriented helping professions (Lippa, Preston, & Penner, 2014). Thus, the way that researchers measure their variables can influence both their findings and conclusions.

Interpreting and Communicating the Results

After data collection and analysis, researchers interpret and communicate their results, which creates yet another opportunity for bias. Here, androcentric thinking can shape how researchers interpret and frame their results. For example, the tendency to state conclusions in the form of the *masculine generic*—that is, using masculine pronouns and nouns (e.g., *he* and *men*) to refer to all people—used to be common but is becoming less so in psychology journals (Hegarty & Buechel, 2006). This decrease can be partly attributed to the APA publishing its "Guidelines for Nonsexist Language in APA Journals" in the 1970s (APA, 1977) and subsequently incorporating these guidelines into the third edition of its *Publication Manual* (APA, 1983). For more on the masculine generic and other gender-related language issues, see Chapter 8 ("Language, Communication, and Emotion").

As you read earlier in this chapter, taking a maximalist approach can sometimes lead people to ignore the overlap that often characterizes female and male distributions. In fact, Hyde (2005) found that 78% of the effect sizes associated with sex differences in cognitive, social, and motor variables were in the small or close-to-zero ranges, with only 8% of the effect sizes in the large or very large ranges. More recently, Zell and colleagues examined 386 effect sizes from 106 meta-analyses (including 12 million participants from over 20,000 studies) and found that 85% of the effect sizes were in the small or close-to-zero range, with a small average overall effect size of *d* = 0.21 (see Table 2.4; Zell et al., 2015).

Table 2.4 How Big Are Psychological Sex Difference Effects?
Summarizing the data from 12 million participants in over 20,000 studies of sex differences, this table shows that the vast majority of sex differences on psychological variables fall into the close-to-zero and small ranges.

Effect Sizes	Effect Size Range				
	0.00–0.10 Close-to-Zero	0.11–0.35 Small	0.36–0.65 Medium	0.66–1.00 Large	> 1.00 Very Large
Number	152	178	46	7	3
% of total	39.4%	46.1%	11.9%	1.8%	0.8%

Source: Zell, Krizan, and Teeter (2015).

STOP AND THINK

What do you think about the finding that sex differences in most psychological (cognitive and social) and motor variables are close to zero or small? Think back to the chapter debate about whether or not researchers should study sex differences. Now that you know that most sex differences are actually quite small, has your opinion changed regarding whether or not gender researchers should study sex differences? Why or why not?

SIDEBAR 2.6

Do Many Small Sex Differences Make a Big Difference?

Results of meta-analyses suggest that most psychological sex differences are small, and that women and men are more similar than different (Hyde, 2005; Zell et al., 2015). However, some researchers challenge the gender similarity conclusion, arguing that these meta-analysis results underestimate real sex differences. How so? Many psychological constructs are multidimensional, meaning that they consist of multiple different aspects. Even if there are only small sex differences on each separate aspect, it can add up to a large overall sex difference when considering the construct as a whole (Del Giudice, 2019). Consider the facial features of men and women. Sex differences on any isolated feature (e.g., eye size, mouth width) are small, making it hard to distinguish between men's and women's faces when viewing just one feature alone. However, we usually do not view facial features in isolation—and when we view whole faces, we can distinguish female from male faces with 95% accuracy (Bruce et al., 1993). Thus, although most individual sex differences are small, clusters of related sex differences might add up to large effect sizes at the level of the overarching construct. This issue continues to be a matter of debate in psychology (Del Giudice, 2013; Stewart-Williams & Thomas, 2013).

HOW DO WE ADDRESS CHALLENGES IN SEX AND GENDER RESEARCH?

Throughout this chapter, we have summarized a host of methodological challenges that sex and gender researchers face. But how do we best address these challenges? Researchers working within a **postpositivism** framework offer a set of guidelines for conducting inclusive and gender-fair research. These researchers view empirical investigation as a useful, although inherently flawed, method for acquiring knowledge (Eagly & Riger, 2014). Their views arose partly in response to a feminist critique of **scientific positivism**, which is the philosophical position that completely objective and value-free knowledge is attainable through empirical investigation. While postpositivists respect science as a process, they disagree that science is entirely objective and value free, and they seek to reduce androcentric biases in scientific research (Chrisler & McHugh, 2018).

In what follows, we summarize methodological guidelines that address some of the challenges outlined in this chapter. We offer these guidelines to raise awareness and stimulate discussion about gender bias and the lack of inclusivity in psychological research, with the ultimate goal of decreasing these biases.

Guidelines for Gender-Fair Research

To promote gender-fair research designs, psychologists offer the following guidelines:

- **Researchers should work to eliminate sex bias from sampling and avoid using men as the standard or norm** (Chrisler & McHugh, 2018). This means that researchers should not generalize findings from single-sex samples to all people and should not select samples based on biased assumptions (e.g., selecting female-only samples when studying contraception).

- **Researchers should use precise, non-gender-biased, nonevaluative terminology when collecting data and describing their participants and research findings.** Researchers should carefully consider the language used in surveys and expand the response options of sex and gender beyond "male" and "female" (and related terms like "husband" and "wife" or "brother" and "sister") to better reflect the diversity of gendered lives (Hyde et al., 2018; Schellenberg & Kaiser, 2018; Westbrook & Saperstein, 2015). In addition, researchers should not use androcentric terms and should avoid interpreting findings from within a female deficit model.

- **Researchers should not exaggerate the prevalence and magnitude of sex differences.** Journal editors and researchers should place more emphasis on publishing studies that find sex similarities, rather than privileging studies that show sex differences. To communicate the magnitude of sex differences, researchers should report patterns and effect sizes across multiple studies via meta-analysis (Hyde, 2018).

Postpositivism
An orientation that views empirical investigation as a useful method for acquiring knowledge but recognizes its inherent biases and values.

Scientific positivism
An orientation that emphasizes the scientific method and proposes that objective and value-free knowledge is attainable through empirical investigation.

- **Researchers should not imply or state that sex differences are due to biological causes when biological factors have not been properly tested.** This guideline is relevant, for example, when evaluating some of the claims made by evolutionary psychologists (whose work you will encounter in Chapter 3, "The Nature and Nurture of Sex and Gender"). David Buss (1989) studied the mate preferences of women and men across 37 cultures and found consistent evidence that men more than women prioritized attractiveness in a mate, while women more than men prioritized wealth and status in a mate. Although Buss did not measure any biological factors, he concluded that sex differences in mate preferences reflect genetically inherited tendencies.

Guidelines for More Inclusive Research

As mentioned earlier in this chapter, much of the existing psychological research on sex and gender relies on making binary, male–female comparisons, while ignoring other relevant demographic variables and power structures that shape identity. To promote more inclusive research that pushes beyond this binary, psychologists offer the following guidelines:

- **Academic psychology would benefit from more ethnic, racial, and class diversity among its professional ranks** (Eagly, 2013). While 60.7% of the U.S. population in 2017 was non-Hispanic White (U.S. Census Bureau, 2017b), 76.6% of the students who earned a PhD in psychology in 2017 were non-Hispanic White (National Science Foundation, 2017). In addition, although not specific to psychology, 76% of the full-time faculty at postsecondary institutions in the United States in 2017 were non-Hispanic White (U.S. Department of Education, 2019). This overrepresentation of non-Hispanic White people in academic psychology likely shapes the questions, methods, and interpretations of research. Increasing diversity would add different voices and perspectives, thereby increasing our understanding of people in general.

- **Academic psychologists should strive to diversify their research samples, not just within the United States but cross-culturally as well.** The majority of participants in psychology research samples are White, Western, relatively wealthy college students (Cundiff, 2012; Henrich et al., 2010). Cross-cultural meta-analyses can allow researchers to examine how sex differences in some variable of interest (e.g., math achievement) relate to gender equity measures (e.g., in educational and job opportunities for girls and women) across different nations (Else-Quest et al., 2010).

- **Researchers should routinely measure and report the demographic characteristics of their samples.** Researchers should expand the number of demographic questions asked of participants in their studies to capture a wider range of identities (Hyde et al., 2019; Sawyer, Salter, & Thoroughgood, 2013). This practice would help build databases for use in meta-analyses that explore

identities at the intersections of sex, race, sexual orientation, and class (Else-Quest & Hyde, 2016).

- **Researchers should avoid language about sex differences that implies generalizability to all people without considering the conditions under which these differences emerge and disappear** (Hyde, 2014). A sex difference found among primarily White, Western, middle-class, heterosexual young adults may not generalize to people of other races, ethnicities, ages, social classes, and so on.

- **Researchers should examine how structural inequalities and power differences associated with sex, gender identity, sexual orientation, race, class, age, ability, religion, and culture interact to shape people's experiences** (Bowleg, 2008; Warner, Settles, & Shields, 2018). Researchers should be careful to examine multiple identities holistically since overall identity is not merely a summation of different demographic characteristics. That is, the experiences of gay, Black men will not be understood well by examining sexual orientation, race, and sex separately. Furthermore, to develop more complete understandings, researchers should examine constructs that are associated with demographic characteristics such as discrimination, stress, health care access, and wages.

STOP AND THINK

What do you think of these gender-fair and more inclusive research guidelines? Do they seem reasonable or overly restrictive? What are some opposing views that researchers might raise in response to these guidelines? Do you think gender researchers should be required to follow these guidelines to get their work published? Why or why not?

Of course, not everyone agrees about which methods will lead to the most complete understanding of sex and gender, but this tension is healthy. As social psychologists, we (the authors of this book) see great value in using the scientific method to answer our research questions while simultaneously remaining aware of and questioning its flaws and imperfections. We welcome a diversity of perspectives and methods in this process. When trying to understand a phenomenon as complex as gender, having a diverse group of individuals approaching it from multiple perspectives and methods makes good sense. In general, gender researchers should continue to engage in a critical reflection about their research questions, methods, and findings by actively examining their underlying assumptions (Chrisler & McHugh, 2018; Schellenberg & Kaiser, 2018). Similarly, we hope that, as you read the studies throughout this book, you will think critically about the methods used and results reported, actively examining your own underlying assumptions as well.

CHAPTER SUMMARY

2.1 Evaluate the meaning of sex differences.

When researchers find that people of different sexes differ significantly on some variable, it means that the average difference found between women and men is unlikely to have occurred due to chance. It does not convey anything about the size, variance, or importance of the sex difference. Still, some gender researchers and consumers of research exhibit a bias in their interpretation of sex differences. Those who take a maximalist approach emphasize differences between sex groups, believing them to be qualitatively different from each other. Those who take a minimalist approach emphasize similarities between sex groups, believing them to be largely alike in their psychological characteristics. Each type of bias may have negative consequences: Maximalist interpretations tend to ignore the large amount of overlap that often characterizes people of different sexes, while minimalist approaches ignore potentially important sex differences. These perspectives are reflected in a long-standing debate in the field about whether or not sex differences should be studied.

2.2 Explain the scientific method and specific quantitative and qualitative methods used in the study of sex and gender.

In adopting the scientific method, researchers test hypotheses derived from theory by conducting studies and interpreting results. The scientific method is defined by its approach rather than the content investigated. It is a systematic, empirical way of investigating the world in order to identify rules and patterns. The researcher makes an observation, generates a hypothesis, tests the hypothesis, analyzes the results, and interprets the results to generate or refine a theory. The process is then repeated to develop theories further and to gather more data about the way the world operates.

Research methods generally fall into one of two categories: quantitative or qualitative. With quantitative methods, researchers turn variables of interest into quantities that are analyzed with statistics. Examples of quantitative methods include experiments, quasi-experiments (person-by-treatment designs and ex post facto designs), and correlational designs. Qualitative methods allow in-depth interpretations of situations, emphasizing how participants make sense of their own experiences in context. Examples of qualitative methods include case studies, interviews, and focus groups. In mixed-methods approaches, researchers use qualitative and quantitative methods within the same study or program of research to seek a more complete understanding of a research topic.

2.3 Describe meta-analyses and explain how to interpret effect sizes of different magnitudes.

Meta-analysis is a quantitative technique for analyzing a collection of results from individual studies on a given topic. It allows researchers to integrate the findings, identify context factors that shape the outcomes, and build theories. The most common effect size measure used in sex and gender meta-analyses is the d statistic, which conveys the magnitude and the direction of sex or gender differences on some variable of interest (in standardized units). A small effect size signifies a relatively large amount of overlap between different sex groups' distributions on some variable, a medium effect size signifies a moderate amount of overlap, and a large effect size signifies relatively less overlap.

For most psychological variables, effect sizes for male–female differences are small, meaning that female and male score distributions overlap substantially. Gender researchers and consumers of their research often emphasize the average differences between people of different sexes (between-sex variability) while ignoring the larger variability within different sex groups (within-sex variability).

2.4 Analyze methodological challenges and biases in sex and gender research.

No research study is free of bias or error, and gender research has its own unique set of challenges. Bias can enter at any step in the research process, from identifying the research question to interpreting and communicating results. Androcentric thinking is a kind of biased, male-centered thinking that assumes men to be the norm and representative of all people. For example, past researchers tested male-only samples and generalized the results to all people. Other common types of bias are researcher bias and participant bias, whereby researchers and participants introduce error into the research process.

2.5 Explain the principles of gender-fair and inclusive research, and describe issues of diversity in sex and gender research.

Postpositivistic gender psychologists view the scientific method as a useful but flawed method for acquiring knowledge. To decrease bias and improve the quality of research findings, they offer guidelines for conducting gender-fair research; these include eliminating sex bias in sampling and using non-gender-biased terminology in describing findings. Gender researchers should also be attentive to issues of diversity in their research and should work to include participants of all backgrounds in their studies. Finally, researchers should recognize that people's experiences are shaped by multiple interconnected identities and by the degree of power and privilege associated with these identities, a concept at the heart of intersectional research.

Test Your Knowledge: True or False?

2.1. Gender researchers disagree about whether or not it is appropriate to study sex differences. (True: There is a long-standing debate among gender researchers about whether or not it is appropriate and ethical to study sex differences.) [p. 46]

2.2. If a study finds that different sexes vary on some variable of interest (e.g., the frequency of smiling behavior), the researcher can therefore conclude that sex (being female, male, or something else) causes differences in smiling behavior. (False: Because sex and gender identity are not true independent variables—that is, they cannot be manipulated, nor can people be randomly assigned to occupy different levels of them—researchers cannot draw cause-and-effect conclusions from studies that compare people of different sexes.) [p. 55]

2.3. Qualitative methods (non-numerical methods that involve in-depth interpretations, such as case studies) are defined as nonscientific. (False: Qualitative methods are scientific. Science is a systematic, empirical way of investigating the world, and it consists of both quantitative and qualitative methods.) [p. 62]

2.4. Across most psychological variables, sex differences are generally small. (True: Sex differences on most psychological variables are in the close-to-zero and small ranges.) [p. 64]

2.5. Psychological science, if done correctly, can be truly objective and unbiased. (False: Psychological science is not truly objective because researchers always bring biases to their studies.) [p. 62]