Review in Advance first posted online on June 26, 2013. (Changes may still occur before final publication online and in print.)

Gender Similarities and Differences

Janet Shibley Hyde

Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706; email: jshyde@wisc.edu

Annu. Rev. Psychol. 2014. 65:3.1–3.26

The *Annual Review of Psychology* is online at http://psych.annualreviews.org

This article's doi: 10.1146/annurev-psych-010213-115057

Keywords

sex differences, gender similarities hypothesis, intersectionality, meta-analysis, effect size

Abstract

Whether men and women are fundamentally different or similar has been debated for more than a century. This review summarizes major theories designed to explain gender differences: evolutionary theories, cognitive social learning theory, sociocultural theory, and expectancy-value theory. The gender similarities hypothesis raises the possibility of theorizing gender similarities. Statistical methods for the analysis of gender differences and similarities are reviewed, including effect sizes, meta-analysis, taxometric analysis, and equivalence testing. Then, relying mainly on evidence from meta-analyses, gender differences are reviewed in cognitive performance (e.g., math performance), personality and social behaviors (e.g., temperament, emotions, aggression, and leadership), and psychological well-being. The evidence on gender differences in variance is summarized. The final sections explore applications of intersectionality and directions for future research.

Contents

INTRODUCTION

For centuries, humans have been fascinated with the idea of psychological gender differences, believing that these differences are both large and immutable. For example, the English clergyman Thomas Gisborne, in his helpful book, *An Enquiry Into the Duties of the Female Sex* (1797), described God's will in creating gender differences.

> The Power who called the human race into being has, with infinite wisdom, regarded, in the structure of the corporeal frame, the tasks which the different sexes were respectively destined to fulfil. . . . He has adopted with the most conspicuous wisdom, a corresponding plan of discrimination between the mental powers and dispositions of the two sexes. The science of legislation, of jurisprudence; the conduct of government in all its executive functions; the abstruse researches of erudition. . . assigned chiefly or entirely to men, demand the efforts of a mind endued with close and comprehensive reasoning in a degree in which they are not requisite for the discharge of the customary offices of female duty. . . to diffuse throughout the family circle the enlivening and endearing smile of cheerfulness, the superiority of the female mind is unrivalled. (Gisborne 1797, pp. 19–22)

Even when formal scientific psychology emerged, around 1879, the consensus remained that psychological gender differences were large (Shields 1975). However, a few researchers, such as Thorndike (1914), Hollingworth (1918), and Woolley (1914), dissented, arguing instead for gender similarities. These debates persist to the present day.

Research on gender differences and similarities is important for several reasons. First, stereotypes about psychological gender differences abound, influencing people's behavior, and it is important to evaluate whether they are accurate. Second, psychological gender differences are often invoked in important policy issues, such as single-sex schooling or explaining why, in 2005,

there were no women on the faculty in mathematics at Harvard; it is crucial to have accurate scientific information available to evaluate such policy recommendations and explanations.

Previous *Annual Review* articles have taken up the question of gender. These include Deaux's review, Sex and Gender (1985); Stewart & McDermott's, Gender in Psychology (2004); and Martin & Ruble's, Patterns of Gender Development (2010). In addition, Zahn-Waxler and colleagues' article in the *Annual Review of Clinical Psychology* (2008) addressed gender and psychological disorders in childhood and adolescence, and Hines's (2011) review in the *Annual Review of Neuroscience* evaluated gender development and the human brain. This review does not repeat material from these previous excellent reviews but instead focuses specifically on the question of gender similarities and differences, with the goal of identifying which psychological attributes show large gender differences, which show small differences, and which show no differences.

The gender similarities hypothesis states that males and females are similar on most, but not all, psychological variables (Hyde 2005). Evidence for the hypothesis came from a review of 46 meta-analyses of research on psychological gender differences that were available at the time. The 46 meta-analyses yielded 124 effect sizes (Cohen's *d* equal to the mean score for males minus the mean score for females, divided by the within-groups standard deviation) for gender differences. Strikingly, 30% of the effect sizes were between 0 and 0.10, and an additional 48% were in the range of a small difference, between 0.11 and 0.35. That is, 78% of the gender differences were small or very close to 0. The gender similarities hypothesis provides important input into the current review.

Whenever possible, conclusions in this review are based on the findings of meta-analyses, which are abundant in the field of gender differences. Meta-analyses confer substantial advantages when evaluating gender differences. First, they evaluate whether a particular gender difference is replicable; too often, individual studies reporting a gender difference capture the imagination of the press and scientists and the finding lives on, despite later studies that yield disconfirming results. Meta-analyses assess whether multiple studies find the same result. Second, meta-analyses go beyond the simple yes/no answer to whether there is a gender difference in a particular psychological attribute by estimating the magnitude of the gender difference. Third, meta-analyses can systematically explore moderators, such as social context, that may contribute to the presence or absence of gender differences.

Although this review is intended to be comprehensive, several topics were beyond the scope of it. These include gender differences in the human brain (Hines 2011, Joel 2012); gender and sex hormones; gender differences in physical health; and studies of gender bias, in which gender is studied not as a person variable but rather as a stimulus variable (Marsh et al. 2009, Swim et al. 1989).

In the sections that follow, several major theories of the origins of psychological gender differences are described first. Next, statistical methods for evaluating gender differences and similarities are summarized. The third section reviews research on gender differences and similarities in average scores on a wide array of psychological attributes (e.g., cognitive variables, personality and social psychology measures, and well-being and psychopathology). The question of gender differences in variability or variance is considered in the fourth section. Next the review addresses the concept of intersectionality and its implications for the study of gender similarities and differences. The final section includes a summary of the findings along with suggestions for future research directions.

Gender similarities hypothesis: the hypothesis that males and females are similar on most, but not all, psychological variables

Meta-analysis: a statistical technique that allows a researcher to combine the results of many research studies on a given topic

Effect size: in the study of gender differences, a measure of how large the difference is, equal to the mean for males minus the mean for females divided by the standard deviation

Intersectionality: an approach that simultaneously considers multiple categories of identity, difference, and disadvantage, such as gender, race, class, and sexual orientation

THEORIES OF THE ORIGINS OF PSYCHOLOGICAL GENDER DIFFERENCES AND SIMILARITIES

A review of all theories and hypothesized causes of psychological gender differences is beyond the scope of this article. Instead, this review focuses on three "grand theories" of differences:

evolutionary theories, cognitive social learning theory, and sociocultural theory. Following that, a more specific theory is presented, expectancy-value theory. Finally, recognizing that most theories have been framed to explain gender differences, I consider the possibility of theorizing gender similarities.

Sociocultural theory: a theory that, because of male-female differences in strength and childbearing, a division of labor by gender arose, creating psychological gender differences

STEM: science, technology, engineering, and mathematics

Evolutionary Theories

Evolutionary psychology has focused on how psychological gender differences are the product of evolutionary selection, based on an assumption that different behaviors are adaptive for males compared with females (e.g., Buss & Schmitt 1993). Two concepts are key to the argument: sexual selection and parental investment. Originally proposed by Darwin, sexual selection, which is distinct from natural selection, consists of two processes: (*a*) Members of one gender (usually males) compete among themselves to gain mating privileges with members of the other gender (usually females), and (*b*) members of the other gender (usually females) have preferences for and exercise choice in mating with certain members of the first gender (usually males). Sexual selection can be invoked, for example, to explain gender differences in aggression.

Parental investment refers to behaviors or other investments of the parent with respect to the offspring that increase the offspring's chance of survival but that also cost the parent something (Trivers 1972). Behaviors are adaptive if they help the individual produce many offspring, but then those offspring must also survive and reproduce if the individual's genes are to be successfully passed on to future generations. Gender enters the picture because human females generally have substantially greater parental investment in their offspring than do human males. Women invest a precious egg (compared with the millions of sperm that men can produce every day) and then invest nine months of gestation, which is costly to the body. At birth, then, women have greater parental investment than men do, and it is to the advantage of the person with the greater investment to care for the offspring, making sure that they survive to adulthood. Herein lies the evolutionary explanation for women's greater involvement in child care, which in turn may have enormous repercussions in other domains, such as the explanation for the dearth of women in certain science, technology, engineering, and mathematics (STEM) disciplines.

It is beyond the scope of this article to review all the evidence for and against evolutionary theories. Briefly, the evidence appears to be mixed (Pedersen et al. 2011). As one example, evolutionary psychologists Buss & Schmitt (1993) predicted large gender differences in the desired number of sexual partners and reported that men, on average, over the next 30 years, desire 16 partners, compared with women's desire for 4. This finding is consistent with evolutionary theories insofar as men are said to increase their fitness by having sex with numerous women. Another team, however, collected similar data and reached very different conclusions (Pedersen et al. 2002). Although they found men desiring more partners than women did, the distributions were highly skewed, with a few individuals wanting hundreds of partners, so the mean is not an appropriate statistic, nor are significance tests requiring an assumption of normal distributions appropriate. Taking this into account, the researchers focused on the median, which for both men and women was one partner. Overall, then, their results indicated gender similarities.

Cognitive Social Learning Theory

Cognitive social learning theory is another approach with broad utility in understanding psychological gender differences. As formulated by Bussey & Bandura (1999), the theory holds that both children's and adults' behavior is shaped by reinforcements and punishments. In addition, people imitate or model others in their environment, particularly if the others are powerful or

admirable. Abundant evidence exists for the processes specified by social learning theory (Bussey & Bandura 1999).

In the more recent versions of the theory, cognitive components have been added, such as attention, self-regulation, and self-efficacy. For example, as children grow, control of their behavior shifts from externally imposed reinforcements and punishments to internalized standards and self-regulation. In particular, children internalize gender norms and conform their behavior to those norms. Self-efficacy, another cognitive component, refers to a person's belief in her or his ability to accomplish a particular task. Self-efficacy may be important in explaining certain gender effects. For example, although girls' math performance is equal to that of boys, generally there is a wider gender gap in math self-efficacy ($d = 0.33$, Else-Quest et al. 2010). Self-efficacy is important because of its power in shaping people's decisions about whether to take on a challenging task, such as majoring in mathematics.

Sociocultural Theory

Sociocultural theory, also called social role theory or social structural theory, was proposed by Eagly & Wood (1999; Wood & Eagly 2012) as an alternative to evolutionary theorizing about gender differences. The essential argument of the theory is that a society's division of labor by gender drives all other psychological gender differences. Psychological gender differences result from individuals' accommodations or adaptations to the particular restrictions on or opportunities for their gender in their society. The theory acknowledges biological differences between men and women, such as differences in size and strength and women's capacity to bear and nurse children. These differences historically contributed to the division of labor by gender. Men's greater size and strength led them to pursue activities such as warfare, which gave them greater status and wealth, as well as power over women. Once men were in these roles of greater power, their behavior became more dominant, and women's behavior became more subordinate. Women's assignment to the role of child care led them to develop qualities such as nurturance and a facility with relationships.

With a few exceptions (e.g., Schmitt 2005), evolutionary psychologists emphasize cross-cultural universals in patterns of gender differences, resulting from evolution many centuries ago. In contrast, sociocultural theory focuses on variations across cultures in patterns of gender differences. Eagly & Wood (1999) reanalyzed Buss's cross-national data on gender differences in mate preferences. Their hypothesis was that the greater the gender difference in power and status (gender inequality) in a culture, the greater would be the gender differences in mate preferences. Using a United Nations database that indexes gender inequality for participating nations, Eagly & Wood (1999) found strong correlations between gender inequality and the magnitude of gender differences in mate preferences (see also Zentner & Mitura 2012). That is, the nations with the largest gender gaps in power also have the largest gender gaps in mate preferences. These findings are consistent with sociocultural theory and are inconsistent with evolutionary theory.

Other researchers have tested the hypothesis that psychological gender differences should be smaller in nations with more gender equality and larger in nations with more inequality. For example, in a cross-national meta-analysis of gender differences in math performance, the researchers used measures of nations' gender equality to predict the gender gap in math performance (Else-Quest et al. 2010). Some indices of gender equality, such as the Gender Empowerment Measure, are global or are composites of multiple indicators, such as women's share of parliamentary seats and the wage gap. Other indices of gender equality are domain-specific, such as women's share of research positions. The domain-specific indicators were most successful at predicting cross-national variations. For example, women's share of research positions in nations significantly predicted smaller gender gaps in math performance as well as smaller gaps in math self-concept and math self-efficacy.

Sociocultural theory is more recent than the other theories reviewed here. Overall, though, evidence is accumulating in support of it.

Expectancy-Value Theory

Eccles's expectancy-value theory (Eccles et al. 1994, Meece et al. 1982), although not proposed as a grand theory of psychological gender differences, nonetheless has broad utility in accounting for multiple phenomena discussed in this review. According to the theory, two categories of factors contribute to a person's decision to take on a challenging task, such as enrolling in calculus in high school: expectancies (expectations for success at the task) and task values (e.g., interest in the task, usefulness of the task for current or future goals). In turn, many factors influence expectancies and values. For example, an adolescent's expectancy of success in a calculus course may be influenced by her self-concept of math ability, past achievement experiences (grades in past math courses, scores on standardized tests), socializers' beliefs and behaviors (e.g., parents' belief that engineering would be a good career for their daughter), and the broader sociocultural milieu (the gender distribution that she observes in her intended occupation). The model has received abundant empirical support (e.g., Eccles et al. 1994).

Expectancy-value theory is especially helpful in identifying the wide array of factors that contribute to the underrepresentation of women in STEM careers. Moreover, it identifies points at which interventions might be effective. It may also be applicable to understanding phenomena such as gender differences in self-esteem and domain-specific self-concept.

Theories of Gender Similarities

The theories reviewed to this point were designed to explain gender differences. Little or no attention has been devoted to theorizing gender similarities. Several possibilities exist.

In the realm of evolutionary forces, theorists have focused on sexual selection and its capacity to create gender differences. Missing from their analysis is a consideration of natural selection, which should act equally on males and females and therefore should create gender similarities (Hyde 2006). Even at the chromosomal level, humans have 23 pairs of chromosomes, only one of which, the sex chromosomes, differ between males and females. The remaining 22 pairs of autosomes are found in both males and females and pass from fathers to daughters, and from mothers to sons, across generations. The evolutionary psychologists' focus on sexual selection should not obscure the greater force of natural selection, which creates gender similarities.

Cognitive social learning theory, sociocultural theory, and expectancy-value theory, although designed to explain gender differences, can easily be adapted to understanding gender similarities, as well as trends over time toward greater gender similarities in the United States. Using cognitive social learning theory, for example, one might speculate that punishments for gender-role violations have declined over time, allowing boys and girls to behave more similarly to each other. As of this writing, it appears that no direct evidence exists that punishment for gender-role violations has declined over time, but it would be an interesting hypothesis to test. Cognitive social learning theory might also account for the influx of women into the biological sciences and medicine since 1970 as resulting in part from the increasing number of female role models; the entry of more women into a field in turn promotes the entry of more women. This same prediction arises from expectancy-value theory.

Sociocultural theory, with its focus on power inequalities and the division of labor by gender, explains gender differences but can also contribute to an understanding of gender similarities. Gender similarities are expected in nations in which there is gender equality. Moreover, as the

division of labor by gender narrows over time, psychological gender similarities should arise. As an example, with the emergence of the women's movement and increased labor force participation of women beginning in the 1970s, family roles have shifted toward equality as least to some extent. A 2001 study found that fathers with an employed wife spent on average 23 hours per week with their children, compared with only 17 hours per week in 1981 (Sandberg & Hofferth 2001). Sociocultural theory predicts that, as men increasingly occupy the child-care role, they will develop more nurturant qualities, reducing the gender gap in nurturance.

Overall, then, an important future research strategy will be to theorize gender similarities. Existing theories may be adapted to this task, yet novel theories may also be needed.

STATISTICAL METHODS FOR THE ANALYSIS OF GENDER SIMILARITIES AND DIFFERENCES

Cohen's *d* (Cohen 1988), equal to the mean for males minus the mean for females, divided by the pooled within-groups standard deviation, is the most common measure of the magnitude of gender differences in psychological research. Cohen's guidelines for interpreting effect sizes are that $d = 0.20$ is a small difference, 0.50 is moderate, and 0.80 is large. Moreover, Hyde (2005) interprets a *d* value ≤ 0.10 as being trivial, as being a gender similarity. Other statistics are available, though, and an exploration of them will be helpful in interpreting the magnitude of gender differences.

An important equivalence formula is $d = 2r/\sqrt{(1-r^2)}$. A quick approximation is $d = 2r$. Therefore a *d* of 0.20, which Cohen interpreted as a small difference, is equivalent to $r = 0.10$. That is, if we computed the correlation between gender (coded 0, 1) and an outcome such as mathematics performance, an *r* value of 0.10 would yield a *d* value of 0.20. Similarly, the *d* value of 0.10, interpreted here as trivial, is equivalent to $r = 0.05$, which most scientists would agree is a trivial relationship.

Cohen (1988) provided another way to understand equivalents of *d*, in terms of the percent overlap of two distributions at various values of *d*, assuming that both distributions are normal and have the same variance. For example, for $d = 0.10$, there is 92.3% overlap between the two distributions, and for $d = 0.20$, there is 85.3% overlap. At a very large value of *d*, 1.0, there is still 44.6% overlap between the distributions. This approach underscores the point that, with small or moderate values of *d*, within-gender variance is much larger than between-gender variance.

Another equivalence provided by Cohen (1988) assesses the percentage of Group B that will equal or exceed the median for Group A at various values of *d*, again assuming normal distributions with equal variance. For example, for $d = 0.10$, 52% of Group B (e.g., males) would score above the median for Group A (e.g., females). For $d = 0.20$, 54% of Group B scores above the median for Group A. These equivalencies again support the reasonableness of interpreting $d = 0.20$ as a small effect. For a very large difference, $d = 1.0$, 69% of Group B scores above the median for Group A.

Del Giudice and colleagues (2012) introduced the use of Mahalanobis *D* to the measurement of the magnitude of gender differences in personality and concluded that the gender similarities hypothesis is incorrect and that there are very large gender differences. A staple of multivariate statistics for decades, *D* in this application measures the distance between two centroids in multivariate space. It is a multivariate generalization of the *d* statistic. What was not made clear in the Del Giudice paper is that *D* is computed by taking the linear combination of the original variables (scores on emotional stability, dominance, vigilance, and so on) that maximizes the difference between groups. What they showed, then, is that, if one takes a large enough set of psychological measures and then takes a linear combination to maximize differences, one can get a big gender difference. Moreover, this difference or distance is along a dimension in multivariate space that is

a linear combination of the original variables, but this dimension is uninterpretable. What does it mean to say that there are large differences in personality, lumping together distinct aspects such as emotional stability, dominance, and vigilance? Certainly contemporary personality theorists do not argue that there is a single dimension to personality. Overall, then, this application of Mahalanobis *D* produces results that are biased toward finding a large difference because of taking a linear combination that maximizes group differences, and it appears to yield results that are uninterpretable.

Equivalence testing: a statistical method to test whether two groups (e.g., males and females) are significantly similar to each other, i.e., whether the two means are within a certain distance from each other

Carothers & Reis (2013) introduced another statistical application for the analysis of gender similarities and differences, taxometric methods. These methods allow the researcher to determine whether the latent structure of a construct is dimensional (continuous) or taxonic (categorical). This approach analyzes whether males' and females' psychological attributes differ in categorical ways or dimensional ways. Tapping multiple data sets, the researchers selected psychological variables that typically are quite gender differentiated, such as masculinity/femininity, relational-interdependent self-construal, and science inclination. With only a few exceptions, the constructs showed gender to be dimensional, not taxonic, contrary to stereotypes that hold that males and females are categorically different. These taxometric methods hold great promise for the future.

Ball and colleagues (2013) introduced equivalence testing as another method for analyzing gender similarities and differences. The problem with identifying gender similarities when using traditional significance testing methods is that one is trying to prove the null hypothesis. Equivalence testing reverses the approach, so that a significant result can indicate significant similarities, e.g., whether the male and female means are within 0.20 standard deviations of each other or some other value.

In summary, the effect size *d* (or Hedges' *g*) will doubtless continue to be a mainstay in evaluating the magnitude of gender differences. Newer methods are becoming available, including taxometric methods and equivalence testing, and they offer additional leverage in gender analyses.

AVERAGE GENDER SIMILARITIES AND DIFFERENCES

Most gender differences research has examined whether average gender differences exist over a wide array of psychological characteristics. The sections that follow review research on gender differences in cognitive variables (mathematical, spatial, and verbal performance), personality and social psychological variables, and well-being and psychopathology. The final section considers the STEM question: Why are women underrepresented in certain careers in science, technology, engineering, and mathematics?

Cognitive Gender Similarities and Differences

In their important review, Maccoby & Jacklin (1974) concluded that gender differences are reliably found in mathematical, spatial, and verbal abilities, with males outperforming females on mathematical and spatial tests and females outperforming males on verbal tests. This view of cognitive gender differences continues to dominate the field to the present day. Here the evidence on each of these gender differences is reviewed, focusing on the magnitude of the differences.

Mathematical performance. A 1990 meta-analysis reviewed 100 relevant studies representing the testing of more than 3 million persons (Hyde et al. 1990a). Overall, the gender difference was trivial in magnitude, $d = -0.05$. However, moderator analyses examining age and cognitive level of the test (from lowest to highest: computation, understanding of math concepts, and complex problem solving) revealed that more substantial differences favoring males emerged in complex

problem solving in high school ($d = 0.29$). This gender gap is of concern because complex problem solving is precisely the mathematical skill needed to enter STEM occupations.

Spurred by then-Harvard President Lawrence Summers's 2005 comments that women lack the mathematical ability to succeed at the highest levels of mathematics and science, Hyde and colleagues conducted additional meta-analyses based on contemporary data. These more recent data indicate that, in general, females have reached parity with males in math performance. One meta-analysis synthesized data from state assessments of US children's math skills (Hyde et al. 2008a). Data represented the testing of more than 7 million students in grades 2 through 11. Across grades, effect sizes ranged between −0.02 and +0.06, i.e., there was no gender difference at any grade level. Because state assessments generally tap only lower-level math skills, the authors turned to data from the National Assessment of Educational Progress, which does include items assessing complex problem solving. For these items, at grade 12, the average effect size was $d = 0.07$, indicating that girls had reached parity with boys even for complex problem solving at the high school level.

In a second meta-analysis based on contemporary studies, Lindberg and colleagues synthesized data from 242 studies appearing between 1990 and 2007, representing the testing of 1.2 million people (Lindberg et al. 2010). Overall, there was no gender difference ($d = 0.05$). The gender difference in complex problem solving was small ($d = 0.16$).

In a third meta-analysis, Else-Quest and colleagues (2010) examined international data on gender differences in mathematics performance, relating it to measures of gender inequality across nations (see also Nosek et al. 2009). Briefly, the findings indicate great variability across nations in the magnitude and even the direction of the gender difference, as noted in the discussion of sociocultural theory.

Overall, then, it appears that girls have reached parity with boys in mathematics performance, at least in the United States. The gender difference in complex problem solving in high school is smaller than it was in the 1990 meta-analysis and has even disappeared in one analysis of National Assessment of Educational Progress data (Hyde et al. 2008a).

Spatial performance. Spatial abilities are assessed in a variety of ways. This review focuses on a particular spatial skill, three-dimensional (3D) mental rotation, which requires the test taker to mentally rotate an object in three dimensions to determine whether it matches one of several other objects. Although many occupations do not require mental rotation skills, several do, including engineering and architecture, areas in which women are underrepresented.

According to an early meta-analysis, the gender difference in 3D mental rotation is large, favoring males, $d = 0.73$ (Linn & Petersen 1985). In a later meta-analysis, the gender difference was moderate in magnitude, $d = 0.56$ (Voyer et al. 1995). These overall effect sizes, however, mask some complexities. Gender researchers have suspected for some time that, for mathematical and spatial performance, tightly timed tests—which measure speed as much as skill—are advantageous to males, whereas untimed tests or tests with ample time provide more opportunity for females to display their skills. One meta-analysis found that, indeed, with short time limits, the gender difference in mental rotation was large ($d = 1.03$), whereas in tests with no time limits the effect size was only moderate ($d = 0.51$) (Voyer 2011; see also Maeda & Yoon 2013).

There is some evidence from individual studies that a male advantage in mental rotation emerges as early as infancy (Moore & Johnson 2008, Quinn & Liben 2008). However, it is a bit too early to tell exactly what these findings mean.

Others note that a spatial training curriculum is notably absent from the schools, so that the gender difference may result from informal, extracurricular learning based on the different experiences of boys and girls in areas such as sports and video game playing. It is clear that scores

on mental rotation tests can be improved with training (Newcombe et al. 2002, Uttal et al. 2013). According to a meta-analysis, the average effect size for the difference between training and control groups was $d = 0.44$ (Uttal et al. 2013). In one experiment, college students were given 10 hours of training on an action video game, Medal of Honor: Pacific Assault; controls played a puzzle game (Feng 2007). Both women and men in the experimental group improved their performance on a mental rotation test. The women improved more than the men, and experimental group women performed as well as control group men. On average, US boys spend 1 hour 37 minutes per day playing video games compared with 49 minutes for girls (Rideout et al. 2010). That is, boys practice with video games about twice as much as girls do, representing a major gender gap in experience relevant to 3D spatial skills.

On the question of the relevance of spatial skills for the underrepresentation of women in engineering, faculty at one college of engineering created multimedia software that provides training in mental rotation (Gerson et al. 2001). Entering students majoring in engineering take the mental rotation training. This training has improved the retention of women in the engineering major from 47% to 77%.

Overall, then, there is a moderate gender difference favoring males in 3D mental rotation. However, this gender difference may result from the absence of spatial training in the schools combined with major gender gaps in relevant out-of-school experiences.

Verbal skills. Contrary to stereotypes that females have better verbal skills than males do, gender differences in verbal skills overall favor females but are quite small, $d = -0.11$ (Hyde & Linn 1988). This overall value, however, masks variations for different types of verbal ability. No gender differences were found for vocabulary ($d = -0.02$), reading comprehension ($d = -0.03$), or essay writing ($d = -0.09$). The largest effect was for verbal fluency, $d = -0.33$.

In an analysis based on well-sampled studies of US adolescents, Hedges & Nowell (1995) computed effect sizes for gender differences in reading comprehension and vocabulary. For reading comprehension, effect sizes ranged from −0.18 (National Longitudinal Survey of Youth data) to +0.002 (High School & Beyond data). For vocabulary, effect sizes ranged between −0.06 (National Longitudinal Study of the High School Class of 1972) and +0.25 (Project Talent). Again, the differences were not large and did not consistently favor females.

In a recent study, international data were analyzed from the Program for International Student Assessment data set, which samples 15-year-olds in participating nations (Reilly 2012). For reading achievement in the United States, $d = -0.26$, a small average difference favoring girls. However, at the lowest level of performance, boys outnumbered girls by a ratio of 4.5 to 1, leading to the question of whether boys with learning disabilities account for the female advantage. Across all 65 participating nations, the gender difference was larger, $d = -0.44$. In an effort to account for variability in the magnitude of the gender difference across nations, effect sizes were correlated with measures of gender equity across nations and yielded significant effects, a finding that is consistent with sociocultural theory.

Taking these three studies together, it is difficult to reconcile the gender gap found in the Program for International Student Assessment data ($d = -0.26$) with the gaps that are close to zero in the other two studies. If there is a female advantage in reading comprehension and other verbal skills, it is a small one.

Attitudes. Attitudes about cognitive abilities and gender differences in abilities are often as important as actual performance in predicting educational and career outcomes. Attitudes toward mathematics have been studied extensively. According to a meta-analysis, for 15-year-olds in the United States, $d = 0.27$ for the gender difference in math self-confidence and −0.23 for anxiety

(Else-Quest et al. 2010). That is, the gender differences in math self-confidence and math anxiety are larger than the gender difference in actual performance.

Personality and Social Behaviors

This section considers gender differences in a wide array of personality measures and social behaviors. Personality domains reviewed here include temperament, personality, impulsivity, emotion, and interests. Social behaviors include aggression, helping behavior, communication, sexuality, and leadership.

Temperament: biologically based emotional and behavioral consistencies that appear early in life and predict outcomes in domains such as psychopathology and personality

Temperament. A meta-analysis of research on gender differences in temperament included data from children between the ages of 3 months and 13 years (Else-Quest et al. 2006; see also Else-Quest 2012). Temperament can be thought of as biologically based emotional and behavioral consistencies that appear early in life and predict outcomes in domains such as psychopathology and personality. In a sense, temperament represents the earliest indication of the individual's personality. Multiple theoretical models are used in temperament research, leading to different measures, but one can think of three basic temperament factors: effortful control (including attention, effortful control, and persistence), negative affectivity (emotionality, distress to limits, fear) and surgency (activity, impulsivity).

For the effortful control factor, girls scored higher on inhibitory control ($d = -0.41$) and attention ($d = -0.23$) (Else-Quest et al. 2006). These average gender differences are in the small-to-moderate range and contrast to larger gender differences at the tail of the distribution, where boys with attention deficit–hyperactivity disorder (ADHD) outnumber girls by ratios of 2:1 to 9:1 across studies (Rucklidge 2010). This distinction between modest average gender differences and marked gender differences at the tail of the distribution illustrates a common problem in which people infer, from marked gender differences in extreme scores, that differences between girls and boys are categorical—i.e., that all boys are more active than girls—when in fact average gender differences are moderate, representing substantial overlap between the male and female distributions.

For the negative affectivity factor, there was no gender difference for negative affectivity ($d = -0.06$), sadness (-0.10), or emotionality (0.01). These findings are perhaps surprising given the power of negative affectivity to predict later depression (Watson & Clark 1984) and the reliable gender difference in depression, discussed in a later section.

In regard to the surgency factor, gender differences in activity ranged between $d = 0.15$ and 0.33, depending on the measure. For impulsivity, $d = 0.18$.

Personality and the five factors. Feingold (1994) conducted a meta-analysis of gender differences in personality based on US test norming data rooted in the Big Five model of personality (McCrae & Costa 2013). For the neuroticism factor, $d = -0.27$ for anxiety. For the extraversion factor, $d = -0.14$ for gregariousness but $+0.49$ for assertiveness. Openness to experience had an effect size of 0.13, and conscientiousness had -0.07. The largest gender difference was for tender-mindedness, which is part of the agreeableness factor, $d = -1.07$.

Although they did not use formal meta-analytic methods, Costa and colleagues synthesized data from the NEO Personality Inventory for more than 23,000 individuals across 26 cultures (Costa et al. 2001). To facilitate comparison between the Costa findings, which are more recent, and the Feingold (1994) findings, **Table 1** shows effect sizes for comparable scales representing each of the five factors. Costa and colleagues provided data for 26 cultures, but because of space limitations, only data for Americans, Japanese, and black South Africans are shown. **Table 1** shows that the

Table 1 Effect sizes for gender differences in personality across two meta-analyses, based on the Big Five model

Personality factor	Feingold (1994)	Costa et al. (2001)US adults	Costa et al. (2001)Japanese	Costa et al. (2001)Black South Africans
Neuroticism, Anxiety	−0.27	−0.40	−0.09	−0.08
Extraversion, Gregariousness	−0.14	−0.21		
Openness to Ideas	0.13	0.32		
Agreeableness, Tender-mindedness	−1.07	−0.31	−0.39	−0.05
Conscientiousness, Order	−0.07	−0.05		

Costa data for US adults generally agree with the Feingold results for the direction of gender differences, if not the magnitude. However, strikingly, the robust finding of greater neuroticism among women was not found for Japanese or black South Africans, providing evidence of the importance of culture in shaping gender differences in personality.

It should also be noted that the gender differences found by Feingold and by Costa and coworkers for the US sample are highly consistent with gender stereotypes in the United States, where women traditionally have been expected to be more neurotic and tender-minded than men (e.g., Broverman et al. 1972). Yet gender stereotypes and roles can vary across cultures. Costa and colleagues' analysis indicated that gender differences in self-reported personality traits are largest in prosperous cultures where women have many educational opportunities. Neuroticism might seem like a luxury for a black South African woman, erasing a gender difference that is found in the prosperous United States.

Impulsivity. Another meta-analysis examined gender differences in impulsivity, distinguishing among reward hypersensitivity, punishment hyposensitivity, and inadequate effortful control (Cross et al. 2011). Women were more sensitive to punishment ($d = -0.33$), but there was no gender difference in reward sensitivity. Men showed greater sensation seeking on questionnaire measures ($d = 0.41$) and behavioral measures (0.36). Contrary to expectations, there was no gender difference on measures of deficits in effortful control. The impulsivity meta-analysis, however, examined data only from samples of persons aged 11 and older, in contrast to the temperament meta-analysis, which focused on younger children.

Emotions. Emotional experience and expression, too, are the object of gender stereotypes in which anger is more acceptable for men, and most other emotions (e.g., sadness, fear, happiness) are considered more acceptable for women (e.g., Durik et al. 2006). Are these stereotypes consistent with actual behavior?

One meta-analysis examined gender differences in emotion expression in children from birth to adolescence (Chaplin & Aldao 2013). The results indicated trivial gender differences for positive emotions ($d = -0.08$), internalizing emotions such as sadness and anxiety ($d = -0.10$), and externalizing emotions such as anger ($d = 0.09$). However, the magnitude of the gender differences varied as a function of age. For example, the gender gap in positive emotions grew larger with age ($d = -0.20$ in middle childhood and -0.28 in adolescence), perhaps reflecting increasing socialization pressure with age. The magnitude of gender differences also varied with context. For example, when children are alone, $d = -0.03$ for internalizing emotions, but $d = -0.16$ when

children are with adults. This meta-analysis again confirms the importance of context in creating or erasing gender differences.

Else-Quest and colleagues (2012) conducted a meta-analysis of gender differences in the self-conscious emotions, which are gender stereotyped in that women are expected to experience more guilt, shame, and embarrassment, whereas men are expected to experience more pride. The results indicated small differences favoring females for guilt ($d = -0.27$) and shame ($d = -0.29$) and, contrary to stereotypes, trivial differences for embarrassment ($d = -0.08$), authentic pride ($d = -0.01$), and hubristic pride ($d = -0.09$),

Overall, then, although stereotypes portray women as the emotional ones and hold that there are large gender differences in emotions such as fear and pride, the data, from both children and adults, indicate that gender differences in emotional experience are small or, in many cases, trivial.

Interests. Su and colleagues (2009) meta-analyzed data on gender differences from 47 interest inventories. The authors summarized their global findings as "Men and Things, Women and People." That is, on the Things–People dimension, the gender difference was large ($d = 0.93$), with women being more interested in people and men more interested in things. Regarding STEM, large gender differences were found for interest in engineering ($d = 1.11$), with more moderate gender differences for interest in science (0.36) and mathematics (0.34).

These findings regarding interests may help to explain the underrepresentation of women in certain STEM fields such as engineering. However, gender differences in interests are not hardwired or immutable. They, too, are shaped by sociocultural factors, a point that is discussed in more detail in the section on STEM.

Aggression. Several meta-analyses are available on gender differences in aggression (Archer 2004, Eagly & Steffen 1986, Hyde 1984). The focus here is on the Archer review because it covers the most recent data.

Gender differences in aggression do appear consistently and are generally moderate in magnitude. For physical aggression, $d = 0.55$ (Archer 2004). Developmentally, this difference appears about as early as children begin playing with one another, around the age of 2 (Alink et al. 2006).

A stereotype about "mean girls" has emerged, implying that, although boys may engage in more physical aggression, girls engage in much more verbal and relational aggression. "Relational aggression" refers to behavior that is intended to hurt others by damaging their peer relationships (Crick & Grotpeter 1995) and is sometimes also called indirect aggression. Contrary to the stereotype, the gender difference is not large, though. According to the Archer meta-analysis, $d = -0.19$ for peer ratings and -0.13 for teacher reports. In both cases, girls score higher, but the difference is small. Boys are quite capable of relational aggression.

It is important to recognize that patterns of gender differences in aggressive behavior are highly context dependent. In one striking experiment, researchers used the technique of deindividuation to produce a situation that removed the influences of gender roles (Lightdale & Prentice 1994). Deindividuation refers to a state in which the person has lost his or her individual identity; that is, the person has become anonymous. Under such conditions, people feel no obligation to conform to social norms such as gender roles. Half the participants were assigned to an individuated condition by having them sit close to the experimenter, identify themselves by name, wear large name tags, and answer personal questions. Those assigned to the deindividuated condition sat far from the experimenter and were simply told to wait quietly. Next, the participants played a video game in which they first defended and then attacked by dropping bombs. The number of bombs dropped was the measure of aggressive behavior. The results indicated that, in the individuated condition, men dropped significantly more bombs than women did, consistent with the gender

differences found in meta-analyses. In the deindividuated condition—that is, in the absence of gender roles—there were no significant gender differences, and, in fact, females dropped somewhat more bombs than males did. In short, a significant gender difference in aggression disappeared when the influence of gender roles was removed.

Findings such as these have led some researchers to conclude that the emphasis on gender differences in aggression research is misplaced and that, instead, context plays a far greater role (Richardson & Hammock 2007). Consistent with this view, a meta-analysis by Bettencourt & Kernahan (1997) of experimental studies indicated that, in the context of violent cues but no provocation, a moderate gender difference is found ($d = 0.41$), whereas when violent cues and provocation are both present, the gender difference is eliminated.

Communication. As with other areas, stereotypes about enormous gender differences in communication abound, exemplified by Deborah Tannen's best-selling book, *You Just Don't Understand: Women and Men in Conversation* (1991). Tannen has argued that women's and men's patterns of speaking are so vastly different that men and women essentially belong to different linguistic communities. In particular, it has been said that women use more tentative speech than men do, as indicated by greater use of tag questions (That was a great movie, wasn't it?) and hedges (I'm kind of interested in biology), none of which convey self-confidence and strength.

According to a meta-analysis of studies of tentative speech, however, the gender differences favor women but are small (Leaper & Robnett 2011). For tag questions, $d = -0.23$, and for hedges, $d = -0.15$. Moreover, the magnitude of the gender difference depends on context. It is larger in lab studies ($d = -0.28$) than in naturalistic studies outside the lab ($d = -0.09$).

Helping behavior. Eagly & Crowley (1986; see also Eagly 2009) conducted a meta-analysis of research on gender differences in helping behavior, yielding $d = 0.34$, indicating that males help more. This overall effect, however, masks wide variations in the magnitude and even direction of the gender difference. Moderator analyses, for example, indicated that the gender difference favoring males was especially pronounced when the situation might involve danger, such as stopping to help a motorist with a flat tire. The gender difference was also larger when the helping behavior was observed by others rather than when the person was alone; heroism is part of the male role, but it is hard to be a hero when no one is watching. Other kinds of helping, such as helping a distressed child, are associated with the female role, and here we see less helping by males and more by females.

Sexuality. Sexuality is another area in which stereotypes hold that women and men are vastly different. A meta-analysis by Petersen & Hyde (2010) paints a complex picture. Across 14 distinct sexual behaviors (e.g., number of partners, casual sex, extramarital sex) and 16 sexual attitudes (e.g., attitudes about extramarital sex, attitudes about homosexuality), no gender differences were large. Four were in the moderate range: males were more likely to masturbate ($d = 0.53$), to use pornography ($d = 0.63$), and to have more sexual partners ($d = 0.36$), and males had more favorable attitudes toward casual sex ($d = 0.45$). Many gender similarities were found as well, with effect sizes ≤ 0.10 for oral sex, attitudes about extramarital sex, attitudes about masturbation, attitudes about condom use, and attitudes about lesbians (but attitudes toward gay men showed somewhat more favorable attitudes among women, $d = -0.18$). Moderator analyses indicated that some gender gaps have been narrowing over time, including attitudes about casual sex.

Unlike many of the other domains reviewed here, for which both self-report and behavioral measures are typically available, most sex research is based on self-report. That leads to the question of whether some gender differences, such as the one in number of sexual partners,

are actual behavioral differences or merely represent different reporting biases among women and men. Consistent with the double standard (Crawford & Popp 2003), might men exaggerate their number of partners, or might women underreport theirs?

Several experiments by Fisher and colleagues have addressed these questions (Alexander & Fisher 2003, Fisher 2007, Jonason & Fisher 2009). Using the bogus pipeline technique, researchers randomly assigned college students to one of three experimental conditions (Alexander & Fisher 2003). In the bogus pipeline condition, participants were hooked up to a fake polygraph while they were completing the sex questionnaire, and told that the machine could detect untruthful answers, presumably eliciting very honest reporting. In the anonymous condition, the student filled out the sex questionnaire anonymously, as is typical of much sex research, and placed the questionnaire in a locked box when finished. In the exposure threat condition, respondents had to hand their completed questionnaire to the experimenter, who was an undergraduate peer, and the experimenter sat in full view while the respondents completed their questionnaire. For reports of the number of sex partners, the gender difference was largest in the exposure threat condition, males reporting more partners. The gender difference was also present in the anonymous condition; however, the gender difference was erased in the bogus pipeline condition, presumably when participants were giving the most honest answers. This research suggests that some gender differences in sexuality do not represent actual differences in behavior but instead are an artifact of contextual factors that cause reporting biases consistent with gender roles.

Leadership. Eagly and colleagues conducted a series of meta-analyses on gender and leadership. One meta-analysis examined data on gender and the effectiveness of leaders (Eagly et al. 1995). Overall, there was no gender difference in leadership effectiveness, $d = -0.02$, which was true whether objective or subjective measures were used. However, male leaders were somewhat more effective in positions that were consistent with the male role, and female leaders were more effective in positions consistent with the female role. Overall, then, there were no gender differences in leadership effectiveness, but women may be more effective than men in certain situations.

A separate question is whether women and men differ in their leadership styles. The styles that have been studied most are transformational (innovative leadership in which the leader serves as a positive role model based on gaining the trust of the followers), transactional (leadership by administering rewards for good behaviors and punishments or corrections for poor performance), and laissez-faire (the leader is neglectful and uninvolved). Transformational style is a concept that has emerged relatively recently and is seen as admirable. For transformational leadership, $d = -0.10$, i.e., a trivial difference (Eagly et al. 2003). For transactional leadership, women have a slight edge in reward-based approaches, $d = -0.13$, whereas men are more inclined to wait until problems crop up and then address them, $d = 0.27$. Men are also somewhat more likely to engage in laissez-faire leadership, $d = 0.16$.

Overall, then, although leadership is stereotyped as masculine, there is no evidence that women are incompetent as leaders, whether measured by effectiveness or by the use of optimal styles such as transformational leadership.

Well-Being and Psychopathology

Some psychological disorders show lopsided gender ratios (Zahn-Waxler et al. 2008). For example, twice as many women as men are depressed (Kessler et al. 1993). Roughly 90% of anorexics are females (Subst. Abuse Ment. Health Serv. Adm. 2009), and, depending on the study, boys with ADHD outnumber girls by 2:1 to 9:1 (Rucklidge 2010). The focus here is particularly on gender differences in depression, rumination, and self-esteem, because they have been studied extensively.

Depression. In adulthood, twice as many women as men are depressed (Kessler et al. 1993). Although girls are no more depressed than boys in childhood, more girls than boys are depressed by ages 13 to 15 (Hankin et al. 1998, Kessler et al. 1993). Although no comprehensive meta-analysis is available, two more specific meta-analyses have been conducted. Twenge & Nolen-Hoeksema (2002) meta-analyzed studies that had used the Children's Depression Inventory (CDI) and found that, between the ages of 8 and 12, $d = 0.04$, whereas between ages 13 and 16, $d = -0.16$. This gender difference in adolescence is not large, which appears to contrast with statistics on the 2:1 ratio of depressed females to depressed males. A resolution probably lies in the fact that the effect sizes are based on mean differences in community samples, whereas the statistics on depression examine extreme scores at the tail of the distribution. Moreover, the CDI measures symptoms of depression, whereas most studies finding the 2:1 ratio assessed diagnoses of depression. In terms of developmental trends, boys' CDI scores held relatively constant from younger to older ages, whereas girls' scores increased, so that $d = -0.22$ by age 14.

The other available meta-analysis explored gender differences in depression at the other end of the life span, among those 75 and older (Luppa et al. 2012). Gender ratios for prevalence rates ranged between 1.4 and 2.2, suggesting that the preponderance of women with depression is a phenomenon that continues into old age.

Hundreds of studies have attempted to identify the factor or factors that account for the gender difference in depression and why it emerges in adolescence. Synthesizing these studies and previous theoretical models, Hyde and colleagues (2008b) proposed the ABC model of gender differences in depression. According to the model, affective, biological, and cognitive factors converge to create an overall vulnerability to depression. Rooted in a vulnerability $\times$ stress approach, the model holds that negative life events interact with the depressogenic vulnerability, yielding increased levels of depression in adolescence, especially for girls. Biological factors in the model include genetic factors such as the 5HTTLPR polymorphism as well as pubertal hormones and pubertal timing (early, on time, or late). The affective factors are several dimensions of temperament including, especially, negative affectivity. Three aspects of cognitive vulnerability are included in the model: cognitive vulnerability as defined in the hopelessness theory of depression (Abramson et al. 1989); objectified body consciousness (a cognitive process in which individuals become observers and critics of their bodies and appearance, McKinley & Hyde 1996; see also Fredrickson & Roberts 1997); and rumination (the tendency to think repetitively and passively about the negative emotions elicited by negative events, Nolen-Hoeksema 2001). According to the ABC model, it is likely that multiple factors contribute to the gender difference in depression, including gender differences in stress beginning in adolescence, girls' greater vulnerability to the negative effects of early puberty, girls' greater objectified body consciousness, and the adverse effects of peer sexual harassment on girls.

Rumination. As noted, some theorists have argued that gender differences in depression can be accounted for at least in part by gender differences in rumination. The one available meta-analysis found that $d = -0.14$ in child samples and -0.36 in adolescent samples (Rood et al. 2009). Thus the gender difference in rumination widens from childhood to adolescence, in parallel to the emerging gender difference in depression.

Self-esteem. The popular media contend that girls have terrible problems with self-esteem beginning in adolescence and, by implication, that boys do not have self-esteem problems. A 1999 meta-analysis tested these claims (Kling et al. 1999). Averaged over all ages, $d = 0.21$, indicating a small difference favoring males but clearly not the large difference that one might expect based on media reports. The effect size increased from 0.16 in elementary school to 0.23 in middle school and 0.33 in high school, but then declined to 0.18 among college students and 0.10 among adults

between the ages of 23 and 59, so that the gender difference was not large for any age group. An analysis by ethnicity for US samples showed that the magnitude of the gender difference was $d = 0.20$ among whites but −0.04 among blacks. That is, the much touted gender difference may be found only among whites and not among ethnic minorities (too few samples of other ethnic minorities were available for analysis).

EMCP: engineering, mathematics, computer science, and physics

The meta-analysis described above synthesized studies that measured global self-esteem. Another approach to studying self-esteem is to measure domain-specific self-esteem or self-concept—for example, self-concept of math ability or self-concept of athletic ability. Another meta-analysis examined studies that had measured domain-specific self-esteem (Gentile et al. 2009). Males scored higher than females on physical appearance ($d = 0.35$), athletic (0.41), and self-satisfaction (0.33) self-esteem, whereas females scored higher on behavioral conduct ($d = -0.17$) and moral-ethical (-0.38) self-esteem. For all other domains, gender similarities were found; effect sizes were close to 0 for academic, social, and family self-esteem.

The STEM Issue

It is well known that women are underrepresented in careers in science, technology, engineering, and mathematics (STEM), although this global statement is not entirely consistent with a more nuanced view of the data. Today, for example, women earn 53% of the PhDs in biology, 48% of the MD degrees, and 78% of the veterinary degrees (Natl. Cent. Educ. Stat. 2012). Those fields display gender similarities, or even a preponderance of women in the case of veterinarians. In contrast, women earn only 18% of the undergraduate degrees in engineering, 31% of the doctoral degrees in mathematics, 22% of the doctoral degrees in computer science, and 19% of the doctoral degrees in physics (Natl. Sci. Found. 2011). It is in these areas that the gender gaps lie.

A variety of explanations have been offered for the gender gaps in engineering, mathematics, computer science, and physics (EMCP, which is a distinct subset of STEM). Then-president of Harvard, Lawrence Summers, famously opined that women lacked the mathematical ability to succeed in these areas, although that argument is contradicted by the evidence for gender similarities in mathematics performance presented previously (but see the discussion of greater male variability below). Other explanations include a chilly institutional climate for women graduate students and faculty in these disciplines (Stewart & LaVaque-Manty 2008); parents' and teachers' gender-stereotyped beliefs and behaviors that discourage girls from these areas (Chhin et al. 2008, Jussim et al. 1996); girls' and women's lack of interest in EMCP and preference for other fields (Ceci & Williams 2011); stereotype threat and lack of a sense of belonging in these fields (Miyake et al. 2010); the distracting effect of romance (Holland & Eisenhart 1990, Park et al. 2011); and women's family roles, which prevent them from working 80 hours per week, or perceptions that these occupations are not family friendly (Frome et al. 2008). A newer and more novel explanation concerns societal values that emphasize self-expressive value systems, prominent in the United States and other postindustrial nations, in which people seek to take courses and find an occupation that is highly interesting and that they will love (Charles & Bradley 2009). In such a societal context, women think that engineering will not be interesting and opt for something that will be. The researchers found that the gender gap in engineering was largest in postindustrial/postmaterialist nations, such as Finland, Germany, Switzerland, and Hong Kong. The gender gap in engineering was smaller in materialist nations with developing economies such as Bulgaria, Colombia, Latvia, and Romania, where the goal is to get a stable job that pays well rather than following one's loves.

It is beyond the scope of this article to review in depth each of these hypotheses and the evidence for them. It is worth noting, though, that the explanation based on women's family roles is inconsistent with the success of women in biology and psychology; surely it takes as

many hours of work per week to succeed in those sciences as it does in EMCP. Moreover, recent evidence confirms that bias against women job candidates continues to exist in the sciences, so that discrimination remains a viable hypothesis (Moss-Racusin et al. 2012).

Greater male variability hypothesis: the belief that the distribution of scores is larger for males than females, on intelligence or other abilities, so that there are more high-scoring males than females, but also more low-scoring males

Variance ratio: the ratio of male variance to female variance, used to test the greater male variability hypothesis

GENDER DIFFERENCES AND SIMILARITIES IN VARIABILITY

The previous section reviewed evidence on gender similarities and differences in average scores. However, another possibility involves gender differences not in means but in variance. The greater male variability hypothesis was originally proposed by Henry Havelock Ellis in 1894 to explain the excess of males among the mentally defective as well as among geniuses (Shields 1975). The idea is that the variance of the distribution of scores for males is larger than the variance for females, creating an excess of males in both the high and low tails of the distribution. It is this hypothesis that Summers apparently had in mind when trying to explain the dearth of women faculty in mathematics at Harvard.

The statistic used to evaluate the greater male variability hypothesis is the variance ratio, equal to the ratio of male variance to female variance. Thus, values greater than 1.0 indicate greater male variance and values less than 1.0 indicate greater female variance.

Several meta-analyses have examined variance ratios for outcomes including mathematics performance (Hedges & Nowell 1995, Hyde et al. 2008a, Lindberg et al. 2010), verbal and spatial performance (Hedges & Nowell 1995), and temperament (Else-Quest et al. 2006). An overview of these findings is provided in **Table 2**.

For mathematics performance, across three meta-analyses and a wide variety of samples, variance ratios consistently range between 1.05 and 1.20. Males display more variability, but the variance ratios are not very far from 1.0, i.e., the variances are not radically different. Similarly, for verbal performance, variance ratios range between 1.03 and 1.16. For temperament, two variance ratios are less than 1.0—for fear and emotionality—indicating greater female variability. However, none of these variance ratios are very far from 1.0, indicating equal variability for males and females.

What is the effect of variance ratios such as these on the number of males and females in the upper tail of the distribution, e.g., the distribution of mathematics performance? Hedges & Friedman (1993) computed the predicted ratio of the number of males to the number of females in the upper tail for various combinations of d and variance ratio values, assuming a normal distribution (which may or may not be accurate for a given data set). For example, for $d = 0.01$ and variance ratio $= 1.05$ (values similar to those shown in **Table 2**), the ratio of males to females above the 95 percentile should be 1.11, i.e., there will be 111 males for every 100 females—not so different from 50–50. At a more extreme point on the upper end of the distribution, the 99.9 percentile, the gender ratio would be 1.33, i.e., 133 males for every 100 females. If we take slightly larger values of d and variance ratio, $d = 0.05$ and variance ratio $= 1.12$, then the gender ratio is 1.34 at the 95 percentile and 2.15 at the 99.9 percentile. Variance ratios such as these cannot begin to explain why only 18% of the undergraduate degrees in engineering go to women, particularly because such a major does not require mathematical skills at the 99.9 percentile.

It is also important to note that, even if there is slightly greater male variability for some cognitive measures, this finding is simply a description of the phenomenon. It does not address the causes of greater male variability, which could be due to biological factors, sociocultural factors, or both (Johnson et al. 2008, 2009).

GENDER SIMILARITIES, DIFFERENCES, AND INTERSECTIONALITY

To this point, this review has described the evidence regarding gender differences and similarities in isolation, neglecting other important social categories such as ethnicity and sexual orientation.

Table 2 Variance ratios (M:F) in meta-analyses of mathematics, verbal, and spatial performance, and temperament

Domain	VR (M:F)
Verbal	
Hedges & Nowell (1995), Project Talent	1.16
Hedges & Nowell (1995), NLS-72	1.03
Hedges & Nowell (1995), NLSY	1.16
Hedges & Nowell (1995), HS&B	1.10
Hedges & Nowell (1995), NELS:88	1.16
Mathematics	
Hedges & Nowell (1995), Project Talent	1.20
Hedges & Nowell (1995), NLS-72	1.05
Hedges & Nowell (1995), NLSY	1.19
Hedges & Nowell (1995), HS&B	1.16
Hedges & Nowell (1995), NELS:88	1.06
Lindberg et al. (2010), NLSY, 2002	1.05
Lindberg et al. (2010), NAEP, 2000	1.18
Hyde et al. (2008a), Grade 2	1.11
Hyde et al. (2008a), Grade 11	1.17
Spatial	
Hedges & Nowell (1995), Project Talent	1.27
Hedges & Nowell (1995), HS&B	1.27
Temperament (Else-Quest et al. 2006)	
Persistence	1.00
Inhibitory control	1.09
Emotionality	0.94
Fear	0.93
Activity	1.00–1.06

Abbreviations: HS&B, High School and Beyond; M:F, ratio of males to females; NAEP, National Assessment of Educational Progress; NELS:88, National Education Longitudinal Study of 1988; NLS-72, National Longitudinal Study of the High School Class of 1972; NLSY, National Longitudinal Survey of Youth; VR, variance ratio (the ratio of male variance to female variance).

Intersectionality is an approach that simultaneously considers multiple categories of identity, difference, and disadvantage, such as gender, race, class, sexual orientation, disability, and religion (Cole 2009). Intersectionality holds that gender effects can never be understood in isolation and must always be examined in context, i.e., in the context of ethnicity and other social identities and categories. One implication is that scientists should not make global statements about gender differences, referring to an entire nation, much less to cross-culturally universal differences. As an example, the case of the intersection of gender and ethnicity is considered in more detail.

In the meta-analysis of gender differences in self-esteem described previously, the magnitude of the gender difference was $d = 0.20$ among whites but -0.04 among blacks (Kling et al. 1999). That is, the small but much publicized gender gap in self-esteem was found for whites but not blacks. Global statements about gender differences in self-esteem miss the mark.

Similarly, an early meta-analysis of research on gender differences in mathematics performance found $d = 0.13$ for whites, but $d = -0.02$ for blacks, 0.00 for Hispanics, and -0.09 for Asian

Americans (Hyde et al. 1990a). The male advantage (although very small) was found for whites and was not present for blacks, Hispanics, or Asian Americans.

The meta-analysis of gender differences in self-conscious emotions also found evidence of the intersection of gender and ethnicity (Else-Quest et al. 2012). For example, for shame, $d = -0.32$ for whites but -0.06 for nonwhites. That is, the overall finding of $d = -0.29$ for shame obscured variations in the magnitude of the gender difference across ethnic groups and, in particular, masked the absence of the difference among nonwhites.

Few gender meta-analyses have looked for variations in the magnitude of the gender difference as a function of ethnicity or other potential moderators such as social class. Meta-analysis provides an excellent method for the analysis of the intersection of gender and ethnicity. Future gender meta-analyses should routinely consider ethnicity and social class as potential moderators. Several factors will complicate this effort. First, ethnic groups are specific to nations. The examples given in the previous paragraphs were for US samples. Other nations have different relevant ethnic groups, with different meanings attached to membership in those groups. At the least, meta-analysts can consider ethnicity in US samples and examine it in samples from other nations based on an understanding of relevant ethnic groups in those nations. Second, many researchers are negligent about reporting the ethnicity of their samples, despite American Psychological Association style guidelines that mandate this reporting. For example, in the math meta-analysis, 70 studies reported ethnicity and the remaining 184 did not (Hyde et al. 1990a). Moreover, in studies with samples covering multiple ethnicities, the data typically are not analyzed for gender differences separately by ethnicity. If we are to realize the potential of intersectional approaches in psychology, researchers at the least must improve their reporting of the gender and ethnicity of their samples and, ideally, analyze gender $\times$ ethnicity interactions.

SUMMARY AND FUTURE DIRECTIONS

Overall, based on the numerous meta-analyses reported here, there is much evidence in support of the gender similarities hypothesis. Domains in which gender differences are small (around $d = 0.20$) or trivial ($d \leq 0.10$) include mathematics performance, verbal skills, some personality dimensions such as gregariousness and conscientiousness, reward sensitivity, the temperament dimension of negative affectivity, relational aggression, tentative speech, some aspects of sexuality (e.g., oral sex experience, attitudes about extramarital sex, attitudes about masturbation), leadership effectiveness, self-esteem, and academic self-concept.

Nonetheless, the gender similarities hypothesis acknowledges exceptions to the general rule. Exceptions to gender similarities, where differences are moderate ($d = 0.50$) or large ($d = 0.80$), include 3D mental rotation, the personality dimension of agreeableness/tender-mindedness, sensation seeking, interests in things versus people, physical aggression, some sexual behaviors (masturbation and pornography use), and attitudes about casual sex.

This review also reveals much evidence of the importance of context in creating or erasing gender differences. For example, deindividuation, which removes the influence of gender roles, erases the gender difference in aggression. Nations with greater gender equality have much smaller gender gaps in mathematics performance and in mate preferences.

Mountains of research have been conducted on psychological gender differences, yielding the patterns of results described above. Given the current state of research, what are the most important directions for future research?

A distinctly different approach holds that the search for gender differences and their causes is not a productive research strategy. For example, in the case of the gender difference in spatial performance, which is moderately large, Newcombe and colleagues (2002) argued that documenting

gender differences and their cause was relatively unimportant and that the real goal should be to find training methods or interventions that would maximize everyone's spatial competence. When researchers find a gender difference, they might productively ask themselves, is this important, and why is it important? Are other issues more important?

Nonetheless, research on psychological gender differences will continue for years to come, given many scientists' firm beliefs that such differences exist and are large and the media's insatiable thirst for new findings of gender differences. What, then, are the best directions to take this research? Two approaches seem especially promising: intersectional approaches and contextual approaches.

Intersectional approaches to the study of gender similarities and differences can profitably investigate the intersection of gender and ethnicity and the intersection of gender and social class. Researchers, reviewers, and editors should no longer settle for a report of a gender difference. They must ask about the ethnicity of the sample and whether the gender difference, if found in a predominantly white sample, is also found in ethnic minority samples. Researchers should chart the magnitude and direction of gender differences across different ethnic groups. The same is true for the intersection of gender and social class. With college-student samples, these analyses are difficult because the participants are predominantly middle class. Nonetheless, even in college samples, some individuals are first-generation college students whereas others have parents who are college graduates, giving researchers some leverage in studying the intersection of gender and social class. If samples of youth are recruited from primary or secondary schools, recruiting should occur over a range of schools that vary in the social class of the students, so that researchers can examine the magnitude and direction of gender differences at various social class levels. Meta-analysts should routinely examine race and social class as moderators of the magnitude of gender differences.

Identifying contexts in which gender differences appear or disappear will continue to be an important strategy, both in primary research and in meta-analyses. Eagly & Crowley's (1986) meta-analysis of research on gender differences in helping behavior provides a model. They considered theoretically derived variables that should moderate the gender difference in helping behavior: The male role emphasizes heroic and chivalrous helping, whereas the female role emphasizes nurturant helping. For example, when the helping behavior was observed (encouraging heroic helping), the effect size for the gender difference was large, $d = 0.74$. In contrast, when the helping was not observed, there was no gender difference, $d = -0.02$. Large gender differences can be created and erased by the context. Situational variables such as these must be taken into account in meta-analyses, and they warrant much more primary research, guided by theories such as sociocultural theory, expectancy-value theory, and cognitive social learning theory.

Above all, researchers should keep the possibilities of gender differences and gender similarities in balance as they report and interpret their findings. There are serious costs to an overemphasis on gender differences, such as beliefs that boys and girls are so different that they must be educated in gender-segregated schools or a belief that marital therapy for heterosexual couples cannot succeed because of profound gender differences in communication styles. A nonsignificant gender difference, that is, a gender similarity, is as interesting and important as a gender difference.

SUMMARY POINTS

1. The gender similarities hypothesis states that males and females are similar on most, but not all, psychological variables. The current review found much evidence in support of gender similarities.

2. Meta-analysis is a statistical technique that can be used to synthesize previous studies of gender differences on a particular psychological attribute. It focuses on the effect size, that is, how large the difference between males and females is.
3. Cognitive social learning theory explains psychological gender differences as being a result of females and males receiving different rewards and punishments for their behaviors, people's tendency to imitate same-gender models, and cognitive processes such as attention and self-efficacy.
4. Sociocultural theory argues that contemporary psychological gender differences have their origins in the prehistoric division of labor by gender; once males and females take on different roles, they develop the psychological qualities that equip them for those roles.
5. Contemporary data indicate that girls have reached parity with boys in mathematics performance, from grades 2 through 11.
6. Gender differences persist in spatial performance, in particular three-dimensional mental rotation. A spatial curriculum should be implemented in the schools that would improve the performance of both girls and boys.
7. Intersectionality is a research approach that emphasizes the simultaneous consideration of multiple categories of identity and disadvantage, including gender, ethnicity, social class, and sexual orientation.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Preparation of this review was supported in part by grants REC 0635444 and DRL 1138114 from the National Science Foundation. Any opinions expressed are those of the author and not the National Science Foundation. I thank Nicole Else-Quest and Amy Mezulis for insightful comments on an earlier draft of this review.

LITERATURE CITED

Abramson LY, Metalsky G, Alloy L. 1989. Hopelessness depression: a theory-based subtype of depression. *Psychol. Rev.* 96:358–72

Alexander MG, Fisher TD. 2003. Truth and consequences: using the bogus pipeline to examine sex differences in self-reported sexuality. *J. Sex Res.* 4:27–35

Alink LR, Mesman J, van Zeijl J, Stolk MN, Juffer F, et al. 2006. The early childhood aggression curve: development of physical aggression in 10- to 50-month-old children. *Child Dev.* 77:954–66

Archer J. 2004. Sex differences in aggression in real-world settings: a meta-analytic review. *Rev. Gen. Psychol.* 8:291–322

Ball LC, Cribbie RA, Steele JR. 2013. Beyond gender differences: using tests of equivalence to evaluate gender similarities. *Psychol. Women Q.* In press

Bettencourt BA, Kernahan C. 1997. A meta-analysis of aggression in the presence of violent cues: effects of gender differences and aversive provocation. *Aggress. Behav.* 23:447–56

Broverman IK, Vogel SR, Broverman DM, Clarkson FE, Rosenkrantz PS. 1972. Sex role stereotypes: a current appraisal. *J. Soc. Issues* 28:59–78

Buss DM, Schmitt DP. 1993. Sexual strategies theory: an evolutionary perspective on human mating. *Psychol. Rev.* 100:204–32

Bussey K, Bandura A. 1999. Social cognitive theory of gender development and differentiation. *Psychol. Rev.* 106:676–713

The fundamental statement of cognitive social learning as it applies to gender development.

Carothers BJ, Reis HT. 2013. Men and women are from earth: examining the latent structure of gender. *J. Personal. Soc. Psychol.* 104:385–407

Ceci SJ, Williams WM. 2011. Understanding current causes of women's underrepresentation in science. *Proc. Natl. Acad. Sci.* 108:3157–62

Chaplin TM, Aldao A. 2013. Gender differences in emotion expression in children: a meta-analytic review. *Psychol. Bull.* In press

Charles M, Bradley K. 2009. Indulging our gendered selves? Sex segregation by field of study in 44 countries. *Am. J. Sociol.* 114:924–76

Chhin CS, Bleeker MM, Jacobs JE. 2008. Gender-typed occupational choices: the long-term impact of parents' beliefs and expectations. In *Gender and Occupational Outcomes*, ed. H Watt, J Eccles, pp. 215–34. Washington, DC: Am. Psychol. Assoc.

Cohen J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum. 2nd ed.

Cole ER. 2009. Intersectionality and research in psychology. *Am. Psychol.* 64:170–80

An explanation of the concept of intersectionality and its applications in psychology.

Costa PT, Terracciano A, McCrae RR. 2001. Gender differences in personality traits across cultures: robust and surprising findings. *J. Personal. Soc. Psychol.* 81:322–31

Crawford M, Popp D. 2003. Sexual double standards: a review and methodological critique of two decades of research. *J. Sex Res.* 40:13–26

Crick NR, Grotpeter JK. 1995. Relational aggression, gender, and social psychological adjustment. *Child Dev.* 66:710–22

Cross CP, Copping LT, Campbell A. 2011. Sex differences in impulsivity: a meta-analysis. *Psychol. Bull.* 137:97–130

Deaux K. 1985. Sex and gender. *Annu. Rev. Psychol.* 36:49–81

Del Giudice M, Booth T, Irwing P. 2012. The distance between Mars and Venus: measuring global sex differences in personality. *PLoS ONE* 7:e29265

Durik AM, Hyde J, Marks A, Roy A, Anaya D, Schultz G. 2006. Ethnicity and gender stereotypes of emotion. *Sex Roles* 54:429–45

Eagly AH. 2009. The his and hers of prosocial behavior: an examination of the social psychology of gender. *Am. Psychol.* 64:644–58

Eagly AH, Crowley M. 1986. Gender and helping behavior: a meta-analytic review of the social psychological literature. *Psychol. Bull.* 100:283–308

Eagly AH, Johannesen-Schmidt MC, van Engen ML. 2003. Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. *Psychol. Bull.* 129:569–91

A complete statement of sociocultural theory together with evidence for it.

Eagly AH, Karau S, Makhijani M. 1995. Gender and the effectiveness of leaders: a meta-analysis. *Psychol. Bull.* 117:125–45

Eagly AH, Steffen VJ. 1986. Gender and aggressive behavior: a meta-analytic review of the social psychological literature. *Psychol. Bull.* 100:309–30

Eagly AH, Wood W. 1999. The origins of sex differences in human behavior: evolved dispositions versus social roles. *Am. Psychol.* 54:408–23

Eccles JS. 1994. Understanding women's educational and occupational choices: applying the Eccles et al. model of achievement-related choices. *Psychol. Women Q.* 18:585–610

A review of research relevant to expectancy-value theory.

Else-Quest NM. 2012. Gender differences in temperament. In *Handbook of Temperament*, ed. M Zentner, R Shiner, pp. 479–96. New York: Guilford

Else-Quest NM, Higgins A, Allison C, Morton LC. 2012. Gender differences in self-conscious emotional experience: a meta-analysis. *Psychol. Bull.* 138:947–81

Else-Quest NM, Hyde JS, Goldsmith HH, Van Hulle C. 2006. Gender differences in temperament: a meta-analysis. *Psychol. Bull.* 132:33–72

Else-Quest NM, Hyde JS, Linn MC. 2010. Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychol. Bull.* 136:103–27

Feingold A. 1994. Gender differences in personality: a meta-analysis. *Psychol. Bull.* 116:429–56

Feng J. 2007. Playing an action video game reduces gender differences in spatial cognition. *Psychol. Sci.* 18:850–55

Fisher TD. 2007. Sex of experimenter and social norm effects on reports of sexual behavior in young men and women. *Arch. Sex. Behav.* 36:89–100

Fredrickson B, Roberts T. 1997. Objectification theory: toward understanding women's lived experiences and mental health risks. *Psychol. Women Q.* 21:173–206

Frome PM, Alfeld CJ, Eccles JS, Barber BL. 2008. Is the desire for a family-flexible job keeping young women out of male-dominated occupations? In *Gender and Occupational Outcomes*, ed. H Watt, J Eccles, pp. 195–214. Washington, DC: Am. Psychol. Assoc.

Gentile B, Grabe S, Dolan-Pascoe B, Twenge JM, Wells BE, Maitino A. 2009. Gender differences in domain-specific self-esteem: a meta-analysis. *Rev. Gen. Psychol.* 13:34–45

Gerson H, Sorby SA, Wysocki A, Baartmans BJ. 2001. The development and assessment of multimedia software for improving 3-D visualization skills. *Comput. Appl. Eng. Educ.* 9:105–13

Gisborne T. 1797. *An Enquiry Into the Duties of the Female Sex*. London: Cadell & Davies

Hankin BL, Abramson L, Moffitt T, Silva P, McGee R, Angell K. 1998. Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. *J. Abnorm. Psychol.* 107:128–40

Hedges LV, Friedman L. 1993. Gender differences in variability in intellectual abilities: a reanalysis of Feingold's results. *Rev. Educ. Res.* 63:94–105

Hedges LV, Nowell A. 1995. Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science* 269:41–45

Hines M. 2011. Gender development and the human brain. *Ann. Rev. Neurosci.* 34:69–88

Holland DC, Eisenhart MA. 1990. *Educated in Romance: Women, Achievement, and the College Culture*. Chicago: Univ. Chicago Press

Hollingworth LS. 1918. Comparison of the sexes in mental traits. *Psychol. Bull.* 15:427–32

Hyde JS. 1984. How large are gender differences in aggression? A developmental meta-analysis. *Dev. Psychol.* 20:722–36

Hyde JS. 2005. The gender similarities hypothesis. *Am. Psychol.* 60:581–92

Hyde JS. 2006. Gender similarities still rule. *Am. Psychol.* 61:641–42

Hyde JS, Fennema E, Lamon S. 1990a. Gender differences in mathematics performance: a meta-analysis. *Psychol. Bull.* 107:139–55

Hyde JS, Lindberg SM, Linn MC, Ellis A, Williams C. 2008a. Gender similarities characterize math performance. *Science* 3 21:494–95

Hyde JS, Linn MC. 1988. Gender differences in verbal ability: a meta-analysis. *Psychol. Bull.* 104:53–69

Hyde JS, Mezulis AH, Abramson LY. 2008b. The ABCs of depression: integrating affective, biological and cognitive models to explain the emergence of the gender difference in depression. *Psychol. Rev.* 115:291–313

Joel D. 2012. Genetic-gonadal-genitals sex (3G-sex) and the misconception of brain and gender, or, why 3G-males and 3G-females have intersex brain and intersex gender. *Biol. Sex Differ.* 3:27

Johnson W, Carothers A, Deary IJ. 2008. Sex differences in variability in general intelligence. *Perspect. Psychol. Sci.* 3:518–31

Johnson W, Carothers A, Deary IJ. 2009. A role for the X chromosome in sex differences in variability in general intelligence? *Perspect. Psychol. Sci.* 4:598–621

Jonason PK, Fisher TD. 2009. The power of prestige: why young men report having more sex partners than young women. *Sex Roles* 60:151–59

Jussim L, Eccles J, Madon S. 1996. Social perception, social stereotypes, and teacher expectations: accuracy and the quest for the powerful self-fulfilling prophecy. In *Advances in Experimental Social Psychology*, ed. MP Zanna, 28:281–388. San Diego, CA: Academic

Kessler RC, McGonagle K, Swartz M, Blazer D, Nelson C. 1993. Sex and depression in the National Comorbidity Survey: I. Lifetime prevalence, chronicity and recurrence. *J. Affect. Disord.* 29:85–96

Presents the original formulation of Hyde's gender similarities hypothesis and the evidence for it.

Reviews research on affective, biological, and cognitive factors that contribute to gender differences in depression.

Kling KC, Hyde JS, Showers CJ, Buswell BN. 1999. Gender differences in self-esteem: a meta-analysis. *Psychol. Bull.* 125:470–500

Leaper C, Robnett RD. 2011. Women are more likely than men to use tentative language, aren't they? A meta-analysis testing for gender differences and moderators. *Psychol. Women Q.* 35:129–42

Lightdale JR, Prentice DA. 1994. Rethinking sex differences in aggression: aggressive behavior in the absence of social roles. *Personal. Soc. Psychol. Bull.* 20:34–44

Lindberg SM, Hyde JS, Petersen J, Linn MC. 2010. New trends in gender and mathematics performance: a meta-analysis. *Psychol. Bull.* 136:1123–35

Linn MC, Petersen AC. 1985. Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Dev.* 56:1479–98

Luppa M, Sikorski C, Luck T, Ehreke L, Konnopka A, Riedel-Heller SG. 2012. Age- and gender-specific prevalence of depression in latest-life—systematic review and meta-analysis. *J. Affect. Disord.* 136:212–21

Maccoby EE, Jacklin CN. 1974. *The Psychology of Sex Differences*. Stanford, CA: Stanford Univ. Press

Maeda Y, Yoon SY. 2013. A meta-analysis on gender differences in mental rotation ability measured by the Purdue Spatial Visualization Tests: Visualization of rotations (PSVT:R). *Educ. Psychol. Rev.* 25:69–94

Marsh HW, Bornmann L, Mutz R, Daniel H, O'Mara A. 2009. Gender effects in the peer reviews of grant proposals: a comprehensive meta-analysis comparing traditional and multilevel approaches. *Rev. Educ. Res.* 79:1290–326

Martin CL, Ruble DN. 2010. Patterns of gender development. *Annu. Rev. Psychol.* 61:353–81

McCrae RR, Costa P. 2013. Introduction to the empirical and theoretical status of the five-factor model of personality traits. In *Personality Disorders and the Five-Factor Model of Personality*, ed. TA Widiger, PT Costa, pp. 15–27. Washington, DC: Am. Psychol. Assoc. 3rd ed.

McKinley NM, Hyde JS. 1996. The Objectified Body Consciousness Scale: development and validation. *Psychol. Women Q.* 20:181–215

Meece JL, Eccles-Parsons J, Kaczala CM, Goff SB, Futterman R. 1982. Sex differences in math achievement: toward a model of academic choice. *Psychol. Bull.* 91:324–48

Miyake A, Kost-Smith LE, Finkelstein ND, Pollock SJ, Cohen GL, Ito TA. 2010. Reducing the gender achievement gap in college science: a classroom study of values affirmation. *Science* 330:1234–37

Moore DS, Johnson SP. 2008. Mental rotation in human infants. *Psychol. Sci.* 19:1063–66

Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. 2012. Science faculty's subtle gender biases favor male students. *Proc. Natl. Acad. Sci.* 109:16474–79

Natl. Cent. Educ. Stat. 2012. *Digest of Education Statistics.* Washington, DC: US Dep. Educ., Inst. Educ. Sci. **http://nces.ed.gov/programs/digest/**

Natl. Sci. Found. 2011. *S&E Degrees: 1966–2008.* Arlington, VA: Natl. Sci. Found., Natl. Cent. Sci. Eng. Statist. **http://www.nsf.gov/statistics/nsf11316/pdf/nsf11316.pdf**

Newcombe NS, Mathason L, Terlecki M. 2002. Maximization of spatial competence: more important than finding the cause of sex differences. In *Biology, Society, and Behavior: The Development of Sex Differences in Cognition*, ed. A McGillicuddy-De Lisi, R De Lisi, pp. 183–206. Westport, CT: Ablex

Nolen-Hoeksema S. 2001. Gender differences in depression. *Curr. Dir. Psychol. Sci.* 10:173–76

Nosek BA, Smyth FL, Siram N, Lindner NM, Devos T, et al. 2009. National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proc. Natl. Acad. Sci. USA* 106:10593–97

Park LE, Young AF, Troisi JD, Pinkus RT. 2011. Effects of everyday romantic goal pursuit on women's attitudes toward math and science. *Personal. Soc. Psychol. Bull.* 37:1259–73

Pedersen WC, Miller LC, Putcha-Bhagavatula AD, Yang Y. 2002. Evolved sex differences in the number of partners desired? The long and the short of it. *Psychol. Sci.* 13:157–61

Pedersen WC, Putcha-Bhagavatula A, Miller LC. 2011. Are men and women really that different? Examining some of sexual strategies theory (SST)'s key assumptions about sex-distinct mating mechanisms. *Sex Roles* 64:629–43

Petersen JL, Hyde JS. 2010. A meta-analytic review of research on gender differences in sexuality: 1993 to 2007. *Psychol. Bull.* 136:21–38

Quinn PC, Liben LS. 2008. A sex difference in mental rotation in young infants. *Psychol. Sci.* 19:1067–70

Reilly D. 2012. Gender, culture, and sex-typed cognitive abilities. *PLoS ONE* 7(7):e39904
Richardson DS, Hammock GS. 2007. Social context of human aggression: Are we paying too much attention to gender? *Aggress. Viol. Behav.* 12:417–26
Rideout VJ, Foehr UG, Roberts DF. 2010. *Generation M^2: Media in the Lives of 8- to 18-Year-Olds*. Menlo Park, CA: Kaiser Fam. Found.
Rood L, Roelofs J, Bögels SM, Nolen-Hoeksema S, Schouten E. 2009. The influence of emotion-focused rumination and distraction on depressive symptoms in non-clinical youth: a meta-analytic review. *Clin. Psychol. Rev.* 29:607–16
Rucklidge JJ. 2010. Gender differences in attention-deficit/hyperactivity disorder. *Psychiatr. Clin. N. Am.* 33:357–73
Sandberg JF, Hofferth SL. 2001. Changes in children's time with parents: United States, 1981–1997. *Demography* 38:423–36
Schmitt DP. 2005. Sociosexuality from Argentina to Zimbabwe: a 48-nation study of sex, culture, and strategies of human mating. *Behav. Brain Sci.* 28:247–311
Shields S. 1975. Functionalism, Darwinism, and the psychology of women. *Am. Psychol.* 30:739–54
Stewart A, LaVaque-Manty D. 2008. Advancing women faculty in science and engineering: an effort in institutional transformation. In *Gender and Occupational Outcomes*, ed. H Watt, J Eccles, pp. 299–322. Washington, DC: Am. Psychol. Assoc.
Stewart A, McDermott C. 2004. Gender in psychology. *Annu. Rev. Psychol.* 55:519–44
Su R, Rounds J, Armstrong P. 2009. Men and things, women and people: a meta-analysis of sex differences in interests. *Psychol. Bull.* 135:859–84
Subst. Abuse Ment. Health Serv. Adm. 2009. Results from the 2008 National Survey on Drug Use and Health: National Findings. *Off. Appl. Stud.*, *NSDUH Ser. H-36*, *HHS Publ. No. SMA 09–4434*. Rockville, MD: SAMHSA
Swim J, Borgida E, Maruyama G, Myers DG. 1989. Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychol. Bull.* 105:409–29
Tannen D. 1991. *You Just Don't Understand: Women and Men in Conversation*. New York: Ballantine
Thorndike EL. 1914. *Educational Psychology*, Vol. 3. New York: Teachers College, Columbia Univ.
Trivers RL. 1972. Parental investment and sexual selection. In *Sexual Selection and the Descent of Man, 1871–1971*, ed. B Campbell, pp. 136–79. Chicago: Aldine
Twenge JM, Nolen-Hoeksema S. 2002. Age, gender, race, socioeconomic status, and birth cohort differences on the Children's Depression Inventory: a meta-analysis. *J. Abnorm. Psychol.* 111:578–88
Uttal DH, Meadow NG, Tipton E, Hand LL, Alden AR, et al. 2013. The malleability of spatial skills: a meta-analysis of training studies. *Psychol. Bull.* 139:352–402
Voyer D. 2011. Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychon. Bull. Rev.* 18:267–77
Voyer D, Voyer S, Bryden MP. 1995. Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol. Bull.* 117:250–70
Watson D, Clark LA. 1984. Negative affectivity: the disposition to experience aversive emotional states. *Psychol. Bull.* 96:465–90
Wood W, Eagly AH. 2012. Biosocial construction of sex differences and similarities in behavior. *Adv. Exp. Soc. Psychol.* 46:55–123
Woolley HT. 1914. The psychology of sex. *Psychol. Bull.* 11:353–79
Zahn-Waxler C, Shirtcliff EA, Marceau K. 2008. Disorders of childhood and adolescence: gender and psychopathology. *Annu. Rev. Clin. Psychol.* 4:275–303
Zentner M, Mitura K. 2012. Stepping out of the caveman's shadow: Nations' gender gap predicts degree of sex differentiation in mate preferences. *Psychol. Sci.* 23:1176–85