

Women's Underrepresentation in Science: Sociocultural and Biological Considerations

Stephen J. Ceci, Wendy M. Williams, and Susan M. Barnett
Cornell University

The underrepresentation of women at the top of math-intensive fields is controversial, with competing claims of biological and sociocultural causation. The authors develop a framework to delineate possible causal pathways and evaluate evidence for each. Biological evidence is contradictory and inconclusive. Although cross-cultural and cross-cohort differences suggest a powerful effect of sociocultural context, evidence for specific factors is inconsistent and contradictory. Factors unique to underrepresentation in math-intensive fields include the following: (a) Math-proficient women disproportionately prefer careers in non-math-intensive fields and are more likely to leave math-intensive careers as they advance; (b) more men than women score in the extreme math-proficient range on gatekeeper tests, such as the SAT Mathematics and the Graduate Record Examinations Quantitative Reasoning sections; (c) women with high math competence are disproportionately more likely to have high verbal competence, allowing greater choice of professions; and (d) in some math-intensive fields, women with children are penalized in promotion rates. The evidence indicates that women's preferences, potentially representing both free and constrained choices, constitute the most powerful explanatory factor; a secondary factor is performance on gatekeeper tests, most likely resulting from sociocultural rather than biological causes.

Keywords: sex differences, spatial ability, mathematics, family versus career trade-offs, women in science

Supplemental materials: <http://dx.doi.org/10.1037/a0014412.supp>

By 2001, women were earning 48% of bachelor's degrees (National Science Foundation, 2007) and 29% of PhD degrees (Hill & Johnson, 2004) in mathematics, representing enormous increases over the prior 30 years. Women's representation among editorial boards in science and awards panels similarly increased (Nelson & Brammer, 2008). These changes are evidence of the strength of cultural factors in determining such outcomes, because biology has not changed over this period. Despite this progress, women's representation among PhD degree holders has not coincided with proportional faculty appointments: Women earned 31.3% of chemistry PhD degrees between 1993 and 2003 but in 2002 were hired for only 21.5% of assistant professorships. Similar disparities exist for new faculty appointments in physics, engineering, and mathematics. In 1976 women represented only 7.5% of the faculty in physical sciences and less than 1% in engineering (Dearman & Plisko, 1979). By 2006 the percentage had increased to 16%–25%, but the hiring of assistant professors in these fields has not been proportional to female PhD pools. This hiring disparity extends beyond math-intensive fields.

Even in less math-intensive fields, such as cellular and molecular biology, fields in which women obtain 46% of all PhD degrees, women disproportionately drop out at multiple points. The picture is the same across many science fields: Women are not being hired as assistant professors at the rate that they are earning PhD degrees. Of course, this does not mean they are being shunned in the hiring process, as they may be less likely to apply for these positions. The situation in nonacademic venues is comparable in that women leave science, engineering, and technology jobs at twice the rate of men, although this figure includes not only jobholders with PhD degrees but also those with bachelor's and master's degrees (Belkin, 2008).

Underrepresentation of women is even worse farther along the science career path. At the top 50 U.S. universities, the proportion of female full professorships in math-intensive fields ranges from 3% to 15% (*Science and Engineering Indicators*; National Science Foundation, 2005, 2006). Moreover, although women obtain nearly 30% of the doctorates in chemistry, "the further you go up the ladder of prestige and seniority, the less encouraging are the numbers" (Cavallaro, Hansen, & Wenner, 2007, p. 21).

Numerous scholars have opined about the causes of the underrepresentation of women in science and particularly in math-intensive STEM (science, technology, engineering, and mathematics) fields. Hypotheses span biological factors (e.g., effects of brain organization, evolutionary pressures, and prenatal hormones; Eals & Silverman, 1994; Finegan, Niccols, & Sitarenios, 1992) to social factors (e.g., effects of cultural beliefs, discrimination, and stereotypes). In this article, we attempt to reconcile conflicting evidence about causes for women's underrepresentation as profes-

Stephen J. Ceci, Wendy M. Williams, and Susan M. Barnett, Department of Human Development, Cornell University.

We are indebted to endocrinologists and behavioral endocrinologists, especially Elizabeth Adkins-Regan, for advice and to David Lubinski for many helpful comments.

Correspondence concerning this article should be addressed to Stephen J. Ceci, Department of Human Development, Cornell University, Ithaca, NY 14853. E-mail: sjc9@cornell.edu

signals. Unlike other efforts to resolve the debate on this topic (Halpern et al., 2007; Rhoads, 2004; Shalala et al., 2007; Spelke, 2005), our approach consisted of developing a framework to organize qualitative and quantitative evidence from the disciplines of psychology, education, sociology, anthropology, neuroscience, endocrinology, and economics into a causal chain and then evaluating this evidence in terms of the importance of each factor and the strength of the evidence for its effect. Over 400 studies served as inputs, including approximately 20 meta-analyses (and several meta-analyses of meta-analyses).

Definitional Issues and Assumptions

At the outset there are several definitional issues and assumptions to be noted.

Mean Versus Right Tail

The evidence on cognitive sex differences can be divided into mean differences (at the midpoint of a distribution) and right-tail differences in proportions in the top 10%, 5%, and 1%. Clearly, the latter are more relevant to the goal of this synthesis because people in STEM professions come predominantly from the right tail. However, data on more mainstream samples is at times valuable, both because they often foreshadow trends at the right tail and because they can reveal sources of contextual effects, thus illuminating possible causal mechanisms in right-tail groups. In the sections that follow, we report mean and right-tail evidence separately except in those domains where no right-tail data exist.

Definition of Math-Intensive Fields

Many of the fields in which women have been most underrepresented are those that are typically heavily involved with advanced mathematics. However, there are many other fields that involve mathematics (accounting, economics, mathematics teaching, biology, finance), although women are not as underrepresented in most of them. Moreover, there is variability within all math-intensive fields as to their mathematics loading. For example, chemistry and engineering have subspecialties that rely heavily on advanced mathematics and other subspecialties that do not. Moreover, there are specializations within the social sciences and business that are heavily mathematical (e.g., financial modeling). Although we have focused on the underrepresentation of women in STEM fields, nothing about our synthesis excludes other fields that involve advanced mathematics.

Skills Required

A related issue concerns the skills needed to become a successful STEM professional. Surprisingly, there is no consensus. Does one have to be in the top 1% in mathematics to become a physicist? Or is the top 10% good enough? Some research suggests that being in the extreme right tail—that is, the top 0.1% or even the top 0.01%—adds incremental predictive power for STEM success. Little is known about the types of mathematics necessary, if indeed a single type is required for all fields. Some types of mathematics favor men and some favor women, and some measures of ability (SAT Mathematics [SAT-M] scores, national aptitude tests) favor men and others (grades, classroom achievement tests) favor

women. Finally, the role of other cognitive abilities, such as spatial cognition, is unknown, notwithstanding its face validity in many professions (radiology, angular laparoscopy, engineering graphics, n -dimensional projections in chemistry, etc.). Because of these lacunae, there is a risk of circular reasoning in which one is tempted to reason that any observed sex difference is causal.

Causal Framework

In Figure 1, the individual and sociocultural factors hypothesized to affect success in math-intensive STEM careers are shown in a circular relationship: Broad contextual influences, including cultural beliefs, frame the proximal environment, which in turn shapes individual motivation, beliefs, and activities. The latter can influence brain development and the consequent abilities that individuals build over time, as demonstrated by Grabner, Neubauer, and Stern's (2006; see also Grabner, Stern, & Neubauer, 2003) findings that superior cognitive performance and the underlying cortical activation are a function of both prior neural efficiency and experience (e.g., expertise at playing chess or navigating a taxi can affect underlying cortical activation, rendering electroencephalography more efficient, with lower event-related desynchronization). Some abilities are manifest in assessed performances such as on the SAT, achievement test scores (e.g., National Assessment of Educational Progress [NAEP]), and grades (grade point average [GPA]), which, in turn, can affect career-related striving and, ultimately, career status.

That aspects of this circular framework may play a role in individuals' success in STEM careers is uncontroversial. Here we address the mechanism by which biological sex affects career success, that is, the extent to which it influences any or all of these factors. It is easy to imagine how biological sex influences cultural expectations and, consequently, indirectly affects the other outer ring factors. However, biology may also directly affect other factors, such as brain development (through hormones). If cultural expectations are taken as a starting point for understanding sex differences in STEM fields, feeding into the proximal environment experienced by the individual, and so forth around the circle, the figure is closed by the influence of the "end point," specifically, career status in STEM fields, on the starting point—cultural expectations in the environment. Feed-forward and feedback loops reflect the leapfrogging influence of various factors. For example, the activities in which individuals engage, such as getting their name on publications and other aspects of academic productivity, can affect how their performance is assessed, even if these activities do not result in changes in brain development or abilities. Similarly, the motivation to pursue a career, reflected in an individual's life choices, may affect an individual's career status, even absent differences in brain development, abilities, or assessed performance. Some support exists for each of these connections. Also, some links may act both clockwise and counterclockwise, for example, the effect of motivation/activities on brain development and the effect of brain developments on motivation/activities. Susceptibility to stereotypes at the individual level, and cultural bias at the societal level, may affect performance, irrespective of ability. Conversely, career status may influence interests and activities by opening or closing opportunities. Biological sex, at the center of the circle, potentially affects any or all of the variables on the outer ring.

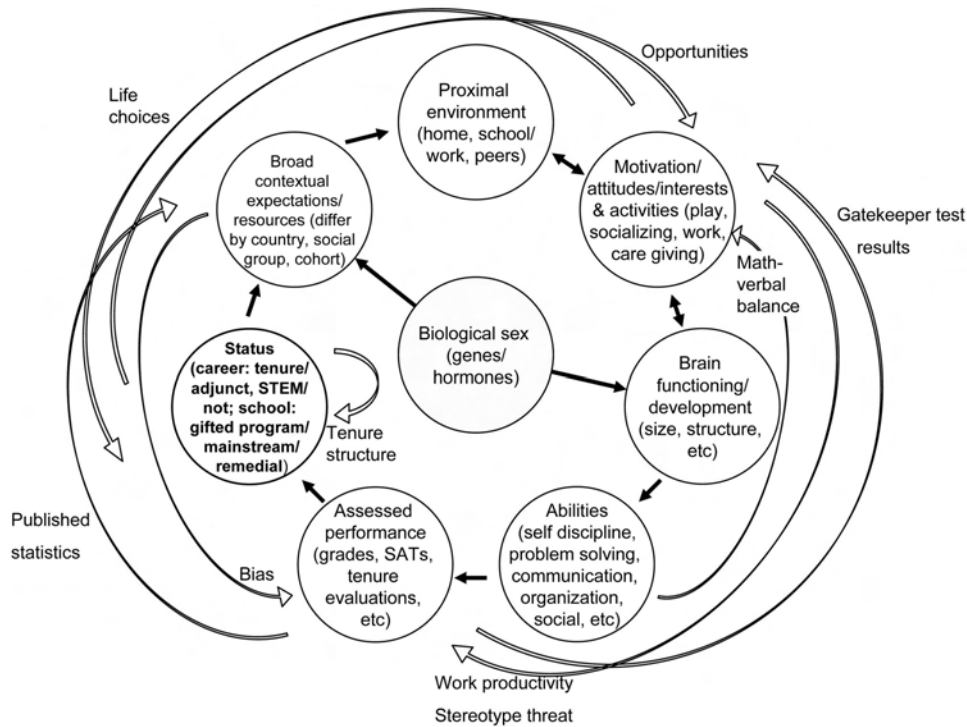


Figure 1. General causal model. STEM = science, technology, engineering, mathematics.

This framework may be used to illustrate potential alternative models that have been proposed to account for underrepresentation of women in STEM fields, all of which are subsumed by the more inclusive model in Figure 1. Versions of these alternatives, representing the most extreme and overly simplified instantiations of the competing hypotheses (intrinsic ability differences vs. intrinsic ability plus differences in interests vs. sociocultural differences) are shown in the online supplemental materials (see Section 1, Figures A1, A2, and A3). Where variables do not play a role in mediating the influence of biological sex on the underrepresentation of women, for simplicity, we do not show them in these straw-person submodels. This does not mean that these variables cease to influence individual success in STEM through non-sex-related factors.

The structure of the remainder of this article is as follows: First, we briefly provide some background and selected data on the performance and representation of women and men. Next, we discuss each component of the framework in Figure 1, starting in the center with the possible effects of biological sex, then moving to the outer circle and working around clockwise, beginning with the effects of broad contextual expectations and resources. Each of the leapfrogging arrows in the framework is discussed at the appropriate point. In summarizing evidence for each of the nodes and links in Figure 1, we show that none of the submodels fully explains the phenomena; thus, we introduce a hybrid model in the Conclusion. Within each of these sections (where data exist), we separate evidence into studies of mean differences, conducted with unselected samples, and studies of selective, right-tail samples. (The exception is the section on hormones in which none of the research deals exclusively with the right tail.) On topics for which

there is a considerable body of evidence, only representative studies are discussed—the bulk of the data are summarized in seven online tables and references. Additional online tables (which we encourage readers to examine) detail the nature of each sample, the measures used, and the results relevant to our thesis. Because of the vast amount of research on some topics, the tables are not exhaustive.

In the Conclusion, we modify the framework in light of the evidence, by boldfacing nodes for which there is evidence of a substantial effect and by varying the width and darkness of each arrow to be congruent with the importance of each link and the convincingness of the evidence.

Background and Contradictions

Hyde (2005) synthesized 128 effect sizes on a broad range of measures from 47 published meta-analyses, and although she concluded that, on net, the sexes were more similar than dissimilar, she reported large effects for mental rotation and mechanical reasoning favoring males (*ds* between .56 and .76), which some have suggested underlie sex differences in advanced math. Hedges and Nowell (1995) examined sex differences in mental test scores in six studies conducted between 1960 and 1992, each based on a national probability sample of adolescents. The distribution of test scores for male and female test takers differed substantially at the top and bottom 1%, 5%, and 10%: Males excelled in science, mathematics, spatial reasoning, social studies, and mechanical skills. Females excelled in verbal abilities, associative memory performance, and perceptual speed. Despite the modest differences at the center of the distribution, the greater variability of male

scores resulted in large asymmetries at the tails, with males outnumbering females by a ratio of 7 to 1 in the top 1% on tests of mathematics and spatial reasoning (see supplemental materials, Section 2, for an explanation of normal curve deviates as they apply to asymmetries at the tails). Such greater male variability has frequently been reported (e.g., Arden & Plomin, 2006; Entwisle, Alexander, & Olson, 1994; Lohman & Lakin, in press; Strand, Deary, & Smith, 2006). For example, Entwisle et al. (1994) found that differences between the variances of boys' and girls' test scores increased significantly with age. In 1st grade, the standard deviations of boys' and girls' scores were similar; however, by 3rd grade, the standard deviation of boys' scores was significantly larger than that of girls (45.6 vs. 37.4), and by 8th grade, the boys' standard deviation was nearly 25% larger. In their analysis of mathematics data for more than 7 million children in 10 states, Hyde, Lindberg, Linn, Ellis, and Williams (2008) reported greater male variance at all grades, on the order of 10%–20%, although the variance ratios did not clearly increase with age. In a recent longitudinal analysis of more than 10,000 British children, Arden and Plomin (2006) reported that when they extracted the first principal component from a battery of cognitive tests, including (at older ages) mathematics tests, greater male variance was found at ages 3, 4, 7, 9, and 10 years. Hence, greater male variance is observed even prior to the onset of preschool.

Hedges and Nowell's (1995) finding of a large male overrepresentation at the right tail in mathematics-related skills was consistent with many, though not all, other studies (cf. Lachance & Mazzocco, 2006; Robinson, Abbott, Berninger, & Busse, 1996): Benbow (1988) reported male–female ratios for the top 0.01% of adolescents (i.e., 1 in 10,000) on the SAT-M of approximately 10:1; moreover, when Benbow and Stanley (1980) screened over 10,000 students for the Study of Mathematically Precocious Youth (SMPY), only 1.7% of girls scored more than 1 *SD* above the female mean, whereas 7.8% of boys scored more than 1 *SD* above the male mean. In earlier work with 450 Baltimore 12- to 14-year-olds, the highest girl's score was surpassed by 43 boys (Stanley, Keating, & Fox, 1974).¹ Finally, Arden and Plomin (2006) reported that on an index of general intelligence, boys were overrepresented in the top and bottom 10% and had greater variance at ages 3, 4, 7, and 10 years. Such findings track with national mathematics achievement test data, including representative samples of British 11- to 12-year-olds in which boys' variability exceeds girls' (Strand et al., 2006) as well as U.S. samples across a wide age range (Lohman & Lakin, in press).

Notwithstanding these findings, there are no longer gender differences in the number of demanding mathematics courses taken in high school, and girls get better grades in such courses than boys (Gallagher & Kaufman, 2005; Kimball, 1989; Mau & Lynn, 2000; Xie & Shauman, 2003). Moreover, in the United Kingdom, the proportion of 16-year-old girls achieving A to C grades in mathematics exceeds the proportion for boys. (The only subject in which boys were noted to outperform girls was physics, in which 90% of boys and 89% of girls achieved an A to C grade; Department for Education and Skills, 2002).

The proportion of women earning bachelor's degrees in STEM fields has increased every year since 1966, and by 2001 women exceeded men earning degrees in some fields. Men and women receive equal grades in college mathematics classes that are of comparable difficulty (Bridgeman & Lewis, 1996), and, as noted,

women now earn almost half of the bachelor's degrees in mathematics. Thus, Spelke (2005) argued, "By the most meaningful measure—the ability to master new, challenging mathematics material over extended time—college men and women show equal aptitude for mathematics" (p. 592).

In addition to their impressive gains in high school and college, women are increasingly attaining doctorates in STEM fields: By 2001, women earned 36.6% of PhD degrees in scientific and engineering fields, up from just 8% in 1966 (Hill & Johnson, 2004), though disproportionately more were earned in less math-intensive fields, such as the social and biological sciences (43.5%–67.1%). Still, women have made impressive gains in attaining doctorates in math-intensive fields as well, obtaining 29% of the PhD degrees in mathematics, 17% in engineering, and 22% in computer sciences. Women's successes have been even greater in other scientific fields, where they have obtained 50% of medical doctor (MD) degrees, almost 75% of doctor of veterinary medicine (DVM) degrees, and 44% of PhD degrees in biological sciences. A generation ago the corresponding percentages were half or less in these fields.

One potential limitation in using the number of men and women who receive bachelor's and PhD degrees as an indicator of expertise in mathematics and science is that it may not represent a sensitive index of the highest potential. We lack data on who become successful STEM scientists and where in the mathematics talent distribution such individuals fell as college students. Of the 48% of undergraduate mathematics majors (National Science Foundation, 2004) and 29% of PhD students who are female, it is unknown what portion is in the talent range of those who become professional mathematicians, or indeed what skills are most relevant. What we do know is that female students' grades in college mathematics classes are as good as those of male students when they take comparable math courses. The aptitude of the subset of such students who go on to become STEM scientists, however, is unknown.

Components of the Framework

To resolve the debate over the causes of sex differences in math-intensive STEM careers, an integrative framework is needed, spanning myriad disciplines, methods, historical epochs, cultures, and developmental levels. In what follows we provide such integration, and return to the framework afterward to adjust it in light of these considerations. In this section, we discuss inputs to the causal framework (Figure 1), beginning with biological sex, at the center of the circle. We then discuss broad contextual factors such as culture and social class, next moving to proximal and motivational inputs, before arguing how they impact ability formation, assessments of ability, and ultimately women's representation in math-intensive STEM careers.

It is important to distinguish between performance at the center of the distribution and that at the right tail. Many studies we review focus on the former. Yet, a small difference at the mean can coexist with a large difference at the extremes—or none at all—depending on the variance and the shape of the distribution. With

¹ SMPY adolescents first qualified by scoring in the top 3%–5% on a grade-based math achievement test given at their schools. Students were then invited to take the SAT. In the beginning, there were slightly more males than females; since 1980, when verbal ability was stressed as much as mathematical ability, the ratio has remained around 50–50.

notable exceptions, studies report fewer women at the right tail of mathematics ability, the part of the distribution where most STEM scientists presumably reside (see Lohman & Lakin, in press). For reading comprehension, perceptual speed, and associative memory, female test takers outnumber male test takers in the top 5% and 10% of scores, and male test takers are 1.5–2.2 times as likely as female test takers to score in the bottom 5% and 10% of the distributions. For both spatial reasoning and mathematics, males are between 1.5 and 2.3 times more likely to be at the high end of the score distribution (including in some analyses ≥ 7 times more likely to be at the top 1%). Where males are hugely overrepresented at the high end is in areas of mechanical/electronic reasoning (by a factor of nearly 10 to 1). Interestingly, there are overrepresentations of male students in top social studies performance by 1.7 to 3.5, which is odd given its similarity to other types of verbal processing on which females excel, although there is no longer an SAT-Verbal [SAT-V] gap between the sexes.

A consequence of focusing on the representations of males and females at the right tail is that ordinary least squares (OLS) procedures that depend on conditional means are not ideal, yet they serve as almost the exclusive basis for analysis. This, in itself, is a source of major inconsistency, with OLS analyses often missing important sex, social class, and ethnic gaps at various quantiles, for example, deciles:

In plotting the OLS results alongside the quantile regression results, we are able to compare the findings obtained by the two methods. We find that OLS results often . . . miss important variation in differences across the distribution. Thus, while differences in the mean are important and tell us about how populations differ on average, differences at the extremes often vary substantially from the mean differences. In cases such as gender differences in mathematics, where the extremes of the distribution are of more interest than the middle, quantile regression or other extreme-sensitive methods should be employed. (Penner & Paret, 2008, p. 249)

Thus, in what follows, we organize the evidence related to each input to the model, starting with biological sex at the core, in terms of whether it pertains to the right tail of the score distribution (e.g., gifted students, top 1%, top 5%), a college sample, or an unselected sample, the former being most relevant for explaining the underrepresentation of women in STEM fields, but the latter shedding light on mechanisms and forecasting trends. As can be seen in Table 1 in the supplemental materials online, there are many more studies of the right tail for some topics (mathematics, bias, specific cognitive tasks) than for others (hormones, stereotype threat). For biological sex, where we start this enterprise, there is no evidence from the right tail, although ample suggestive evidence can be gleaned from animal, clinical, and unselected samples.

Biological Sex

At the core of the model in Figure 1 is the role of biological sex. Two long-standing candidates for causal roles of sex differences in STEM fields are brain structure differences (volume, organization) and the organizing effects of prenatal sex hormones on the brain—specifically, testosterone (see Brosnan, 2006)—and postnatal activating hormonal effects (e.g., puberty, menstruation, contraceptives, circadian and seasonal fluctuations, menopause; Van Goozen, Cohen-Kettenis, Gooren, Frijda, & Van de Poll, 1994,

1995). The hormone research examines the relationship between sex hormones and spatial and cognitive abilities. We discuss this research here. The other body of research concerns male–female brain size and structure differences, which we discuss in a later section on brain functioning and development with other research on brain functioning because, although it is presumed that brain differences would mediate any hormonal effects on cognition, those brain differences are not the primary subject of the hormone research. Here we discuss the role of hormones in sex differences in which nearly all of the evidence comes from clinical and unselected samples and very little from right-tail samples; thus, no right-tail subsection is included in this section. Hormones are hypothesized to affect cognitive abilities, particularly spatial skills, through processes associated with averting programmed cell death and enhancing connectivity in structures such as the hippocampus that are associated with spatial memory in mammals (Baron-Cohen et al., 2005; Giedd et al., 1996). Table 2 in the supplemental materials summarizes much of the peer-reviewed research on cognitive sex difference and hormones. The studies span analyses of special populations with hormonal disorders (e.g., congenital adrenal hyperplasia [CAH], hypogonadotropic hypogonadism), hormone interventions with animals and humans (such as transsexuals seeking sex change, laboratory manipulations), and analyses of normal variations in free-circulating testosterone levels.

Evidence based on populations with hormonal disorders is mixed. Several studies show that spatial abilities are unambiguously related to androgen levels. For example, Resnick, Berenbaum, Gottesman, and Bouchard's (1986) study showed that girls with CAH achieved significantly higher spatial scores than control girls, whereas boys with CAH showed significantly lower spatial scores than control boys, a finding also reported by many others (e.g., Hines et al., 2003; Hines & Kaufman, 1994), although exceptions to this finding have been numerous (e.g., Malouf, Migeon, Carson, Petrucci, & Wisniewski, 2006; Ripa, Johannsen, Mortensen, & Muller, 2003; see Caplan, McPherson, & Tobin, 1985; Schattman & Sherwin, 2007). In their recent meta-analyses, Puts, McDaniel, Jordan, and Breedlove (2008) reported moderate associations between CAH and spatial ability. Males with low androgen levels (due to hypogonadotropic hypogonadism) have lower spatial ability than controls. However, Hines et al. found that differences in spatial ability among females did not covary with their degree of androgen, and Resnick et al. (1986) reported that boys with CAH who had high early testosterone performed similarly to control boys. As Puts et al. argued, excess early androgens might directly affect CAH individuals' spatial ability by altering the neurocognitive systems that subservise spatial cognition such as the hippocampus. Alternatively, androgens might influence CAH children's proclivity to engage in experiences that shape spatial skills (e.g., rough and tumble play; Hines & Kaufman, 1994) or, by masculinizing appearance, might indirectly affect treatment that influences spatial ability, although CAH individuals' external genitalia are often surgically repaired and females are feminine in appearance (see Puts et al., 2008, p. 108).

Evidence based on hormone interventions is also mixed. Slabbekoorn, van Goozen, Megens, Gooren, and Cohen-Kettenis (1999) demonstrated that androgen therapy for genetic female transsexuals led to higher three-dimensional rotation ability compared with their pre-androgen ability, and a number of studies find a U-shaped association between activation levels of testosterone

and mental rotation and mathematics scores (see Hampson & Moffat, 2005, for a review). However, there are other studies of comparable quality that fail to find such relationships (e.g., Christiansen & Knusmann, 1987; McKeever, Rich, Deyo, & Conner, 1987; see Hogervorst, Bandelow, & Moffat, 2005, for a review). As one example, a 3-month cross-over random trial showed that 200 mg of testosterone enanthate had no effect on older men's mental rotation (reviewed in Hogervorst et al., 2005).

Studies of normal variations in hormone levels examine both prenatal organizing hormones and postnatal activating hormones. With regard to the former, Baron-Cohen, Lutchmaya, and Knickmeyer (2004; see also Knickmeyer & Baron-Cohen, 2006) reviewed the effect of male hormones in fetal and amniotic fluid on later spatial and mathematical ability, and Fink, Brookes, Neave, Manning, and Geary (2006) showed a correlation between numeric competencies and finger-length digit ratios such that higher digit ratios (i.e., the second to fourth finger [2D:4D ratio] being more similar in length for women) correlated with lower numerical ability for males. The digit ratio is a marker for prenatal testosterone for the fourth finger and estrogen for the second finger (Manning, 2002; Sanders, Bereczkei, Csatho, & Manning, 2005); in gay men and women, the second finger is typically shorter than the fourth (McFadden et al., 2005; T. J. Williams et al., 2000).² However, others have not found an effect of prenatal hormone levels on cognitive ability (e.g., Finegan et al., 1992), and Puts et al. (2008) found only very small correlations between the 2D:4D ratio and spatial ability in their meta-analyses. Studying normal variations in postnatal activating hormones, Moffat et al. (2002) found a strong effect of testosterone on visual-spatial tests, whereas Davison and Susman (2001) found a relationship between testosterone and spatial cognition for boys in all six contrasts but for girls in only one of the six. Thilers, MacDonald, and Herlitz (2006) failed to find any association between spatial cognition and testosterone.

Further, despite some studies finding effects of testosterone on spatial performance, such findings cannot be readily generalized to differences at the right tail of the spatial distribution. The optimal level of testosterone for spatial performance appears to be near the low end of the normal male range (Brosnan, 2006) or the high end of the female range. Very high testosterone is often associated with reduced, rather than enhanced, spatial scores, producing an inverted U-shaped function, with both very low and very high levels being deleterious for mental rotation and math performance. Physical scientists' digit ratios are closer to female ratios than to male social scientists' ratios (Brosnan, 2006), consistent with the suggestion that the middle of the inverted U-shaped function is associated with STEM success.

Across the three categories of hormone studies—disorders, artificial interventions, normal variations—a number of studies suggest a U-shaped relationship between activational effects of testosterone on mental rotation and mathematics performance (see Hampson & Moffat, 2005, for review); however, other studies failed to find such a relationship (e.g., Christiansen & Knusmann, 1987; McKeever, Rich, Deyo & Conner, 1987; see Hogervorst et al., 2005, for a review). Table 2 in the supplemental materials summarizes much of this evidence and reveals additional inconsistencies within and across the three types of studies.

There are also methodological issues that arise in the hormone studies: Some studies that have reported a relationship between

testosterone and spatial cognition have relied on marginal trends (e.g., Neave, Menaged, & Weightman, 1999) or have reported such findings only after removal of aberrant subjects (Grimshaw, Sitarenios, & Finegan, 1995). Sample sizes sometimes have been understandably small, especially for studies of CAH, Turner's syndrome, and transsexual changes. Furthermore, the results have often been quite nuanced in terms of the changes that result from shifting measures, such as spatial targeting with darts versus balls (e.g., Hines et al., 2003) or observations that the relationship between testosterone and spatial cognition is much stronger for males than females only for some measures and at certain times (Davison & Susman, 2001; see Section 3 in the supplemental materials for emblematic inconsistencies). Another example of ambiguity occurs around the question of why the relationship between CAH and spatial ability has been moderately strong in meta-analyses, whereas the relationship between the 2D:4D ratio and spatial ability has not, suggesting that perhaps the critical period for androgens to affect brain regions that subserve spatial ability is later than the first trimester when the 2D:4D ratio occurs (Malas, Dogan, Hilal Evcil, & Desdicioglu, 2006), inasmuch as the CAH effects occur in the later part of pregnancy or even postnatally (Puts et al., 2008). Finally, if elevated prenatal androgens are responsible for spatial ability differences between CAH and non-CAH individuals, this is complicated by the fact that CAH individuals also differ in glucocorticoid levels, which could influence spatial cognition.

Further, despite some studies finding effects of testosterone on spatial performance, such findings cannot be readily generalized to differences at the right tail of the spatial distribution. Recall that the optimal level of testosterone is near the low end of the normal male range (Brosnan, 2006), the high end of the female range, or the moderately high end of the overall range when sexes are combined. As noted, very high testosterone is often associated with reduced spatial scores, not enhanced ones, producing an inverted U-shaped function with both very low and very high levels being deleterious for mental rotation and math. One data point possibly relevant to the right tail is available, the finding that physical scientists' digit ratios are closer to female ratios than to male social scientists' ratios (Brosnan, 2006). This is consistent with the suggestion that the middle of the inverted "U" is associated with STEM success.

The confusion in the data is reflected in reviews and commentaries. On the basis of extensive reviews of scientific studies, Kimura (1996, 2000, 2002) has argued for the role of hormones on spatial cognition. Hampson and Moffat (2005) reviewed some of the literature on activation effects of hormones (postnatal fluctuations due to exogenous dosage, circadian or menstrual cycles,

² Limb-related markers for prenatal testosterone include finger ridge counts and the 2D:4D finger length ratio. Such characteristics are markers for prenatal testosterone levels, as gonadal and limb development are genetically linked. The Homeobox genes (*Hoxa* and *Hoxd*) are critical for the development of the urogenital system, limbs, and digits (see Sanders, Sjodin, & deChastelaine, 2002). Hence, prenatal gonadal growth is genetically tied to the development of the hands and feet, supporting the view that distal limb characteristics reflect prenatal testosterone levels. Therefore, performance on sex-dimorphic tasks being associated with limb markers is consistent with a prenatal organizational effect of testosterone on brain development and certain cognitive abilities.

time of day, menopause), arguing that the data were mostly consistent with spatial enhancement by steroid hormones, although the authors also noted inconsistent studies. However, the evidence for prenatal organizational effects is even less consistent than the evidence for activational effects (see Grimshaw et al., 1995). In a collection of chapters on sex differences (Ceci & Williams, 2007), hormone researchers (Kimura; Hines; Berenbaum & Resnick) expressed different views about the role of hormones. Berenbaum and Resnick (2007), for example, argued that differences between females with and without CAH might be due to nonandrogen factors. CAH girls' sex-atypical behavior and superior spatial ability could result from parents treating them like boys because of their masculinized genitalia as well as from reactions of others to their growing competence in stereotypically male activities. Bhasin and colleagues' conclusion about the lack of a dose-dependent relationship between testosterone and visual-spatial memory (for a review, see Bhasin et al., 2005) seems prudent: "Although men, on average, perform better on tests of spatial cognition than women, testosterone replacement has not been consistently shown to improve spatial cognition in hypogonadal men. We did not find changes in spatial cognition at any dose" (Bhasin et al., 2001, p. 1178). On methodological grounds, the Bhasin et al. (2001) findings are compelling (see online Section 4 for a description.) In sum, interpretation of the extensive and complex human literature on the effects of sex hormones on spatial abilities is not straightforward. Positive findings are often offset by studies representing challenges or problems. Like other hormones, androgen's actions on the brain are mediated by different types of receptors, which vary in their proportions in various neural tissues. These differences in receptor type and density can lead to different pathways to cognitive performance that evolved to permit distinct regulation of the effects of sex hormones in different cell and tissue types. Given that mathematics and mental rotation tasks can be performed using different strategies and processes (e.g., feature analysis that is more verbally mediated vs. a strictly visual approach), the density and type of hormone receptors in the regions subserving these processes and strategies would influence performance even with the same nominal task. This would lead to inconsistencies unless the tasks were defined at a more granular level (e.g., specific type of cognitive processes employed to rotate or calculate).

Animal studies show the most pronounced hormone effects (Adkins-Regan, 2005). For instance, male rats have been found to outperform female rats on a water maze (and human males excel on a computer version of this task; Astur, Ortiz, & Sutherland, 1998), and male rat superiority is nullified by castration or by administering testosterone neonatally to female rats (De Vries & Simerley, 2002; Isgor & Sengelaub, 2003; Roof, Zhang, Glasier, & Stein, 1993). However, although animal studies seem most robust, they are also less applicable to humans.³ Lacking are large-scale, representative human studies of individuals at the right tail of the ability distribution that unequivocally demonstrate a relationship. Clinical studies of individuals of unknown representativeness along biological dimensions (e.g., timing and dosing of prenatal and postnatal hormones) are fascinating bases for generating hypotheses but must await randomized experiments and large-scale population studies that report data for right-tail groups. We conclude that hormones account for some of the sex-related variance at the center of the math and spatial distributions but have not been

established as a primary cause of sex differences at the right tail. Inconsistencies exist that need to be reconciled and extended to very-high-ability groups before one can regard them as more than suggestive in assessing the extent to which hormones cause sex asymmetries at the extreme right tail. Thus, the final causal model depicts the evidence for a hormonal basis (through brain functioning and abilities) of the dearth of female scientists as weaker than the evidence for other factors we review.

Broad Contextual Influences

We continue our discussion of the circular elements of the framework with an examination of broad contextual influences—culture, race, social class, and cohort—that moderate cognitive ability formation. Cognitive operations such as memory, visualization, and quantitative reasoning are influenced by the physical, cultural, and emotional context (Ceci, 1996). That sex differences are affected by contextual influences is unsurprising. Studies in this area compare sex differences in performance across groups. As such, they show that context matters, but they do not generally illuminate the mechanisms by which broad contextual factors affect sex differences.

Broad contextual resources and expectations, as well as the proximal environment embedded in that context, are likely to vary according to who is being studied, where they are being studied, and even when they are studied—that is, by culture, social group, and cohort. This variability ripples through the causal chain presented in Figure 1 and is eventually reflected in assessed performance and status. Although we know little about how such differences in the broader context and proximal environment lead to differences down the causal chain that, in turn, lead to developing abilities, we do have data on proxies for those eventual abilities in measures of assessed performance and status and how they vary by context. Below we discuss each of these types of variability in turn along with the resulting observed differences in performance and status. Unlike the hormone studies, all of which derived from non-right-tail samples, the following studies can be divided into those that sampled mean differences and those based on right-tail differences.

Mean Differences in Broad Cultural Expectations and Contextual Influences

Research is available to illustrate the effects of both cultural context and historical context.

Cultural context. Kimura (2007), arguing against cultural explanations of sex differences in STEM professions, suggested that cognitive sex differences "are present across cultures that vary in social pressures to conform to a gender norm. This has been documented for both mathematical reasoning and spatial ability (e.g., Geary & DeSoto, 2001)" (p. 41). However, researchers often place differential emphasis on the same evidence, as we show later, and some have argued that mean sex differences do vary

³ Several strong hormone effects in animals were not found in humans, such as the failure of DHEA (dehydroepiandrosterone), a weak androgen secreted by the adrenals and converted to testosterone, to predict cognitive declines in older persons, despite strong effects in animals (Hampson & Moffat, 2005).

across cultures. For example, Schratz (1978) reported that among African American and Hispanic high school students, girls scored higher than boys. Brandon, Newton, and Hammond (1987) found that Hawaiian girls in Grades 4–10 scored higher than boys, and the female advantage was larger among Hawaiian, Japanese, and Filipino students than among Caucasians. African American girls match or outscore African American boys on every assessment (American Association of University Women, 1998), and effect sizes for quantitative ability sex differences are smaller among minority students in the United States (Friedman, 1989). Additionally, in Iceland, high school girls are superior to boys on spatially loaded subtests (Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005).

In one transnational comparison, U.S. 5th-grade boys' mean score on spatially loaded subtests was 13.1, whereas girls' mean score was 12.4. However, 5th-grade girls from Japan and Taiwan outscored U.S. boys dramatically, with mean scores of 18.1 and 16.1, respectively (Lummis & Stevenson, 1990). Guiso, Ferdinando, Sapienza, and Zingales (2008) reported enormous transnational differences in 15-year-olds' math scores on the 2003 Program for International Student Assessment, which was given to more than a quarter million students in Organisation for Economic Co-operation and Development (OECD) countries, ranging from a mean male advantage of 22.6 points (Turkey) to a mean female advantage of 14.5 points (Iceland). Beller and Gafni (1996), analyzing data from national samples of 9-year-olds, found the effect size for sex differences in math to range from -0.06 in Ireland (female superiority) to $+0.28$ in Korea, and Penner (in press) reported a similar wide range of effect sizes for high school students in 22 countries, ranging from highs of $.63$ (Netherlands), $.62$ (Denmark), and $.60$ (Norway) to lows of $.05$ (Hungary) and $.13$ (United States). Penner argued that the large cross-national variation in sex differences in the Third International Mathematics and Science Study (TIMSS) suggests that culture rather than biology is involved because the observed patterns are not otherwise explicable. For instance, the magnitude of sex differences remains the same throughout all points in the distribution for about half of the countries (favoring males), but several countries' sex differences are larger at either the left tail (Netherlands, Lithuania) or the right tail (e.g., Sweden) of the distribution; additionally, for some countries, girls do as well as or better than boys at the left tail but worse at the right tail (United States, Hungary). Finally, in some countries sex differences are most pronounced in the middle of the distribution (Russia, Austria).

Further calling into question claims of culture-independent sex differences is a lack of consistent sex differences between kindergarten and 3rd grade (Lachance & Mazzocco, 2006). In addition, trends are not always found in the same direction: for example, in Korea the effect size decreases with age (from 9 to 13 years), whereas in Ireland and Spain it increases with age. In the United States, few gender differences in mathematics are usually found among primary school children (Friedman, 1989; Lachance & Mazzocco, 2006). (For reviews showing no systematic gender gaps until early adolescence when boys begin to excel, see Fox and Cohn, 1980; Friedman, 1989; and Kimball, 1989, although Rathbun, West, & Germino-Hausken, 2004, analyzed the Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999, and reported sex differences in math achievement as early as 1st grade, and Penner and Paret, 2008, found sex differences in math ability

among kindergartners, prior to curricular streaming.) Wise, Steel, and MacDonald (1979) documented, with a nationally representative sample of Project Talent participants tested in the early 1960s, that U.S. sex differences in math accelerated greatly during high school in that era. Male superiority among adolescents does not always appear, however, with most studies finding no differences on algebra skills but boys outperforming girls on three-dimensional solid geometry (Kimball, 1989). Finally, Mullis, Martin, Fierros, and Goldberg (2000) reported small and inconsistent sex differences on TIMSS at Grade 8 but consistent male superiority by Grade 12, particularly among the highest quartile of scorers. Ultimately, the inconsistency about how early mean sex differences occur depends on the tests used, specifically, their content and their difficulty level. Hyde et al. (2008) reported findings of math achievement for over 7 million U.S. students in 2nd through 11th grade on the No Child Left Behind tests. There were no substantial sex differences at any grade, with $d_s < .10$ across the board, though variance ratios were larger for boys, a topic we discuss in the section on right-tail differences.

Female students in some nations outscore U.S. and Canadian male students on mathematics tests (Valian, 1998), sometimes by greater margins than those by which U.S. males outperform U.S. females. For example, Taiwanese and Japanese females vastly exceed males in the United States: 8th-grade Japanese girls scored 569, Japanese boys scored 571, and American girls and boys scored 502 and 507, respectively (Mullis, Martin, & Foy, 2005). Girls in Singapore scored 1 *SD* higher than Americans (611 and 601 for Singapore girls and boys; Valian, 2007). Additionally, there are far smaller sex differences between 5th-grade boys and girls in the United States ($d = 0.18$) than between U.S. boys and Japanese boys ($d = 1.42$) (Hyde & Linn, 2006). Penner (in press) reported d_s for within-country sex differences in mathematics ranging from $.05$ to $.63$ across 22 countries. The magnitude of differences across countries, however, is at least as large as the size of sex differences within the United States.

Within the United States, sex differences are inconsistent across cultural/ethnic and socioeconomic status (SES) groups. In one meta-analysis, the gender gap in math was larger for Whites ($d = 0.13$) than for Blacks ($d = -0.02$), Hispanics ($d = 0.00$), and Asian Americans ($d = -0.09$) (values from a meta-analysis by Hyde, Fennema, & Lamon, 1990). However, Penner and Paret (2008) used quantile regression to focus on specific deciles in the distribution and found that, in contrast to OLS regression procedures, the largest sex differences were found for Asian American children, starting as early as kindergarten. Among low-SES 3rd graders, Levine et al. (2005) found that girls and boys did not differ notably in spatial skills, and middle-class girls were at least as proficient in these skills as were lower class boys, although the sample size and design did not afford an optimal test of the interaction and the effect has failed to show up in a large Project Talent reanalysis (Wai, Lubinski, & Benbow, in press), which admittedly came from a different era, tested different-aged children, and used different tests, making comparisons with Levine et al. problematic. Casey, Andrews, Schindler, Kersh, and Samper (2008) found sizable SES differences in two- and three-dimensional mental rotation among kindergartners: 6.95 correct transformations out of 10 for the higher SES children versus 4.35 for the lower SES ones in the baseline or control groups. Using a nationally representative sample, Penner and Paret (2008) reported

that the largest sex differences found in kindergarten through 5th grade at the right tail occurred among children whose parents had advanced degrees (42.4 vs. 38.7 for boys and girls, respectively). In juxtaposition, one Australian study of four high schools found that the magnitude of sex differences in mathematics was far greater among low-SES than among middle-class students (Lamb, 1996). In general, sociologists have reported that although sex differences in mathematics exist among high-SES older students, the size of the gender gap is greatest among low-SES older students, and disadvantaged males and African American males score lowest of all, thus reversing the usual male superiority (American Association of University Women, 1998; Catsambis, 2005). Thus, the size of the minority effect is smaller (Friedman, 1989), although Hyde et al.'s (2008) analysis revealed small mean sex differences for all racial groups.

Taken together, these cultural and sociodemographic differences suggest that culture may play a major, though poorly understood, role in creating proximal differences that lead to differences in STEM fields. In the Guiso et al. (2008) analysis, cultures that valued egalitarianism exhibited narrower sex differences in math achievement, a finding independent of genetic differences between cultures, thus reinforcing the cultural mechanisms themselves. As we describe later, others have not found such effects (Charles & Bradley, 2006).

Historical context. In contrast to sex differences amongst those currently at the peak of their career—born in the 1940s and 1950s—are females who have grown up in today's more egalitarian world. If there are differences between these cohorts, they are likely due to broad contextual factors. This is important because a lot of data on which conclusions have been drawn about sex differences in mathematics and science are 20 or more years old, with some being 50 years old. For example, some of the most striking findings in both Wise et al. (1979) and Hedges and Nowell (1995) are based on Project Talent, a study of children born around the end of World War II and assessed during adolescence in the late 1950s. Similarly, the original cohort of SMPY data was analyzed in the early 1970s (Stanley et al., 1974). Would the same results be found today?⁴ Hyde, Fennema, and Lamon (1990) reported substantial declines in effect sizes over time, a finding also found by others (Friedman, 1989; Hyde, 1990; Hyde, Fennema, Ryan, Frost, & Hopp, 1990). Friedman (1989) reported a correlation of .75 between year of publication and the magnitude of effect size for studies conducted in the 1980s. Friedman (1989) made this point with a comparative meta-analysis of the data used in Hyde's (1981) own meta-analysis of the high school studies cited by Maccoby and Jacklin (1974); the .43 median effect size for those older studies shrank to .24 for post-1974 studies. In their meta-analysis of spatial ability, however, Linn and Petersen (1985) found no changes in the size of sex differences between the ages of 10 and 60 years, suggesting that historical time may not be a moderator for all variables. Similarly, Kimura (2007) argued that cognitive sex differences are mostly unchanged in magnitude over the past 30–40 years, a period in which women's roles and access to higher education have changed substantially.

A striking example of a secular change in sex differences was provided by Shayer, Ginsberg, and Coe (2007), who analyzed the Science Reasoning Test II, used in Britain. In Table 3 of the supplemental materials online, we highlight three points revealed by Shayer et al.: (a) In 1975 there was a substantial male advantage

in the mean scores; (b) this advantage disappeared by 2004; (c) although both boys and girls have shown large drops in performance during this epoch, the drop was greatest for boys. Moreover, although males have outperformed females on the SAT-M for more than 30 years ($d = 0.39$), the magnitude has shrunk from 40 points to 33 points (Royer & Garofoli, 2005). Further, in a meta-analysis of 286 effect sizes on sex differences in spatial abilities, Voyer, Boyer, and Bryden (1995) found most differences declining but some, such as mental rotation, increasing (see also Nordvik & Amponsah) in contrast to claims that it is decreasing (e.g., Feingold, 1988). Of course, change in either direction underscores the malleability of sex differences. Finally, for mathematics, Hyde, Fennema, and Lamon (1990) reported a mean effect size for studies published prior to 1973 of $d = 0.31$ (boys superior) but only 0.14 for studies since 1974, a decline also found by others (Friedman, 1989).

In some cases, changes over time swamp remaining sex differences, calling into question their meaningfulness. For example, Freeman (2004) reported that the percentage of girls taking calculus courses in U.S. high schools rose from 4% in 1982 to 11% in 2000, whereas the percentage of boys enrolled in such courses rose from 6% to 12% over this period. Although the remaining 1% difference is significant on account of the large sample, it is trivial compared with the magnitude of changes over time, rendering explanations in terms of stable biological sex differences questionable.

Right-Tail Differences in Broad Contextual Influences

Sex differences at the right tail attributable to broad cultural influences are also inconsistent: Males are overrepresented, but the degree depends on the measure used and the culture and cohort studied. For example, some researchers claim that sex differences among adolescents are smaller today than they were a generation ago (Feingold, 1992; see Friedman, 1989, for a meta-analysis), whereas others (e.g., Nordvik & Amponsah, 1998) argue otherwise.

Cultural context. Although in the United States, males outnumber females at the extremes of mathematics and mental-rotation ability, there is inconsistency in the ratios of males to females across cultural contexts. In some countries, the ratios are smaller than in others, with some showing nonexistent differences (Feingold, 1994), whereas in other countries, females excel over males in both the top 5% and the top 1% (Guiso et al., 2008). In the latest analyses of U.S. data, Hyde et al. (2008) reported that although sex differences at the mean have disappeared in mathematics, White males outnumber females at the top 1% by a ratio of 2.06:1, although among Asian Americans, females slightly outnumber males 0.96:1. However, with regard to one type of mental-rotation ability—three-dimensional rotations—the male advantage is much more consistent (Linn & Petersen, 1985). For example,

⁴ Only in the 1970s did Harvard University / Radcliffe College admit men and women on an equal access basis; now half of undergraduates are female (Harvard University, 2005). The University of Cambridge did not admit women to its traditionally all-male colleges until 1972, although they were granted membership in 1947 and admitted to two women-only colleges with partial university rights previously. Enrollment of women increased from 11% in 1968 to 49% in 2004 (Cambridge University, 2005).

Nordvik and Amponsah (1998) reported $d = 0.85$ among Norwegian university science majors whose high school concentration had been physics and mathematics.

Charles and Bradley (2006) analyzed international data on university degrees awarded in 2001, including those in math-intensive fields. Women predominate in traditionally female-typed fields, such as education and health, whereas men predominate in stereotypically masculine fields, including all of the math-intensive STEM fields. For instance, in computer science, females are underrepresented in all 21 OECD countries; however, their degree of underrepresentation varies greatly across countries. In Turkey, men are overrepresented among computer science graduates by a factor of only 1.79, whereas in the Czech Republic they are overrepresented 3.59 times more, by 6.42. In the United States, the “male overrepresentation factor” is 2.10 (as noted; the d for Advanced Placement [AP] Computer Science AB scores has shrunk dramatically between 1984 and 1996), and in the United Kingdom the figure is 3.10. Charles and Bradley reported no correlation between females’ math achievement and entry into stereotypically male fields ($r = -.04$ between females’ achievement in 8th-grade math and their representation in computer science). Interestingly, the most economically developed countries do not produce the greatest ratios of women in computer science. Nor is there a strong correlation with number of women in the workforce or in high-status jobs or in higher education. Penner’s (in press) data revealed that ideological differences concerning the importance of home and children for women do not track well with transnational variation in sex differences in math. National beliefs in equal opportunity also are not a consistent predictor of female entry into male fields: None of the highest scoring nations—Turkey, South Korea, and Ireland—in Charles and Bradley’s study are known for gender-egalitarian attitudes or practices. When these authors calculated a correlation between women employed in computer science and the percentage of the national population disagreeing with the statement that “a university education is more important for a boy than for a girl,” there was an inverse association between these two variables ($r = -.47$). However, when Guiso et al. (2008) examined each culture’s endorsement of anti-egalitarian gender views (e.g., giving scarce jobs to men over women), they found it to be a powerful predictor of the sex gap in math.

Historical context. In a meta-analysis, Linn and Hyde (1989) concluded, “Gender differences in spatial ability are heterogeneous and declining. Differences that remain are responsive to training” (p. 19). B. J. Becker and Hedges (1984) regressed effect sizes for quantitative ability on year of publication, to arrive at a positive yearly coefficient of 0.01, suggesting that in each decade the effect size would decrease by 0.10. Notwithstanding these reports, others have argued that differences have not been declining, as we detail below. As was the case for mean sex differences on the SAT-M, which have not changed during the past 35 years (Halpern et al., 2007), many have found that differences at the right tail are fairly stable. In Stumpf and Stanley’s (1998) analyses, males had an advantage at the right tail in virtually all math-intensive areas of AP exams, an observation that has remained fairly stable during the 1980s and the 1990s. Thus, their trend analyses show, for the most part, good consistency in the effect sizes over time for sex differences in the right tail of the SAT and AP examinations, with some notable exceptions (i.e., d for computer science has shrunk dramatically between 1982 and 1996). Similarly, Lohman and

Lakin (in press) found that the proportion of high-scoring males was relatively constant across levels and forms on the Cognitive Abilities Test, Quantitative Battery, spanning national U.S. samples from 1984 to 2000. Strand et al. (2006) had a similar finding of male overrepresentation at the right tail of this test for a national sample of 320,000 11-year-olds from the United Kingdom. As can be seen in Figure 2, boys are significantly more likely to score at stanine 9 (≥ 2 SDs above the mean) as well as at the bottom stanine in quantitative ability, and this has been fairly stable over more than a 16-year period. Hedges and Nowell’s (1995) analyses of six national data sets mentioned earlier also showed consistency in the sex ratios over a 32-year period. In contrast to these demonstrations of impressive consistency, there are a number of examples of inconsistency in the gender ratio at the right tail, which are also based on large national samples or meta-analyses (B. J. Becker & Hedges, 1984; Friedman, 1989; Hyde, Fennema, & Lamon, 1990; Linn & Hyde, 1989). In Lohman and Lakin’s (in press) data showing impressive consistency in male advantage at the right tail on many cognitive measures, females appear to have narrowed the gap at the right tail on the Cognitive Abilities Test Nonverbal Battery (Figure Classification, Figure Analogies, Figure Analysis) over the same time period: 9th stanine female-to-male ratios changed from 0.72 in 1984, to 0.83 in 1992, to 0.87 in 2000. Thus, again, findings regarding stability over time are mixed.

The resolution to the question of whether sex differences in math and spatial ability have been consistent or narrowing over time requires consideration of a number of factors, many of which are discussed later. Factors such as (a) the composition of the tests (consistency is more likely when the test content has remained consistent over time, as changes in its composition can lead to shifts in the proportion of problems that favor each sex), (b) changes in the proportions of each sex taking the test, because as one group becomes more numerous in its participation, its scores go down (and there have been increases in female students taking some tests such as the SAT [Nie & Golde, 2008]), (c) changes in analytic approaches, for example, extreme-tail-sensitive approaches versus OLS (see Penner, 2005), and (d) changes in the type and number of math courses each sex has taken (which has occurred; Hyde et al., 2008). (See Stumpf & Stanley, 1998, for a

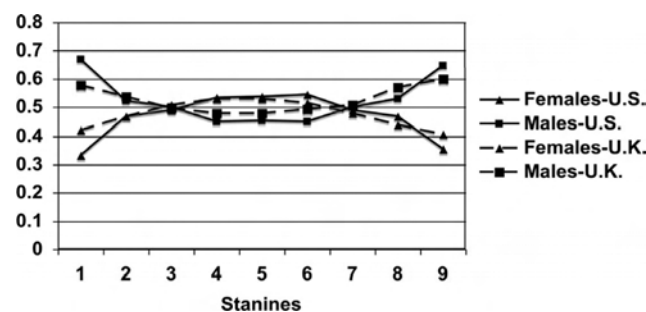


Figure 2. Male-to-female proportions at each stanine ($M = 5$, $SD = 2$) based on data from over 320,000 students in the United States and 320,000 10- to 11-year-olds in the United Kingdom. Data are from “Consistencies in Sex Differences on the Cognitive Abilities Test Across Countries, Grades, Test Forms, and Cohorts” by D. Lohman and J. Lakin in press, *British Journal of Educational Psychology*. Adapted with permission of the authors.

discussion of additional factors that may be related to inconsistency over time.)

In addition to their analysis of mean differences, Shayer et al. (2007) also examined differences in scores on the Volume and Heaviness test at the right tail of the distribution. The authors found substantial changes (i.e., declines) over time in average scores, but the change at the extreme right tail was much larger. Virtually no child in 2003 scored in the top 10% of the range by 1976 standards. This finding demonstrates the importance of specifying the relevant population sample and epoch.

There are many other cases of change over time, and we describe some to illuminate the causes of current sex differences, as opposed to understanding the phenomenon in other historical periods. As Figure 3 shows, in the United States, the ratio of boys-to-girls performing in the top 1% on measures of advanced mathematics has declined consistently. The sex ratio of adolescents who scored > 700 on the SAT-M at age 13 years, a feat achieved by only 1 in 10,000 students, has shrunk from 13:1 in 1983, to 5.7:1 in 1994, to 4:1 in 1997, to 2.8:1 by 2005 (Julian Stanley, quoted by R. Monastersky, 2005; see also Gates, 2006b). If sex differences were primarily biologically driven, one might expect greater consistency across cultures and time (Spelke, 2005).

In addition, Stumpf and Stanley (1996) found that women narrowed the performance gap in AP Computer Science AB courses between 1984 and 1996 ($d_s = .59$ and $.16$, respectively); the number of women scoring greater than 700 on the College Board's Mathematics II Achievement test increased by 150%, and the number of women with high scores on the AP Physics Test increased by 142% as a result of increased female participation in these male-dominated subjects. If such scores reflect the ability necessary for eligibility in STEM fields, then suddenly there are many more eligible women.

Cohort differences have also been found on measures of career discrimination among professionals in STEM fields. Ginther's work (2001) found significantly lower odds of women on the tenure track in scientific disciplines being promoted, controlling for demographics, productivity, and other factors, in the 1972–1979 cohort of those with PhD degrees, but no significant sex difference in the 1980–1989 cohort. Although many of these findings imply sociocultural rather than biological causation, they do not prove that there is not a biological component to male–female differences in extreme mathematics ability; they merely

show that, irrespective of any biological influence, there has been substantial environmentally induced variance. These data call into question predictions for the future that are based on data from individuals who are now reaching retirement age and who grew up without the multitude of female role models in government, athletics, management, and science that girls have today.

In short, the results of the various cohort and historical trends indicate that the performance gap between males and females in advanced math—both at the middle and at the right tail—fluctuates in response to various factors, sometimes due to males getting worse, sometimes due to females getting better, and sometimes due to changes in the composition of the test or measure.

Proximal Environment

Continuing around the causal model in Figure 1, broad contextual attributes feed into the proximal environment. Proximal sociocultural influences have received the attention of researchers for over 50 years. Of greatest relevance are school, peer, and parental influences, and the environmental differences they create. Many studies in this area investigate the mechanisms by which contextual factors may influence sex differences.

In Table 4 of the supplemental materials online, there are several studies showing a sex-differentiated pattern of teacher and parent behavior, although some of the evidence begs for replication and further exploration, as well as updating. Nearly all of this evidence comes from unselected samples, not right-tail ones. Kelly (1988) meta-analyzed over 80 studies of teacher–student interactions and reported that boys drew more attention from teachers. Although girls received less criticism, they also received less instruction even though they raised their hands more often. The magnitude of these differences was not great, and they held regardless of the gender of the teacher. Beaman, Wheldall, and Kemp (2006) updated this analysis.

Mean Differences

First, there is evidence of differential treatment of girls in high school mathematics classrooms on many measures (see above), although some of this is old data. J. R. Becker (1981) gathered frequency counts of teacher–student interactions coupled with in-depth observational data. She studied 10 high school geometry teachers in the late 1970s, an era corresponding to the generation now reaching the top of STEM fields. Seven of the 10 teachers were women. Girls were largely ignored: Teachers provided boys with more formal and informal reward and support, and a good affective environment in which to learn, and male students answered more open as well as direct questions, process questions, and callouts, even though there were no differences in student-initiated interactions. Despite a lack of difference in student-initiated interactions, 63% of the teacher-initiated academic contacts were with boys, whereas “females, relatively speaking, were treated with benign neglect” (pp. 50–51).

Sex differences have also been found in the attitudes and perceptions of parents and teachers: Sixth-grade girls' mathematics abilities are underestimated by their mothers, whereas boys' abilities are overestimated (Frome & Eccles, 1998); in addition, abilities of high school girls are viewed less favorably than boys' abilities by parents and teachers (Hyde, Fennema, Ryan, et al.,

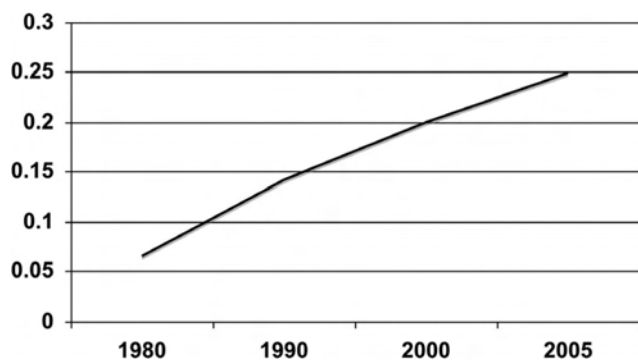


Figure 3. Changes in female representation at the extreme right tail of mathematics score distribution for 13-year-olds.

1990). Further, J. E. Jacobs and Eccles (1985) found that children are influenced more by their mothers' perceptions than by actual grades when developing opinions of their own abilities. Parents provided more mathematics-supportive toys and opportunities for sons than daughters and held higher perceptions of sons' abilities, and children's past perceptions were found to predict later interests and GPA (J. E. Jacobs, Davis-Kean, Bleeker, Eccles, & Malanchuk, 2005; see also Davis-Kean et al., 2007). It is easy to imagine how such differential treatment at school and home might have led to differences in performance, such as boys excelling on far-transfer problems (Gallagher & DeLisi, 1994; Gallagher, Levin, & Cahalan, 2002; see Gallagher & Kaufman, 2005). Ravitch (1998) criticized claims that sex differences in scientific careers are the result of teachers showering more attention and praise on boys, or of higher self-esteem among boys, arguing that far from failing girls, the schools are doing a good job in closing gender gaps in mathematics and science. With the sole exception of high school physics, where 27% of boys versus 22% of girls enrolled, girls were taking as many courses in mathematics and science as boys, and this state of affairs has been true at least since the late 1980s. For example, female high school graduates in 1990 had higher enrollments than boys in 1st- and 2nd-year algebra and in geometry; among the graduates of 1994, there were few sex differences in precalculus, AP calculus, statistics, trigonometry, and a host of science courses; girls were more likely to enroll in chemistry and biology than were boys, and 43% of girls took a rigorous college-preparatory program in 1994 versus only 35% of boys. These observations do not support the claim that teacher treatment of girls has led to lack of motivation to take science and math courses.

Moreover, the findings are correlational, and myriad factors could be at work to explain teacher and parent behavior that seems to favor boys, such as greater attention directed toward them to control obstreperous behavior or sex differences in talking to parents of the same sex. In Muller's (1998) analyses of a large representative database of public school 8th–12th graders, little evidence was found that parental behavior of either gender contributes to sex differences in mathematics achievement, and Catsambis (1994, 2005) and others have shown that girls display less interest in mathematics even when their ability is comparable to boys'. Finally, as J. E. Jacobs et al. (2005) noted, in their regression predicting a child's interest in mathematics, neither gender nor mathematics activities predicted interest. Moreover, the cohort difference they found (younger cohort more interested in mathematics) does not correspond to national changes in the number of women majoring in mathematics, as the younger cohort reached college age in the mid- to late 1990s, and yet by the mid-1990s, gender differences in mathematics were already negligible among both middle school and high school students; thus, their putatively lower interest in mathematics was nevertheless associated with a rise in majors (Catsambis, 2005).

Right-Tail Differences

The findings at the extreme right tail do not indicate a strong causal role for parental encouragement in math. In analyzing data from the TIMSS project, Penner (in press) found that the students who claimed that their mathematics achievement was important to their parents actually exhibited larger sex differences at the right tail than at the left tail: Girls in the top 5% of mathematics

achievement score 3% worse than boys among those who say that mathematics achievement is unimportant to their parents, but girls in this group score 6% worse than boys among those who say their achievement in mathematics is important to their parents. (Mathematics scores at the left tail of the mathematics distribution—the bottom 5%—are enhanced among those who report that their parents value mathematics achievement.)

The hypothesis that boys would report greater parental encouragement in mathematics than girls was also not supported by the data collected on two groups of talented adolescents. One group scored in the top 5% of either the SAT-M or the SAT-V before age 13 years (only 1 in 10,000 children of this age score this highly). The second group of children scored above the 97th percentile on a school achievement test, making them talented but not nearly to the degree of the 1 in 10,000 group. Raymond and Benbow (1986) collected questionnaire data on both groups (and their parents) to ascertain whether parental encouragement was associated with extreme talent in math or verbal domains and whether sex typing of math as a male domain correlated with math scores. The researchers found that little variation in maternal or paternal support was accounted for by child gender. Intellectually talented boys did not perceive greater mathematical encouragement than did intellectually talented girls. Instead, mathematical encouragement varied by the child's actual mathematical ability (1 in 10,000 vs. less talented). A measure of sex typing did not correlate with SAT-M scores or SAT-V scores in either the 1 in 10,000 group or the less talented group.

Although there may be some evidence for an effect in the middle of the distribution, sex differences at the right tail do not strongly relate to differential parental and teacher attitudes. Therefore, in our final model, although proximal processes are assumed to mediate the effects of broad contextual factors on motivation and behavior, they are downplayed as a primary causal factor in women's current underrepresentation in math-intensive fields, because the mechanism by which they act is unclear. On both theoretical and empirical grounds, the proximal sociocultural processes are conceptualized as the engines that drive developmental outcomes, that bring genetic differences to fruition (or not, depending on their adequacy), and that determine how much of a child's cognitive potential gets realized. Bronfenbrenner and Ceci (1994) summarized the empirical and theoretical evidence that proximal sociocultural processes underpin broad cultural and social class influences, with cultural–contextual variables predicting cognitive outcomes because they are proxies for proximal processes (see online Section 5).

Motivation, Attitudes, and Interests

The proximal sociocultural environment feeds, in turn, into motivation/attitudes/interests and activities (Figure 1). Potentially causal sex differences also occur in this component. For example, differences in motivation/interests and activities may lead directly to differences in status, through life choices, and indirectly, through differences in productivity and stereotype threat affecting performance, as well as through the circular chain of effects on brain development and abilities.

Mean Differences

Baron-Cohen (2007) has argued that females are born with an innate motivation to orient toward people, whereas males have an orientation toward objects, which leads the sexes down differing paths of interests. As evidence, Connellan, Baron-Cohen, Wheelwright, Ba'tki, and Ahluwalia (2001) found that male newborns looked longer at an object but that female newborns looked longer at a person. Spelke (2005), however, criticized this view, suggesting "male and female infants are equally interested in people and objects" (p. 951); she views the Connellan et al. results as an aberration amidst hundreds of other infancy studies, criticizing it on methodological grounds.⁵

Some have claimed that mean differences in the ways the sexes spend their time (e.g., playing with Legos vs. dolls; Bornstein, Haynes, Pascual, Painter, & Galperin, 1999), toy purchases (Davis-Kean et al., 2007; Goldstein, 1994), and computer video game experience (Quaiser-Pohl, Geiser, & Lehmann, 2006; Terlecki & Newcombe, 2005) may contribute to differences in mathematical and spatial abilities. Levine, Huttenlocher, Taylor, and Langrock (1999) opined that girls' lack of outdoor exploratory play leads to their lower spatial cognition scores, although low-SES children do not show this gender effect, and White middle-class girls are superior to low-SES boys who would seem to engage in more unsupervised outdoor play. Hence, it is unclear whether outdoor play occupies a causal role in mental rotation and other spatial and mathematics skill differences. Jahoda (1979) found that Scottish 7- to 11-year-olds' ability to replicate structures with cubes was related to their mental-rotation ability (see also Kersh, Casey, & Mercer Young, 2008, for related evidence), and Brosnan (1998) reported a correlation between 9-year-old children's replication of a Lego structure and scores on a mental rotation task. Wolfgang, Stannard, and Jones (2001) demonstrated a relationship between block playing and mathematical ability.

A recent, well-controlled study found that merely playing action video games can narrow gender differences in mental rotation (Feng, Spence, & Pratt, 2007). Terlecki, Newcombe, and Little (2008) found mental-rotation performance of college women who had one semester of video game training (Tetris) to be only marginally lower than that of men who had no training (but were given repeated testing on rotation ability), and this was true regardless of pre-existing spatial experience. Perhaps with an even longer intervention, more complete gender convergence would occur. Sorby and her colleagues designed a training program to aid female engineering students.⁶ Female students score lower than male students on visual cognition scores both before and after participation in an engineering graphics course (Sorby, 2001, 2005). Although both sexes made gains in spatial ability over the semester, women's posttest scores were typically lower than men's pretest scores. In response, Sorby and her colleagues designed a course in spatial reasoning that included teaching and practice of complex spatial rotations. She demonstrated that sex differences in mental rotation can be closed among math-intensive STEM majors. By identifying male and female engineering majors who failed a spatial pretest that included mental rotation along one or more axes, Sorby exposed as many such students as possible to a visual cognition course in their freshman year. (Sorby also reported data for a randomly selected cohort to minimize selection bias.) Female students who took the course achieved significantly

higher spatial posttest scores, had higher GPAs several years later by about 0.3 of a grade (e.g., 2.7 to 3.0), and were more likely to remain engineering majors (63.6% vs. 53.1% of control group female students who had not taken the course; comparable figures for male students were 69.2% vs. 62.5%; the difference was not significant because of the small sample size.) Spatial ability has been singled out as a significant predictor of sex differences in other fields as well, such as medicine (Tendick et al., 2000), chemistry (Carter, LaRussa, & Bodner, 1987; Pribyl & Bodner, 1987), and mathematics (Tartre, 1990). Thus, spatial ability can exert direct and indirect effects, the former in professions that are spatially loaded, such as radiology, angular laparoscopy, engineering graphics, and *n*-dimensional projections in chemistry, and the latter in its role in certain types of mathematics, which in turn may be important in math-intensive careers.

Kersh et al. (2008) reviewed several studies of the relationship between block play and spatial ability, concluding that "these studies suggest a link between block building activities and the development of spatial competency in boys and girls. Nonetheless, more research is needed to understand and document the mechanisms involved in this relationship" (p. 11). Notwithstanding the demonstrated benefits on females' performance of playing spatial games, including occasional demonstrations that the entire gender gap can be closed as a result of such experiences, two meta-analyses found that men and women improve in parallel in response to practice and training, so that gender differences remain constant in size (Baenninger & Newcombe, 1989; Marulis, Warren, Uttal, & Newcombe, 2005, cited in Terlecki et al., 2008). Table 5 of the supplemental materials online summarizes some of the most prominent examples of this work. Finally, even if spatial activities enhance girls' spatial ability, better evidence is needed than is presently available showing that spatial ability underpins sex differences in advanced mathematics or the types of activities needed for success in certain professions.

These studies, particularly Kersh et al.'s (2008) investigation, provide some empirical support for the relationship between early block play and subsequent spatial and mathematical abilities, but other studies often have been inconsistent and gloss over social class differences. For instance, it seems odd to see team sports credited with developing mathematics ability in contrast to activities classified as low complexity, such as chess and imaginary play, as the stereotype of the sports-mad "jock" is not usually associated with being a "mathematics/science whiz," and games such as chess and imaginary play are thought to promote cognitive development (e.g., Ferguson, 2006; Fischer, 2006; Keach, 2003).

⁵ Nettle (2007) validated this distinction among adults, showing that two dimensions are reliable correlates of personality traits. He opined that sex differences in empathizing and systemizing can explain men's greater interest in science. Our own view of this controversy between Spelke and Baron-Cohen is that the putative biological roots of infant differences need not be the basis of the adult sex differences, as adult agreeableness and sociability (which Nettle finds related to empathizing) can also be influenced by early experiences. Resolving this issue would require longitudinal analysis. Having stated this, we note that there is ample evidence of gender differences in people versus object orientation among adolescents and adults (see Lippa, 2005, for review).

⁶ We are indebted to Nora S. Newcombe for bringing this work to our attention.

Hence, it is unclear whether play occupies a causal role in mental rotation and other spatial and mathematics skill differences.

Right-Tail Differences

The motivational research at the right tail deals with professionals in academia rather than with unselected samples of children and concerns how men and women spend their time building their careers and families.

Lubinski (2004) reported the amount of time that nearly 2,000 33-year-olds, who during their adolescence were in the top 1% of quantitative ability, devote to their current jobs and the amount of time they would devote to their ideal jobs. Roughly twice as many high-aptitude men reported working at their jobs > 50 hr per week at age 33 years, and three times more women reported working <40 hr. Lubinski reported a similar sex difference in a study of nearly 10,000 high-aptitude math scorers, leading him to suggest that “one only needs to imagine the differences in research productivity likely to accrue over a 5- to 10-year interval between two faculty members working 45- versus 65-hour weeks (other things being equal) to understand the possible impact” (Lubinski & Benbow, 2007, pp. 90–91). These findings have been replicated on two independent cohorts that are even more exceptional (Lubinski et al., 2006, p. 198): top mathematics/science graduate students identified in their mid-20s and tracked for 10 years ($N \approx 700$) and profoundly gifted participants identified before age 13 and tracked for 20 years ($N \approx 400$).

Mason and Goulden’s (2004) analysis of a nationally representative sample of individuals with PhD degrees, as well as an analysis of 4,459 tenure-track faculty at the nine University of California campuses, reveals that factors affecting women’s success and satisfaction spill over into the family, or the reverse, the family spills over into the job: Although 66% of faculty fathers report working > 60 hr per week at their careers, only 50% of faculty mothers do; mothers report working more hours per week than fathers when combined across career, housework, and caregiving: an average of 101 hr for women with children versus 88 hr for men with children. (Childless men and women both work an average of 78 hr across all domains.) Faculty mothers work 4 hr less at their academic jobs than do childless women faculty (J. A. Jacobs & Winslow, 2004, p. 117). (The fact that childless men and women both report working 78 hr gives credibility to the sex-neutrality of self-report measures.)

Some research indicates that sex differences in productivity among senior scientists disappear when the type of institution and available resources are taken into account (Xie & Shauman, 2003). But what are the origins of the resource differences in Xie and Schauman’s data? One possibility is that current institutional resources of senior professors are partly the consequence of prior productivity differences between men and women. If this is true, then adjusting productivity for level of current resources will underestimate the role of past productivity. (Repeatedly, one confronts the ambiguity of correlational data, and this is a case in point.) Consistent with this suggestion, in his large-scale survey, Davis (in press) reported that sex differences in hours worked and productivity were observed among postdocs after controlling for institution, family structure, and levels of supervision and training. Davis reported that male postdocs worked, on average, 52 hr per week and that female postdocs worked 49.4 hr, a male advantage

that also shows up in the National Science Foundation’s Survey of Doctoral Recipients for women with children under age 18 (Hoffer & Grigorian, 2005). The male standard deviation for hours worked was greater (12.1 vs. 11.1), indicating that more men probably worked the highest numbers of hours per week, accounting for their greater productivity. Consider the following: sole-authored peer-reviewed articles, 0.3 versus 0.2 for men and women, respectively; coauthored peer-reviewed articles as first author, 1.7 versus 1.2; coauthored peer-reviewed articles as non-first author, 1.7 versus 1.3; non-peer-reviewed papers, 1.7 versus 1.3. However, the number of chapters written was identical, 0.2 versus 0.2, as were patents and conference papers. Female postdocs submitted more grant proposals on which they were the principal investigator than did male postdocs (1.1 versus 0.8), whereas men submitted slightly more as co-principal investigators, 0.7 versus 0.6, for men and women, respectively, which suggests that some of the difference in publication rates may be the result of different resource allocation strategies. (There were no reported sex differences in teaching or service responsibilities.) Although one can always question the validity of self-report data, such as the Sigma Xi postdoctoral survey by Davis (e.g., perhaps men overestimate their accomplishments), it is probably not a systematic bias because women report excelling at grants and other accomplishments.

Relatedly, Lubinski, Benbow, Shea, Eftekhari-Sanjani, and Halvorson (2001) reported that although 77% of the female graduate students and 81% of male graduate students expressed the view that a full-time career was “important” or “extremely important,” significant sex differences subsequently emerged in the importance of having a part-time career for some time period and having a part-time career always.

In regard to working part-time for a limited period of time, about one third (31%) of female graduate students responded that this option was “important” or “extremely important,” compared with 9% of male graduate students. For having a part-time career always, the respective proportions were 19% for females and 9% for males (Lubinski et al., 2001, p. 312).

These findings accord with evidence from a recent national survey in Britain (Hakim, 2006) showing that more women prefer home-centered lifestyles than do men, and conversely that far more men prefer committed work-centered lifestyles than do women.

However, if women are not succeeding in math-intensive fields as a result of prioritizing family over career, then it is of interest to ascertain whether they are similarly prioritizing in such fields as medicine, law, veterinary medicine, dentistry, biology, and psychology. Being a physician or a veterinarian makes inroads into family life every bit as great as holding math-intensive careers, with 48-hr shifts and an inability to stay home with sick children. Even though women have flocked to these fields, being nearly at parity or in the majority among new doctorates in them, some have argued that women have not progressed to the top of these fields either (for a European contrast, see Nelson & Brammer, 2008), despite their ample presence in the PhD pool. Women tend to drop out of tenure-track positions at higher rates than do men (see online Section 6). For example, the proportion of women in senior faculty positions at Harvard University still averages only 13% across all disciplines, not just in STEM fields (Harvard University, 2005). In medicine at the University of Pennsylvania (Gender Equity Committee of the University of Pennsylvania, 2001), “the

resignation rate among female Assistant Professors-CE [clinician educator] (16%) is significantly higher than that of male Assistant Professors-CE (9%)"; the committee's Executive Summary summarizes the situation: "This large group of junior faculty women is experiencing particular difficulty reconciling their professional responsibilities with the demands of family and home life, resulting in an unusually high resignation rate" (p. III) (see online Section 6). Even in fields in which women are well represented (e.g., medicine, law, and veterinary science), they are not found in the top positions commensurate with their numbers. They are either not on tenure track (Mason & Goulden, 2004), dropping off tenure track for part-time work until their children get older (Leslie, 2007), or stalled at the associate professor stage. According to Hamel, Ingelfinger, Phimister, and Solomon (2006), in 1960, only 5% of medical students in the United States were women, whereas today they comprise approximately 50%; despite these gains, women who enter academic medicine are less likely than men to be promoted or to serve in leadership posts. As of 2005, only 15% of full professors and 11% of department chairs were women (Hamel et al., 2006). The humanities are also affected: Only 18.7% of the tenure-track faculty in the top 20 philosophy departments are women (Haslanger, 2007). Even in cell and molecular biology, fields where women obtain nearly half of all PhD degrees, they drop out of the pipeline at multiple points, as postdocs and as assistant professors, though Strenta, Elliott, Adair, Matier, and Scott (1994) showed that in four highly select institutions, there are no sex differences in dropouts from undergraduate programs in biology. Leboy (2007) reviewed 20 medical schools and found that female assistant professors lagged behind the PhD pool from a decade earlier by 10–15%. Thus, the proportion of women in the PhD pipeline does not predict the proportion of those to be hired as assistant professors on tenure track within the next decade. For example, in 1994–1996 women comprised 45% of all PhD degrees awarded in the biomedical sciences but just 29% of the tenure-track assistant professors hired within the following decade. According to Leboy, women opt out of tenure-track positions for alternatives, finding the schedules too restrictive to allow them to care for children and older parents. Leboy's analysis suggests that many women who drop out of tenure track, or do not proceed on it in the first place, have not forsaken research, as they continue to publish and present at conferences. Methods for continuing to invest in and capitalize on these women's resources merit further investigation.

On the basis of our review of the literature, much of the explanation for women's underrepresentation in math-intensive fields can be found in the career–family trade-off and in a greater preference for what Hakim (2006, 2007) termed "home-centered" lifestyle (as well as in sex differences in career preferences that we detail below). This is as opposed to looking for causes of differences in mental rotation, hormones, or teacher expectancies. Therefore, in our final causal model, we have accorded this factor primary status and the others lower status. There is more to success as a scientist than being in the right tail of the cognitive distribution, no matter how important this may prove to be, and we agree that there is a significant underrepresentation of women on some high-aptitude spatial and mathematical tests. If women were as successful as their cognitive ability predicted, they would be far more numerous in math-intensive fields than they are. It appears that the family–career trade-offs constitute a major factor in the

dearth of women in fields such as engineering, physics, computer science, and so forth.

In several large-scale surveys of academics, women's success in academia seems to be on a collision course with their success as parents. Above, we described evidence from surveys showing that women work fewer hours per week at their academic jobs compared with men with children as well as men and women without children (Jacobs & Winslow, 2004; Mason & Goulden, 2004). In the survey by J. A. Jacobs and Winslow (2004) women also reported lower rates of marriage and fewer children: 41% of female academics are married with children versus 69% of male academics. Among academics hired within the first 12 years of earning their PhD, only 30% of tenure-track women have children versus 50% of tenure-track men. Among older academics, 40% of women express the wish for more children versus 29% of men. Finally, female academics are more likely to be unmarried, 28% of women versus 11% of men, and to be divorced (144% of the men's rate). This means that women are more likely to be sole wage earners than men, and they probably are more likely to care for elderly parents even when they have no children of their own. As Leslie (2007) showed, in his analysis of four surveys by the National Center for Education Statistics, there is a linear trend between the number of children and the number of hours worked at an academic job, with more children reducing women's hours of working at their academic jobs but actually increasing men's hours on the job (Figure 4). Leslie (2007) concluded, "It is increasingly clear that having children has a particularly serious effect on women's careers" (p. 12). (Lest this be interpreted that having and investing in children is somehow bad, one could argue that the "bad news" is that women scientists do not have as many children as they prefer. Findings indicate that women who work less or opt out of STEM careers are as satisfied with their life as men who work 60 hr per week. In other words, these women do not interpret their transfer of hours from career to family as negative; Lubinski et al., 2006, pp. 197–198).

To this survey evidence, we can add Hakim's (2006) argument:

There is solid evidence that men and women continue to differ, on average, in their work orientations and labour market behaviour, and that these differences are linked to broader differences in life goals, the relative importance of competitiveness versus consensus-seeking values, and the relative importance of family life and careers. (p. 280; see also Hakim, 2000, 2003, 2004, 2005)

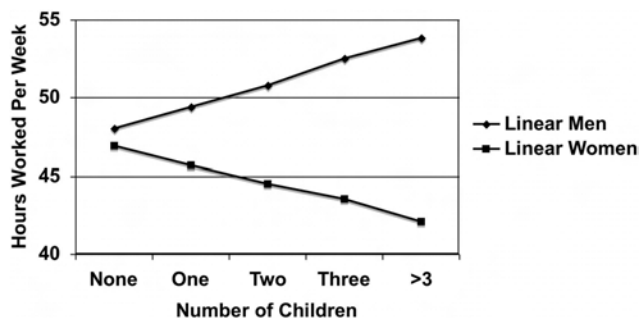


Figure 4. Number of dependent children and hours worked per week (linear trends, 2004). Data are from "The Reshaping of America's Academic Workforce" by D. W. Leslie, 2007, *Research Dialogue*, 87, p. 13. Adapted with permission.

Hakim has presented data showing that 10%–30% of women in various surveys she has summarized prefer home-centered lifestyles and prefer not to work outside of the home and about 60% prefer adapted work lifestyles; however, only approximately 20% prefer work-centered lifestyles in which the main commitment is to career. In contrast, more men are work-centered, and far fewer of them are home-centered according to a national survey in Great Britain (Hakim, 2006).

Ginther (2001) analyzed data from a nationally representative sample receiving science doctorates from 1975 to 2000. She found sex differences in promotion to tenure, after controlling for productivity (number of papers published and presented) and demographics (promotion probability was not significant for the most recent cohort by itself). Women were paid less, though some of the difference disappeared when productivity measures were controlled. However, pay differences for full professors could not all be explained by controlling these factors. Differences in the model's coefficients for full professors suggested that any given level of productivity yields less for women than for men. However, salary gaps have been closing over the past 30 years, and when Ginther (2004) decomposed the factors into observable versus unexplained portions, the former referring to characteristics that affect pay, such as rank, years since earning the PhD degree, type of institution, and field, she found that there was an overall unexplained gap of only 2%, mostly due to higher unexplained differences among full professors. Among assistant and associate professors, the unadjusted 2001 salary gender gaps of 5% and 4%, respectively are virtually completely explained by observables; hence, wage gaps of the magnitude commonly reported are unlikely to be a reason that fewer women enter or remain in math-intensive careers. In an extension and update of this analysis, Ginther and Kahn (2006) found that the gender difference in hiring, tenure, and promotion (after controlling for demographics, family, productivity, type of institution, and field of study) vanished. (In Rudd et al.'s recent analysis of social scientists, the median salary for women faculty was equivalent to men's.) Interestingly, family characteristics have a different impact for men and women, including the likelihood of being promoted in some fields. For example, in the physical sciences, having school-age children lowers the probability of women being promoted to full professor by 9.6% while having no effect on men. As Ginther and Kahn (2006) noted, although

women [doctoral recipients] are less likely to take tenure track positions in science, the gender gap is entirely explained by fertility decisions . . . Children create a marked divergence between men and women. The presence of a pre-kindergarten-aged child lowers women's likelihood of having a tenure track job by 8.2 percent . . . [while having] no effect on men's likelihood" (pp. 8–9).

(Rudd et al. showed that in the social sciences, women are slightly more likely to take tenure-track posts than are men.) An older child increases a man's likelihood of attaining a tenure-track job but has no effect on a woman's. The beneficial effect of children for men may reflect characteristics of men who marry and have children compared with those who stay single. This effect may also reflect societal biases favoring men with children as employees while disfavoring mothers as employees because of their presumed lower productivity.

Women with families are less likely to be on tenure track and are also more likely to be at small colleges and/or in adjunct positions (Leslie, 2007; Mason & Goulden, 2004). As Ginther and Kahn's (2006) analysis reveals, and as Mason and Goulden's survey data show, once a woman opts to go off tenure track or delays going on tenure track, the chances of getting on tenure track are reduced. Some argue that this is an important reason women earn less and are less often in tenured positions in scientific fields. In sum, the claim that women are affected by family demands that statistically are less likely to affect men is regarded by some as an institutional barrier that has kept women from rapid promotion and advancement and a more likely reason for their underrepresentation in STEM careers than the smaller number of women at the right tails of mathematics and science distributions (Halpern, 2007; Shalala et al., 2007). However, one could interpret this as an admission that extant salary and rank discrepancies are justified because men invest longer hours in their careers, uninterrupted by family needs. When women and men put in similar hours, there do not appear to be rank gaps (Mauleón & Bordons, 2006). In his controversial remarks, former Harvard University president Lawrence Summers invoked this as a possible explanation for the dearth of women in math-intensive careers:

We would like to believe that you can take a year off, or two years off, or three years off, or be half-time for five years, and it affects your productivity during the time, but that it really doesn't have any fundamental effect on the career path. A whole set of conclusions would follow from that in terms of flexible work arrangements and so forth. The question is, in what areas of academic life and in what ways is it actually true. (Summers, 2005)

McDowell (1982) studied this issue by classifying the shelf life of research publications to determine how often old articles were cited. He took this as evidence that it was possible to take a child care leave and return without falling far behind in the knowledge needed to publish. He reported that research in the STEM fields becomes obsolete particularly fast. Correspondingly, McDowell showed a bigger child penalty in these fields in terms of research productivity for women than for men who were in other ways comparable to them. Women in math fields may thus pay a bigger "current knowledge" price for taking family leaves.

However, most of the abovementioned factors associated with women's lower success are equally relevant to women in the humanities and social sciences, and indeed to women in medicine, veterinary medicine, and dentistry. Clearly, the few factors that are specific to the math-intensive fields (such as the presence of school-age children reducing the probability of promotion to full professor for women by 9.2%) are insufficient to fully account for women's underrepresentation. Another factor that needs to be added to the mix concerns personal choices and preferences. There are pronounced sex differences in occupational preferences that occur along a "people-to-object" dimension (Lippa, 1998): Women are more likely to pursue people-oriented or organic fields, whereas men with similar mathematics and science ability tend to pursue object-oriented fields (Webb, Lubinski, & Benbow, 2002; see Lippa, 2005, for a review). Sex differences on the people-to-object dimension are quite large: $d = .62$ for people and $d = 1.29$ for object (Lippa, 2006), and they are longitudinally stable (Lippa, 1998, 2005). Sex differences in occupational preference account for more variance in the prediction of later careers

than does the SAT-M or the Graduate Record Examinations Quantitative section (Achter, Lubinski, Benbow, & Eftekhari-Sanjani, 1999; Wai, Lubinski, & Benbow, 2005). In a tracking study of 1,100 high-mathematics aptitude students who expressed a goal of majoring in mathematics or science in college, Achter et al. and Wai et al. found that many later switched to non-mathematics majors and that such individuals were more likely to be women. Although all of these 1,100 students came from the top 1% in mathematics aptitude, they manifested both ability and interest differences that were evident long before they began taking different courses that led to different college majors. One determinant of who switched out of math/science fields was the asymmetry between their verbal and mathematics abilities. Women's verbal abilities on average were nearly as strong as their mathematics abilities (only 61 points difference between their SAT-V and SAT-M), leading them to enter professions that prized verbal reasoning (e.g., law), whereas men's verbal abilities were an average of 115 points lower than their mathematics ability, possibly leading them to view mathematics as their only strength.

Numerous researchers (Eccles, 2007b; Eccles, Barber, & Jozefowicz, 1999; Halpern, 2007; Hines, 2007; Hyde, 2005; Kimura, 2007; Lubinski & Benbow, 2006, 2007) have urged factoring students' interests into the predictive mix, because the sexes' different interests propel them into different careers (see Lubinski & Benbow, 2007, for a review). Historically, there have been pronounced swings away from male dominance in fields such as teaching and secretarial work, and more recently medicine and veterinary science, explained in terms of changes in prestige and income rather than by differences in hormones, aptitude, or genes. Eccles (2007a) showed that young women were more likely than men to aspire to health careers, because they place higher value on people-oriented jobs, and this remained true after their mathematics ability was controlled.

Sex differences in career aspirations can be seen as valuable rather than as a deficit: Talented men and women choose how they would like to develop, even if it leads them to excel in different areas. It seems benign if females are, on average, interested in different fields than males, with more women going into biology, law, and medicine and more men opting for physics, computer science, and engineering. One might ask why it is more valuable to encourage women to shift from their dominance in biology to mathematics—so that they can end up working on a search algorithm for Google rather than on a cure for AIDS? In the debate over women's underrepresentation in math-intensive fields, there is often the implied assumption that if a field does not have a certain percentage of women, it is a sign that societal barriers exist, rather than that women are making informed choices and thus the field must remedy the situation (Gates, 2006a). As Webb et al. (2002) concluded,

Contemporary discourse regarding the mathematics–science pipeline implies that a great societal loss is encountered when a person with high mathematics–science potential chooses to apply those talents outside engineering or the physical sciences, but why? Is an individual's contribution to environmental law, for example, less valuable than a contribution to chemistry?" (p. 791)

However, even in fields in which women are well represented (e.g., medicine, law, and veterinary science), they are not found in the top positions commensurate with their numbers. They are

either not on tenure track (Mason & Goulden, 2004), dropping off tenure track for part-time work until their children get older (Leslie, 2007), or stalled at the associate professor stage. According to Hamel et al. (2006), in 1960 only 5% of medical students in the United States were women, whereas today they comprise approximately 50%; despite these gains, women who enter academic medicine are less likely than men to be promoted or serve in leadership posts. As of 2005, only 15% of full professors and 11% of department chairs were women (Hamel et al., 2006). The humanities are also affected: Only 18.7% of the tenure-track faculty in the top 20 philosophy departments are women (Haslanger, 2007). Even in cell and molecular biology, fields where women obtain nearly half of all PhD degrees, they drop out of the pipeline at multiple points—as postdocs and as assistant professors. Leboy (2007) reviewed 20 medical schools and found that female assistant professors lagged behind the PhD pool from a decade earlier by 10%–15%. Thus, the proportion of women in the PhD pipeline does not predict the proportion hired as assistant professors on tenure track within the next decade. For example, in 1994–1996 women comprised 45% of all PhD degrees awarded in the biomedical sciences but just 29% of the tenure track assistant professors hired within the following decade. According to Leboy, women opt out of tenure track positions for alternatives, finding the schedules too restrictive to allow them to care for children and older parents. Leboy's analysis suggests that many women who drop out of tenure track, or do not go on it in the first place, have not forsaken research, as they continue to publish and present at conferences. Methods for continuing to invest in and capitalize on these women's resources merit further investigation.

The tenure system requires a young professor to show excellence at a young age and work full-time (or more). This makes it particularly difficult for women, who, according to surveys, still carry the major share of child rearing, care of older relatives, and (although this is not a topic that many find politically appealing to mention) often defer their careers to those of their male partners. None of this makes universities' tenure tracks easy workplaces for women, even with progressive child-care benefits and family leave policies. However, before we assume that enacting more progressive family leave policies would change the representation of women in STEM careers, data are needed that presently do not exist to answer questions such as, Can scientists be productive in the long haul if they reduce their scientific effort for a period of time early in their careers? Is scientific output in the first few postdoctoral years predictive of lifetime output, as assumed by the tenure system, or does striving for tenure provoke only a temporary spurt of activity that is unrepresentative of later productivity? Do fields differ in the consequences of delayed start-up or part-time work to raise children? As noted earlier, McDowell (1982) found that fields differ in the speed of obsolescence of their cited research; hence, some fields could be associated with greater penalties for taking a break or reducing time spent on research. Of course, it is not possible to conduct the relevant experiment to answer such questions because those denied tenure are not given the opportunity to show what they could have done later in their career, although there may be some merit in studying countries with different employment systems that would permit an examination of the prediction of early productivity. These are the sort of data that are lacking, with the exception of only a few studies that

we describe next, none of which spans multiple fields or types of institutions.

An investigation of careers of biochemistry PhD recipients by Long (1992) argues that the timing of tenure is tilted against women. The tenure schedule punishes temporary reductions in productivity by a permanent withdrawal of opportunities at the time when many women have to make such a reduction in productivity for child rearing. Long (1992) analyzed the scientific productivity of a cohort of researchers who graduated in the 1950s and 1960s. The annual number of articles and citations was tracked over their careers. The curve of citations by year for men rose steeply in the first few years, then leveled off, whereas the curve for women fell in the 4th year before leveling and then rising steeply in later years, catching up with men. Although men produced more articles overall, women produced higher impact articles (higher citations) throughout their careers, and by year 17 the average paper by a woman is cited 1.5 times more often than the average paper by a man. However, because of the small proportion of women entering these fields at that time, it may be that women in Long's sample were selected from the most extreme right tail of the ability distribution; the far higher proportion of men who entered these fields and received PhD degrees in the late 1950s and 1960s probably means they spanned a wider range of ability than did the women who were more cognitively select, given that so few women went into biochemistry at that time. This would bring into question the generalizability of the study's finding that women had higher average citation counts, although not the demonstrated productivity dip.

Similar findings have been reported for other fields. Mauleón and Bordons (2006) analyzed the bibliometric records of 333 Spanish materials scientists between 1996 and 2001. Like Long, these authors found that women published in higher impact journals than did men at their rank, particularly at the highest rank. However, because there was a lack of significant differences between male and female scientists within a given rank, Mauleón and Bordons inferred that the promotional system was not sex-based but rather was due in part to women's lower overall productivity, especially at the lower ranks. These researchers concluded the following:

Productivity increases with professional category (rank) for both men and women . . . gender differences in productivity within each professional category (rank) were not found, an issue that might indicate that scientific requirements for promotion are independent of sex. However, a different "lifecycle" of productivity for men and women is found in the area. The lower productivity of women as a group can be due to their lower presence in the upper and most productive categories, but also to their lower productivity at specific age classes, whose reasons would require further analysis. (Mauleón & Bordons, 2006, p. 215)

One could argue that the impact ratings of journals or mean number of citations to articles are not what the system selects for but rather productivity, total number of citations, and number of sole-authored articles and invitations to publish in special issues (an indication of status and visibility that might be rooted in early productivity but which results in low-impact articles in terms of citations or low-impact journals), and citation rankings are heavily influenced by factors unrelated to journal quality, such as the

number of active researchers in an area, the number of journal pages devoted to a topic, and the half-life of citations in a subfield.

Taken together, the studies of Long (1992) and Mauleón and Bordons (2006) reveal a structural tilt in the tenure system in the timing of promotion decisions: The biggest gender gap in productivity occurred at the time at which tenure decisions were made. Predictions of future impact based on assessments made 5 to 7 years after attaining the doctorate were often not good. If the goal of the tenure process is to select those expected to have future impact, these studies—particularly Long's—suggest that the process is flawed. Of note, in both of these studies, early productivity did predict later productivity, although it did not predict later impact.

On the basis of our review of the literature, much of the explanation for women's underrepresentation can be found in the career–family trade-off and in a greater preference for home-centered lifestyles (Hakim, 2005, 2006, 2007) (as well as to sex differences in career preferences). This is as opposed to looking for causes of differences in mental rotation, hormones, or teacher expectancies. Therefore, in our final causal model we have accorded this factor primary status and the others lower status in explaining the underrepresentation of women at the top in STEM fields.

Brain Development

We turn next to research on sex differences in brain development that have been linked to women's underrepresentation in STEM fields. All of this evidence comes from studies of mean differences rather than right-tail gaps.

Continuing around the causal circle in Figure 1, differences in motivation/interests and activities may, in turn, lead to differences in brain development, manifest in observed differences in brain size, structure, and function, or in consequent differences in ability. A biopsychosocial view of sex differences postulates that biological factors are enmeshed with social forces at every step, in an iterative unfolding (Berenbaum & Resnick, 2007; Bronfenbrenner & Ceci, 1994; Guo & Stearns, 2002). Thus, the social context brings biological potential to fruition and, to the extent that the context is more supportive for one sex than for the other, that group's spatial and mathematical ability will be a better reflection of its biological potential. For example, spatial skills of females with non-right-handed relatives benefit more from spatial experiences than skills of females whose relatives are right-handed, illustrating the interaction of biology and experience in the development of spatial skills (Casey & Brabeck, 1989, 1990). Casey, Nuttall, and Pezaris (1999) hypothesized that genes affect brain organization (reflected in handedness) and are manifest in the ability to capitalize on experiences to develop spatial skills. The hemispheric brain organization of girls from all-right-handed families is less optimal for developing spatial skills. One source of spatial experiences is playing with male siblings, who typically engage in more play drawing on such skills. According to Casey et al. (1999), the results of their study show that "children from all right-handed families do not appear to be able to use their spatial experience with male siblings to increase their spatial skills" (p. 1237). However, this interpretation is complicated by the finding that girls from all right-handed families who did not have brothers did just as well as girls from mixed-handedness families who did

have brothers. Thus, the conclusions regarding the role of this gene–environment interaction on spatial skills are unclear (see Section 7 in the supplemental materials).

The direct evidence for the role of activities on brain development is clearer. For example, in a group of 21 female and 3 male 22-year-olds, a 3-month period of practice to master a three-ball juggling cascade for at least 60 s in the air resulted in an increase in gray-matter density in the temporal cortex, human middle temporal/V5 complex (hMT/V5), which reverted to pre-juggling morphology 3 months after juggling ceased, hence revealing a direct relationship between emerging ability and brain morphology (Draganski et al., 2004). Similarly, magnetic resonance images (MRIs) of the posterior hippocampi of 16 London male taxi drivers showed that they are larger than those of nondrivers, whose anterior hippocampal region was larger than taxi drivers'. Hippocampal volume correlated with the amount of time spent as a taxi driver: positively in the posterior and negatively in the anterior hippocampus (Maguire et al., 2000). These data accord with the view that the posterior hippocampus can expand regionally to accommodate those with a high dependence on navigational skills. (The hippocampus is an area of the brain that continually produces new neurons in adulthood, and thus the extent to which implied plasticity in this brain region generalizes to other regions that subserve other cognitive functions is unknown.)

Differences in brain development may also be caused directly by biological sex (Gur & Gur, 2007; Haier, 2007), but the etiology is unclear. Differences in head volume and perimeter, which correlate very highly with head mass at autopsy, suggest to some that women are biologically less capable at scientific reasoning (see Scheibinger, 1987). Rushton (1992a, 1992b) studied the head measurements of several thousand Army personnel and found that women had smaller brains than men, even when adjusted for body size. He reported that, after adjusting for stature and weight, and then for sex, rank, and race, the cranial capacity of men was larger by 110 cm³. Despite great variability in brain morphometry, boys' cerebral volume is 9% larger than girls, on average (Giedd et al., 1996). When brains are at their maximum size (age 25 years), men's are 175 g (17%) heavier than women's. Ankney (1992) found that a 142-g difference remained, in favor of men, after correcting for body size differences. Analyzing 55 studies, Rushton and Ankney (2007) found that only 30% of the sex difference in brain size is due to differences in body size. Such results have led some to suggest that males' greater brain mass is genetic and is responsible for their superior mathematical and spatial ability (Ankney, 1992).

However, although brain size differences might have cognitive consequences, it is unclear why such a gross difference would result in a particular deficit in mathematical and spatial skills rather than other cognitive processes, or for that matter why women excel in so many cognitive domains. Perhaps the argument is that structural differences in areas subserving spatial or math ability covary with size; however, the sex size difference is pervasive and is not confined to a specific region (Rushton & Ankney, 2007). Further, given that women achieve better grades in mathematics than men, one must argue that brain size specifically affects, say, SAT-M scores, but not school mathematics grades. As Rushton and Ankney (2007) noted, in both MRI studies that have examined mental rotation and brain size, no correlation between them was significant.

Brain imaging studies have identified more subtle male–female brain differences (e.g., Gazzaniga, Ivry, & Magnum, 1998; see

Table 6 in the supplemental materials). Haier, Jung, Yeo, Head, and Alkire (2004, 2005) examined whether sex differences in the amount of gray and white matter in different brain areas are related to general intelligence in 47 volunteers from the general population. The researchers found structures distributed throughout the brain in which the amount of gray matter or white matter predicts IQ scores. Specific areas associated with language in the frontal and parietal lobes seem especially important (see also Hugdahl, Thomsen, & Erslund, 2006): “When we reanalyzed our MRI data separately for men and women, we found completely different brain areas correlated to IQ . . . The amount of gray and white matter in the frontal areas seems more important in the women; the gray matter in the parietal areas seems more important in the men” (Haier et al., 2005, p. 146).

Others have shown that the volume of these same areas appears to be under genetic control. Thus, it could be that the sexes achieve the same cognitive capability using different brain structures. Research with fMRI during mental rotation tasks (see Gur & Gur, 2007; O’Boyle et al., 2005) reveals sex differences in brain optimization. A number of studies have specifically provided evidence for organizational differences in male and female brains during mental rotation (Gur et al., 2000; Haier et al., 2004, 2005; Halari et al., 2006; Hugdahl et al., 2006), although the modal study investigated small, unrepresentative samples with mostly correlational data (with some exceptions, e.g., Krendl et al., 2008). Hugdahl et al. (2006) argued that males utilize a parietal lobule “gestalt” perceptual strategy, whereas females may utilize a frontal lobe “serial” reasoning strategy, suggesting that males are biased toward a coordinate approach and females toward a categorical approach, showing more left-sided activation during mental rotation. Thus, the two sexes can achieve the same level of mental rotation using different strategies, which lead to activation of different brain areas. As with the brain size work, the etiology of these differences—whether driven directly by biological sex, by sociocultural factors, or by an interaction—is unclear, and the implications for the dearth of women in STEM is unknown because this research was not conducted with participants from the extreme right tail.

In sum, there is support from different approaches for the view that there are brain-related sex differences in many behaviors that might be related to performance in STEM fields. However, etiology is unclear, and the studies do not examine extreme right-tail samples.

Abilities

Differences in brain development feed into potential differences in abilities. Much of the debate about sex differences in STEM fields centers on differences in precursor abilities that are purported to lead to differences in the abilities needed to perform high-level STEM jobs. Unfortunately, there is little direct research on the actual abilities involved in performing STEM jobs successfully at the highest levels: We are looking for sex differences in precursors without knowing the criterion: how much and what kind of math or spatial ability is needed. Researchers have studied a variety of possible precursors, including (a) global characteristics such as intelligence, (b) specific cognitive tasks (spatial cognition, particularly three-dimensional mental rotation), and (c) mathematical aptitude tests, such as the SAT-M.

General Intelligence

Some have invoked general intelligence to account for the dearth of women in math-intensive fields. For example, Lynn (1991) argued for an evolutionary account of what he claimed is lower general intelligence of women. He opined that during the evolution of hominids, intelligence became an important determinant of male success. Relatedly, Lynn and others (e.g., Eals & Silverman, 1994; Geary, 1996, 1998; Sanders, 2007) have suggested that male specializations in hunting and making artifacts may have been more cognitively demanding than female specializations in gathering foods and child rearing. Although he did not discuss general intelligence, Geary (1998, 2002) has provided a unifying framework for sexual selection that stresses hormonal, experiential, and evolutionary influences on cognitive sex differences and described a rationale for why sex differences may have evolved (see Section 8 of the online materials). Critics of evolutionary theory's role in sex differences in spatial reasoning have pointed to counterexamples (e.g., Newcombe, 2007; see commentaries following Geary's, 1998, target article; see Halpern et al.'s, 2007, critique) and reported instincts that reflect equivalent evolutionary pressures for the sexes (e.g., Hrdy, 1999). The Halpern et al. (2007) team differed among themselves, with evolutionary psychologists maintaining that "the male brain is naturally better prepared to perform some spatial tasks and others who feel the weight of the evidence is clearly on the environmental side" (p. 24). We endorse their summary position that the available evidence is insufficient to determine the impact of evolution on sex differences in cognitive ability, although it presents intriguing suggestions.

Mean differences. Two meta-analyses reported that male adult means on general intelligence tests (Raven's Progressive Matrices) are approximately 5 IQ points above female means (Irwing & Lynn, 2005; Lynn & Irwing, 2004). However, other evidence suggests that when measurement controls are exerted and sampling is representative, women perform as well as men on general intelligence, including the Raven's Progressive Matrices (Flynn & Rossi-Case, 2008). Recently, Brouwers, Van de Vijver, and Van Hemert (in press) conducted a meta-analysis of cross-national Raven's scores between 1944 and 2003. When the sum scores of the Advanced, Colored, and Standard test versions were transformed to a single 0–100-point scale, Brouwers et al. found no sex differences: mean for male test takers ($N = 175$) = 61.71, mean for female test takers ($N = 113$) = 62.76, $F(1, 286) = 0.33$, $p < .564$, partial eta-squared = .01. (Including various country-level covariates did not alter this result.) At the outset of the mass testing movement, sex differences in general intelligence were small to negligible (in fact, the early Stanford–Binet and Wechsler–Bellevue tests yielded small but insignificant advantages for female test takers; Macintosh, 1996), although differences on later measures were eliminated by design (see Ackerman, 2006). For example, a population study of 87,400 children born in Scotland in 1921 showed that when they were 11 years old, the children's mean intelligence score was 43.1 for boys and 43.5 for girls (Deary, Whalley, Lemmon, Crawford, & Starr, 1999). Raven scores of 97 of these surviving individuals tested in 1998 revealed differences favoring males that were not reliable (30.2 vs. 27.5, respectively, $p = .10$). Notwithstanding equivalent levels of general intelligence in the sexes, Spinath, Spinath, and Plomin (2008)

found, in a large British sample of 9-year-old twins, that general intelligence was the strongest predictor of sex differences in math (favoring boys) and in English (favoring girls) as well as of boys' and girls' perceptions of abilities in these respective domains. Taken together, however, these results urge caution in the interpretation of differential intelligence as an explanation for sex differences in STEM fields.

Right-tail differences. Although Deary, Thorpe, Wilson, Starr, and Whalley's (2003) test results of 11-year-olds showed no sex difference in the center of the distribution, there was a larger male standard deviation and an excess of boys at both the low and high extremes: the ratio of girls to boys was 1:1.4 at the right tail ($IQ \geq 130$).

Specific Cognitive Tasks

Other researchers have investigated sex differences on specific rather than general ability measures, including a large body of research on spatial tasks. Many have argued for the spatial basis of mathematics (e.g., Fias & Fischer, 2005; Geary, 1998; McGee, 1979), reporting that kindergartners' visual integration and discrimination of geometric forms predicts unique variance in 4th-grade mathematics (Kulp, 1999; Kurdek & Sinclair, 2001) and that block-playing skill in kindergarten correlates strongly with later mathematical ability (Wolfgang et al., 2001). Factor analysis of the standardization data from the Wechsler Preschool and Primary Test of Intelligence by LoBello and Gulgoz (1991) showed that math performance loads .40 on the visuoperceptual organization factor at every age. Thus, some evidence exists that early spatial cognition underpins later math ability, a position that had long been hypothesized (e.g., Sherman, 1967). Recently, Kersh et al. (2008) have reviewed additional evidence for the spatial–mathematical causal connection.

Mean differences. Sex differences are complex. Far from the monolithic stereotype of female superiority in verbal domains and male dominance in quantitative ones, females excel at some forms of arithmetical calculation (number operations), verbal fluency (vocabulary, writing), perceptual speed, associative memory, and some forms of nonverbal reasoning; males excel at spatial reasoning, as well as at social studies, history, and math word problems (Hedges & Nowell, 1995, Table 2). The magnitudes of the differences on most of these measures are small ($d_s \leq .2$), though several are quite large (e.g., mental rotation of three-dimensional items can range between 0.85 and 1.06; Nordvik & Amponsah, 1998) and have not declined between 1974 and 1992 (Masters & Sanders, 1993; see meta-analyses by Hyde, 2005; Linn & Petersen, 1985; Masters & Sanders, 1993; Voyer et al., 1995; for a discussion of complexity, see Feingold, 1992, 1994; Halpern et al., 2007).

Mean sex differences change with development, with girls initially better at computation, but the difference fades by adolescence, and although there is no initial difference in complex mathematics problem solving, boys surpass girls by high school (Hyde, 2005; Wise et al. 1979). As noted, Rathbun et al. (2004) reported sex differences in math achievement among first graders using the Early Childhood Longitudinal Program—Kindergarten Cohort (ECLS-K) data set.

Notwithstanding the dispute over whether sex differences have been narrowing, one skill stands out as a large-magnitude effect: mental rotation. This occurs on tasks in which two- and three-dimensional perspective drawings are shown at different orienta-

tions and one must determine if they are the same object as well as on tasks in which one is asked to judge whether a two-dimensional piece of paper can be folded into a three-dimensional shape. Response times increase as the angle of disparity between the two shapes increases, suggesting that an image must be mentally rotated to be superimposed on the reference shape (Hugdahl et al., 2006). On such tasks, the size of the mean sex gap is large ($d \approx 0.7$ – 0.8 ; Hyde, 2007; Voyer et al., 1995). However, explanations are complicated by the fact that men are likelier to form an image of one object and rotate it mentally to see if it aligns with the other; in contrast, women are likelier to engage in a feature-by-feature comparison of the objects (Spelke & Grace, 2007). Sometimes one strategy is more effective than the other, and both males and females can use both strategies. When they are constrained to use only one strategy, males and females tend to perform somewhat more similarly. For spatial targeting tasks (e.g., dart throwing), the effect size is larger ($d = 1.3$ – 1.9) than for mental rotation tasks (Jardine & Martin, 1983; Watson & Kimura, 1991), although, the conceptual connection with mathematics skills is less obvious, as motor skills complicate matters.

However, sex differences in adult mental rotation are not necessarily biological. As mentioned in the section on motivation, a lifetime of different experiences could yield differences in brain and/or behavior that are not necessarily biological in origin. If boys were typically spending childhood building with blocks while girls were playing with dolls (e.g., Kersh et al., 2008), it would not be surprising to find spatial and social skill differences, but this would not prove that their origin was innate. A number of studies have shown that mental rotation can be improved by either repeated testing or playing video games, and although the magnitude of sex differences is sometimes unchanged by such experiences (Terlecki et al., 2008), the rate of improvement is slower for females, leading to the suggestion that longer interventions might result in greater gains for them, as males reach asymptote early. Data are needed before the onset of experiential differences. Robinson et al. (1996) have reported the longitudinal relationship between spatial cognition and mathematics in precocious kindergarten and primary school children. There is some suggestion that correlations between mathematical and spatial abilities are higher in girls than in boys (Friedman, 1995), leading Robinson et al. to conclude that “the question of early gender differences in both mathematical and spatial precocity is thus still open” (Robinson et al., 1996, p. 342). Sex differences in spatial cognition emerge early, according to new research by two independent labs. Using a simple habituation paradigm with a Shepard and Metzler (1971) object, Moore and Johnson (2008) showed that, after habituation, when infants see the same object in a new perspective alternating with its mirror image, 5-month-old boys looked longer at the mirror image, but girls showed no preference. Using static drawings of a two-dimensional object rotated in a two-dimensional (frontal) plane, P. C. Quinn and Liben (2008) also found mental rotation in 3- to 4-month-old male infants but not in female infants.

Although Maccoby and Jacklin (1974) reported sex differences in spatial skills by adolescence, Levine et al. (1999) reviewed studies documenting sex differences among preschoolers: “Uttal, Gregg, and Chamberlain (1999) found that 5-year-old boys were better at interpreting a map of a space than 5-year-old girls, particularly when the map was rotated with respect to the space it represented” (p. 940). Levine et al. (1999) also highlighted find-

ings that “boys as young as 4 years of age performed better than girls on a task that involved replicating spatiotemporal patterns tapped out by the experimenter on a set of blocks, and the size of this sex difference remained constant across the 4- to 10-year age range (Grossi, Orsini, Monetti, & De Michele, 1979; Orsini, Schiappa, & Grossi, 1981)” (p. 940). Finally, Casey et al. (2008) reported sex differences in three-dimensional mental rotation among kindergartners. Additionally, Waber, DeMoor, Forbs, Almi, Botteron, and Leonard (2007) reported the first wave of findings from the large National Institutes of Health (NIH) Brain Development Study, in which boys outperformed girls on the Block Design subtest, which requires three-dimensional construction of blocks to match a two-dimensional perspective drawing.

However, other studies of spatial differences have reported inconsistent results. Although the Uttal et al. study is cited above as demonstrating early sex differences, this does not appear to be the case: In all four of Uttal et al.’s (2001) experiments, there are no reliable sex differences, including Experiment 4, which involves rotation. Siegel and Schadler (1977) suggested that studies that failed to find sex differences employed easier tasks that did not overwhelm the spatial processing system and that sex differences are most likely when the spatial system is taxed. (In accord with this suggestion, there are several studies, some with rhesus monkeys, such as Lacreuse et al., 2005). Some studies have found early male superiority on two-dimensional spatial tasks, whereas others found differences only on three-dimensional tasks. Cronin (1967) found that male kindergartners and first graders scored higher than girls on a task matching triangles with mirror images. More recently, Levine et al. (1999) investigated early male superiority on another two-dimensional spatial task, recognizing transformations of shapes. Boys and girls scored equally well in the youngest age group, which could be due to the task being quite difficult for 4-year-olds (mean correct = 9.96, with chance being 8), but a significant male advantage developed later.⁷ Siegel and Schadler (1977) asked 5-year-olds to place 40 items in a three-dimensional model of their classroom. On all dependent measures, boys exceeded girls by very large amounts. Johnson and Meade (1987) administered a battery of seven spatial tasks to 1,800 students in kindergarten through 12th grade. They reported that boys were favored on 78 out of 96 contrasts, though there was no pattern of male superiority in grades kindergarten through 4th grade, including on several tasks that tap rotation and spatial relations; male superiority was consistently observed only from 5th grade and up. Other studies of similar age groups failed to find sex differences on two-dimensional tasks. McGuinness and Morley (1991) found no difference on a jigsaw task with 3-, 4-, and 5-year olds. However, they did find a difference in favor of boys on a three-dimensional Lego-building task: Boys ages 4–5 years were 1 year more advanced than girls, but girls caught up by kindergarten, which the authors suggest was due to a ceiling effect.

⁷ However, a 15-min testing session another day improved both sexes’ performance, with a magnitude roughly equal to the difference between the sexes. This suggests that it might not require a large difference in experience to create sex differences in performance. However, that a 15-min experience one day can, on another day, elevate both sexes’ scores by an amount equal to their original difference says nothing about the cause of those differences.

Unfortunately, few of these studies with young children focused on three-dimensional mental rotation, the skill invoked to explain adult sex differences in mathematics. Most focused on either two-dimensional rotations of maps, puzzles, and photographs or on three-dimensional tasks that did not involve rotation. Levine et al. (1999) did compare rotation and nonrotation tasks (the former involving dynamic mental transformations), and although they found a male advantage on both types of task, they found no difference in the size of the male advantage between them, a finding at variance with the adult literature. Also Waber et al. (2007) reported the first wave of findings from the large NIH Brain Development Study, and boys outperformed girls on the Block Design subtest, which requires three-dimensional construction of blocks to match a two-dimensional perspective drawing. The abovementioned study by Casey et al. (2008) provides evidence of sex differences in three-dimensional rotation ability among 5-year-olds (boys solved 5.39 mental rotations out of 10 vs. 4.08 for girls), and Wolfgang et al. (2001) reported strong correlations between block-playing skill in kindergarten and later mathematics ability.

A meta-analysis of older individuals' spatial ability conducted by Linn and Petersen (1985) found sex differences but no changes in effect sizes between the ages of 10 and 60 years. The authors hypothesized several potential causes of sex differences, including females being more cautious and taking longer because they double-check answers. They concluded that "males tend to outperform females on mental rotation at any age where measurement is possible . . . Sex differences may result from differential rate of rotation, differential efficiency in strategy application, differential use of analytic processes, or differential caution" (pp. 1488–1489).

Right-tail differences. The meta-analysis by Hyde, Fennema, and Lamon (1990) of over 100 studies of mathematics found no sex difference ($d = -.05$ in favor of females) for nonselective samples from the general population but significant male superiority ($d_s > .3$) for college-bound youths and precocious individuals. Hyde et al. (2008) found small but consistent male overrepresentation at the right tail in an analysis of 10-state data. In addition, Friedman reported an effect of 0.348 for a male advantage at the right tail, far higher than at the midpoint. For the purpose of understanding sex differences in STEM fields, these select groups are the most relevant. Royer and Garofoli (2005) reviewed the persistent evidence for greater male variability among select samples. Penner (2005) also has shown that the effect size for sex differences in mathematics increases as the samples move deeper into the right tail. Penner (2003) reported Gender \times Item Difficulty interactions in mathematics and science in the United States and 10 other countries: Male advantages that were minimal on easy questions became larger as questions grew more difficult.

Other studies have also looked at the right tail. For example, Robinson et al.'s (1996) study of mathematically precocious children, and Lachance and Mazzocco's (2006) analysis of the top quartile, showed no consistent sex differences and no evidence of male superiority on most spatial measures, whereas other studies of similarly aged children have found spatial skill differences favoring boys. However, the measures differed across these studies, making meaningful comparisons difficult. As documented in Linn and Petersen's (1985) meta-analysis, four common mental rotation tests do not correlate with each other as highly as desired. Whereas many studies of nonselect samples (see Lachance & Mazzocco, 2006) reveal that girls, if anything, excel over boys in

mathematics until 3rd or 4th grade, among high-scoring students, boys outperform girls in 3rd through 6th grade (Swiatek, Lupkowski-Shoplik, & O'Donoghue, 2000). Leahey and Guo (2001) found that sex differences in mathematics were greatest for high-scoring elementary students through high school, and Reis and Park (2001) found that high-scoring male students exceeded high-scoring female students from the 8th grade through high school. In an exception to the failure to find sex differences among young children, Penner (2005) used a nationally representative sample of children and found mathematics advantages for boys at the right tail (top 5%) as early as kindergarten. Benbow (1992) found that males outnumber females in mathematics from adolescence through adulthood, particularly in the highest scoring groups.

Some of these findings must be interpreted with caution, however, because, when differences in samples get amplified at the extremes, this can render conclusions about small subsets of very high performers less reliable.⁸ Yet, a recent analysis of a subset of more than 17,000 high school seniors' mathematics scores on the NAEP revealed that among a select sample who had taken advanced mathematics and science courses during high school, boys outscored girls significantly, despite receiving lower grades (and also despite an absence of mean sex differences; Nations Report Card, 2007). Although Robinson et al. (1996) found that young boys outscored girls on several numeracy measures, on the majority of the spatial measures there were no sex differences. Boys were significantly better only on the spatial memory measure, usually an area of female superiority (e.g., Hyde, 2005). There is some evidence that precocious boys gain more than precocious girls on spatial and mathematics over the kindergarten–second grade period, under both normal and enrichment conditions (Robinson et al., 1996; Robinson, Abbott, Berninger, Busse, & Mukhopadhyah, 1997). This study has the advantage of being longitudinal, permitting the use of growth modeling to detect small differences that accumulate over 4 years of primary school. These researchers reported effect sizes for grades kindergarten through 3rd grade on most of 24 tests of mathematical and spatial ability. Of the resultant 73 effect sizes, over 70% (52) were <0.25 , with 42% of them $<.10$. There was no pattern of boys or girls being superior on these, even among the top quartiles of boys and girls in math and spatial ability. The measures of the latter did not include mental rotation, but they did include two spatial skills that would seem related (the Form Constancy and Position in Space subtests from the Developmental Test of Visual Perception, 2nd ed.).

In sum, although there is a pattern of male advantage on some spatial and mathematical measures, results are mixed, particularly at younger ages. Whether such spatial ability is a cause of adult male dominance in STEM careers remains unclear.

⁸ A complexity in comparing differences at the right tail is the incomparability of tests purporting to measure the same construct, such as math achievement. Linear equating techniques to map the score differences of one group on one test onto those from another work well in the interior of a distribution but poorly at its tails (Kolen & Brennan, 1995). This complicates the discussion because it is difficult to equate across the different measures among the top 1% or the top 0.1%.

Aptitude Tests

In addition to research into general intelligence and specific cognitive abilities, another line of inquiry has attempted to pin the cause of sex differences in math-intensive fields on the skills demonstrated in so-called aptitude tests. These are standardized paper-and-pencil tests of various forms of reasoning ability that are not supposed to be directly taught in school, though psychometric researchers disagree on the extent to which they emphasize aptitude over achievement (Ceci, 1996).

Mean differences. Although elementary school-age girls score as well as or better than boys in science and mathematics, there is some slippage by high school, when fewer girls take AP Chemistry and AP Physics courses. Girls begin to score lower on some science and mathematics aptitude tests around this time (e.g., the SAT-M), and the sex gap expands during high school (e.g., Wise et al., 1979), though on some math tests, such as the NAEP, there are no sex differences in mean performance among 12th graders (Nations Report Card, 2007). In mathematics and science, collapsing across ethnicity, boys in all grades receive disproportionately more scores in the top categories of Advanced and Proficient on the National Assessment of Education Progress (NAEP), a nationally representative test of subject knowledge. Scores of Proficient were obtained by 35% of boys versus 30% of girls in 4th grade, 30% of boys versus 27% of girls in 8th grade, and 19% of boys versus 14% of girls in 12th grade (*Science and Engineering Indicators*; National Science Foundation, 2006). Notwithstanding any aptitude gap, female students achieve higher grades than male students in most science and mathematics courses (Nations Report Card, 2007; Young, 1991, 1994), so some have looked to aptitude tests (at which male test takers often do better) as a possible explanation for male dominance in STEM professions (see Gallagher, 1998; Lubinski & Benbow, 2007; Mau & Lynn, 2000; Xie & Shauman, 2003). This discrepancy between aptitude scores and grades has led to claims of bias from all sides (Halpern, 2007). One side argues that teachers are biased against boys because they give them lower grades than their aptitude warrants: In a study of 67,000 college calculus students, male students who received grades of Ds and Fs had SAT-M scores that were equal to those of female students who received grades of B (Wainer & Steinberg, 1992). Proponents of the other side of the argument, however, argue that aptitude tests are biased against girls because they underpredict their grades in college, although not in graduate school (Kuncel & Hezlett, 2007). When mathematics aptitude is equated by matching on SAT-M scores, female students outperform male students in college math courses (Royer & Garofoli, 2005). The correct interpretation depends on which of these is a better measure of the mathematics skills needed to succeed in STEM fields. This is not the same thing as the best predictor, because a societal bias in favor of males could yield a higher correlation between aptitude scores (which are higher for males) and success in STEM fields (in which males are more numerous) than between grades (where females excel) and success, without the relationship carrying any causal weight. Arguments become circular: Most STEM professors are men, and men do better on “aptitude” tests such as the SAT-M; therefore, such tests must measure mathematical “aptitude” better than grades (where females do as well or better), so males have greater “aptitude” for mathematics, and this explains why a disproportionate

number of STEM professors are men. What is needed is a theoretically grounded measure of math aptitude that can be causally related to success in STEM careers. Although mental rotation has been the single best candidate in the literature, it is not theoretically grounded. It has not been shown that individual differences in mental rotation are any more important in fields where women are underrepresented than in other fields. In fact, mental rotation has not yet been shown to be a predictor of success in math-intensive fields, although it has been shown to predict math scores, as noted earlier.

Others have suggested that performance in the aptitude testing situation may draw on different motivational resources than performance in the classroom, with the former (SAT-M) drawing on self-efficacy and the latter (grades) drawing on mastery-based strategies (asking for help, preparing, doing homework). “As a consequence, the testing situation may underestimate girls’ abilities, but the classroom may underestimate boys’ abilities” (Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006, p. 22). Complicating this argument, however, is evidence that homework and preparation actually make an independent contribution to female students’ aptitude test performance, not only to their grades (e.g., Mau & Lynn, 2000). Even if it is found to be important, the SAT-M alone cannot tell us whether there are specific items that must be answered correctly by successful STEM scientists and, if so, whether they tap spatial abilities involved in mental rotation, or what variance in scientific success such skills account for when added to measures such as creativity, diligence, risk taking, other forms of mathematics skill, communicative ability, and preferences.

Another approach has involved looking at prior mathematical knowledge and strategies. Byrnes and Takahira (1993) showed that high school students’ prior knowledge and strategies explained much of the variance in SAT-M scores. No sex differences were found in amount of prior knowledge, but a sex difference was found for the conditional probability of getting an item right, if one had the prior knowledge and constructed an effective strategy: For male students, the probability was .91, whereas for female students it was only .72, suggesting that “female students spend too much time on individual items or are more likely to fall prey to misleading choices” (p. 809). Perhaps female students’ greater time reflects slower processing, less automated knowledge, or simply greater cautiousness, the latter being consistent with recent research we review later showing that stereotype threat leads to anxiety that impedes working memory and consequently lowers mathematics scores by approximately 11% (Beilock, Rydell, & McConnell, 2007). Females may be more cautious overall, an observation often reported in an older literature, and this may be at play here, too, although there is no direct test of this. However, Linn and Petersen’s (1985) meta-analysis of mental rotation suggested that one basis for sex differences is greater cautiousness of women, leading to women’s increased reviewing of choices and double-checking of mental transformations.

Researchers have investigated females’ underperformance on ill-defined problems or ones requiring unconventional solutions, which are examples of far transfer (Barnett & Ceci, 2002). Gallagher and DeLisi (1994; Gallagher, 1992) found that males excelled on SAT-M questions when solutions were not clearly defined and that females did equal or better on problems requiring familiar strategies learned in school (Gallagher et al., 2000). Fen-

nema et al. (1998) investigated antecedents in younger children and found that girls used more standard algorithms and concrete strategies, whereas boys used more abstract, invented algorithms, which are associated with being able to solve problems requiring flexible use of knowledge. To some, this difference in abstract thinking may be caused partly by differences in how teachers behave towards girls and boys (Fennema et al., 1998; J. R. Becker, 1981). However, in contrast to Gallagher et al. (2000, 2002), Harris and Carlton (1993) found that girls actually excel at abstract math items and that boys excel at applied items. Gallagher et al. (2002) applied a similar approach in their analysis of the GRE-Q. They found that it is possible to manipulate the performance of males and females by changing the mix of question types and criticized the lack of a theory-driven basis for the SAT-M and GRE-Q tests (see pp. 16–17).

Spelke (2005) has also suggested that the SAT is arbitrary; changes are sometimes made to its content, some of which favor male test takers and some of which favor female test takers. Drawing on Gallagher et al.'s (2002) findings, Spelke described changes made to items on the SAT-M that could have resulted in bias against women: Girls consistently outperform boys on items where one must determine whether the data provided are sufficient to solve the problem, yet these items have been removed from the SAT-M because they benefit from coaching. As Spelke argued, removal of items without a theory of the underlying construct can result in biases. It is unknown what each item on the SAT-M measures and how it relates to STEM success. Similarly, changes to the NAEP hinder interpretation of trends in sex differences. The most recent 2005 NAEP included more questions on algebra, data analysis, and probability and fewer on numeracy and used a different format (Mervis, 2007), any of which could affect sex differences and which preclude interpretations of temporal changes. Note that the mere existence of items favoring one sex is not evidence of bias unless the differences that result from their inclusion lead to irrelevant test difficulty, which unfairly affects one sex (Gierl, Khaliq, & Boughton, 1999). Similarly, Harris and Carlton (1993) compared male and female test takers with equal SAT-M scores to determine which items presented gender-specific difficulties. Consistent with some (but not all) research, male test takers performed better on geometry and trigonometry items and female test takers performed better on arithmetic/algebra items.

Reports of sex differences in math subskill areas is complicated by the level of the student (high school, college, graduate school) and the dependent measure (grades vs. test scores): Female students receive better math grades starting in middle school and continuing through college (Hong, O'Neil, & Feldon, 2005), and their advantage widens as the math becomes more advanced (calculus, elementary functions, probability theory) as opposed to less complex algebra, plane geometry, and trigonometry (Kimball, 1989); on the NAEP, geometry, mathematical operations, and number properties are areas of male superiority (Nations Report Card, 2007), with all other content associated with no sex differences. These sex differences were quite small, however. In contrast, female students excelled on items that were abstract or included variables such as X or $a(b)$, whereas male students did better on items embedded in applied contexts; the magnitudes of these differences were fairly large and led to the suggestion that males use math to solve everyday problems to a greater extent than

do females, supporting the argument that males perform better in math because they view it as more applicable.

Some data show that sex differences are apparent in solving speeded test problems but that when instructed to use the same strategies, males and females do so. In other words, females, while preferring certain strategies that are nonoptimal on speeded tests such as the SAT-M, can utilize optimal strategies as well as males can. This suggests that they have the cognitive ability but choose, for whatever reason, to use different strategies (Spelke, 2005). Training on appropriate strategy use can have large effects, again calling into question the nature of sex differences in performance. However, results are mixed, and Kimura (2007) suggested that adult sex differences are uninfluenced by training: Although both sexes benefit from short-term intensive training on spatial tasks, their scores do not converge (Baenninger & Newcombe, 1995; Sorby, 2001). (However, see Section 9 in the supplemental materials online for a cautionary note about the effect of training to narrow gaps.)

In sum, males usually score higher on mathematics aptitude tests (e.g., SAT-M), both as children and as adults, but the patterns of performance by type of question are complex, their relation to success in STEM fields remains unspecified, and their theoretical rationale is undetermined.

Right-tail differences. Several right-tail analyses that bear on these questions have been conducted by Lubinski, Benbow, and associates (Lubinski et al., 2006; Park, Lubinski, & Benbow, 2007; Wai et al., 2005). The SMPY program was one of the first major modern research efforts to focus specifically on the extreme right tail and conduct periodic follow-ups (e.g., Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Park, Lubinski, & Benbow, 2008; cf. Wise et al., 1979). Young adolescents were screened for admission (the cutoff was scoring above the 97th percentile in a classroom math test and then scoring > 700 on the SAT-M at age 13 years, which occurs at the rate of 1 in 10,000). There was a higher proportion of boys who made the > 700 cutoff. At the end of high school, these students took the SAT-M again, and again there was a preponderance of boys at the right tail (Benbow & Stanley, 1980, 1983). Because the initial difference occurred before students began to select their courses, and because there were few sex differences in attitudes toward math, it was suggested that the sources of sex difference were probably, in part, biological (Benbow, 1988; Benbow & Stanley, 1980; see also Pinker, 2002). This assumes that the only sex-differentiated sociocultural variable that could explain such differences is course choices, which may not be the case, as there are other potential non-biologically driven experiential differences, as discussed earlier.

Several SMPY longitudinal analyses contrasted students who scored in the top quartile of the top 1% of the SAT-M with those who scored in the bottom quartile of the top 1%. When assessed 20 years later, the former received significantly more PhD degrees in science, were more likely to be tenured STEM professors at top universities, and had more inventions/patents. This suggests that successful STEM scientists disproportionately come from the very top of the SAT-M distribution, assuming that the top and bottom quartiles of the top 1% were equally motivated to rise to the top of STEM professions, took equal numbers of math classes; came from similar SES backgrounds, and were subject to similar cultural beliefs and pressures.

We lack a comparable analysis of school and college test performance on which females excel. We have found no prospective studies analyzing the future careers of students at the right tail based on scores on classroom tests. Perhaps such school-based measures can also predict future success. Without them, it is difficult to resolve the discrepancy between school grades and SAT scores.

In contrast to these longitudinal analyses, Shalala et al. (2007) argued that sex differences in aptitude cannot explain the underrepresentation of women in STEM careers and that aptitude scores do not predict success in STEM careers:

Notably, it is not just the top SAT scorers who continue on to successful careers; of the college-educated professional workforce in mathematics, science, and engineering, less than one third of the men had SAT-M scores above 650, the lower end of the threshold typically presumed to be required for success in these fields. The differing social pressures and influences on boys and girls appear to have more influence than their underlying abilities or their motivations and preferences. (Shalala et al., 2007, p. 24)

This assertion is based on analyses by Weinberger (2005). However, it is important to add a qualifier Shalala et al. mentioned later: The careers analyzed by Weinberger were not exclusively, or even primarily, professorial or high level; they included master's-level technical workers in laboratories. There is no evidence in her analysis that the SAT-M scores of two thirds of individuals holding professorial positions in math-intensive fields are below 650. Although Weinberger reported that White women enter these fields at half the rate of men with the same math test scores, this could reflect women's reluctance to work at laboratory technical jobs rather than a barrier preventing them from doing so. We are not claiming that the SAT-M scores of professors are 3, 4, or 5 *SDs* above the mean, as Summers (2005) opined; we are simply noting that it is misleading to imply that two thirds of them have SAT-M scores below 650, because they almost surely do not. (See the Cornell GRE-Q data in Figure C of Section 10 of the supplemental materials for an explanation of why this oft-repeated claim is probably incorrect.)

Recently, research has examined possible causes of the disjunction between school grades and aptitude tests by examining specific SAT-M items on which the sexes differ. One approach has been to explore the contribution of specific abilities such as mental rotation to overall SAT-M scores. Casey, Nuttall, Benbow, and Pezaris (1995) found that, although mental rotation scores were predictive of overall SAT-M scores for highly math-talented younger females, they were not predictive for highly math-talented younger males. Casey and her colleagues (Casey et al., 1995, 2001) with a less select, though somewhat right-tail-oriented group—the top third of the college-bound sample—found that roughly two thirds of the gender-SAT-M relationship was explained by mental-rotation ability and that one third of the relationship was explained by mathematics self-confidence; thus, both are highly important. These researchers also found that most of the mediational effects of mental rotation and self-confidence were not explained by grades (Casey, Nuttall, & Pezaris, 1997): Mental rotation appears to be tapping different abilities. Thus, mental rotation may be important for most students, but perhaps not for the very top male students who comprise the STEM subset. Relatedly, Webb, Lubinski, and Benbow (2007) found that more than

half of the top 1% of spatially gifted students are not identified when screening is conditioned on the top 1% of math ability. Clearly, spatial ability is complex, and more research is needed before it can be claimed to be a precursor of the male overrepresentation in math-intensive STEM careers.

Gallagher and DeLisi (1994) followed up their finding that girls do better on well-defined problems and boys outperform girls on ill-defined problems. The authors found that among right-tail students, the use of conventional strategies, which are good for solving well-defined problems, was correlated with negative attitudes toward math. Gallagher and DeLisi suggested that female students use conventional strategies because their lack of confidence in math discourages them from experimenting. However, given that the sexes achieved their high scores by answering different types of questions correctly, it could be that girls' lack of confidence stemmed from lack of the deeper understanding that permitted boys' experimenting to extend algorithmic knowledge.

A number of investigators have controlled for differences in mental rotation ability, with the result that the male advantage on the SAT-M disappears (Burnett, Lane, & Dratt, 1979; Casey et al., 1995). Conversely, when mathematics aptitude scores are covaried out of mental rotation scores, the male advantage on mental rotation remains for right-tail samples. Such findings have led some to conclude that sex differences in mental rotation mediate sex differences in advanced mathematics (Casey et al., 1995). Recently, Casey and her colleagues have reviewed the evidence for a causal connection between spatial and mathematical ability, noting that it is most evident among older children when the mathematics content is less focused on numerical learning and more focused on geometry and problem solving (e.g., Kersh, Casey, & Young, in press; see also Battista, 1990; Delgado & Prieto, 2004; Friedman, 1995). Kersh et al. (2008) noted that:

Battista (1990) found that among high school students, scores on mental rotation were significantly correlated with geometry performance, and related to students' choice of geometry problem-solving strategies. Delgado and Prieto (2004) found that, in a sample of 13-year-old students, mental rotation ability added significant unique variance to performance on geometry and word problems, but not arithmetic. Finally, in a sample of 6th grade boys, Hegarty & Kozhevnikov (1999) found moderately high correlations between math achievement and scores on two measures of spatial ability, one a measure of spatial visualization and the other, mental rotation. (p. 6)

These analyses are the closest approximation of a construct validation in the literature, one that elevates mental rotation to explain sex differences in math aptitude. Again, however, the relation of such skills to success in careers in STEM fields is unknown.

Mental rotation can be measured in a number of ways, the most common being the original three-dimensional perspective figures of Shepard and Metzler (1971), in which participants must judge whether pairs of perspective objects are the same or mirror reversals or, in an alternative variation, whether a target figure is matched with a pair of rotated figures. Other mental-rotation figures are also commonly used, including the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, & Harman, 1976), the Vanderberg test (Vanderberg & Kuse, 1978), Thurstone and Thurstone's (1962) Primary Mental Abilities test, the Mental Cutting Test (College Entrance Examination Board, 1939), the Purdue

Spatial Visualization Test—Revised: Rotations (Guay, 1977), and the Block Design subtest from the Wechsler Adult Intelligence Scale (see Sorby, 2001, for examples of different tests). The results from these various measures are not always consistent (see Linn & Petersen's, 1985, meta-analysis showing the lack of consistency across the Primary Mental Abilities test and Vanderberg and Kuse's measures), and others have shown that making small scoring adjustments and adding items can alter results significantly (Peters et al., 1995). There may be sensible reasons for the divergence of results in response to different scoring methods (e.g., whether 20 or 40 rotations are used), but collapsing across such differences may account for the inconsistencies noted above.

To summarize, we find evidence that males and females score differently on mathematical and spatial ability tests and evidence suggesting that differences on ability measures are related to success in STEM fields. However, the causal relationship between these two observations is unclear.

Assessed Performance

As with any competence–performance relationship, the mapping from abilities to assessed performance in STEM-related abilities is not straightforward and may be moderated by other factors. Thus, aside from issues of whether the abilities described are relevant to STEM careers, the evidence discussed above may be an inaccurate representation of true ability differences. Sex differences can be affected by superficial aspects of the testing situation, including format, content, instructions that make gender salient (stereotype threat), and by societal bias. Next we review these sources of error in performance assessments.

Test Format and Context

Assessment of abilities can be affected by superficial aspects of test content and context, yielding systematically misleading performance measures. All of this evidence comes from studies of unselected samples rather than from studies of right-tail groups, although several studies have reported both right-tail and mean differences separately (see Stumpf & Stanley, 1998). In some cases, differences are smaller when testing in a group rather than an individual setting (Voyer et al., 1995), and, if individual testing is used, when the tester is female rather than male (McGuinness & Morley, 1991). Findings of social context sensitivity of sex differences are widespread in other areas as well (see Hyde, 2005, for examples). The temporal context may also affect mean sex differences. For example, tests in which there is time pressure (e.g., mental rotation, SAT-M) tend to result in a larger male advantage than when speed is not crucial (e.g., tests of spatial visualization, school grades; e.g., Linn & Hyde, 1989; Spelke, 2005). Aspects of modality also matter: Stumpf and Stanley (1996) and others have found that boys do better than girls on multiple-choice measures but not on free response measures in some subjects. Thus, sex differences may be greater on the SATs than on AP exams because SATs are composed of multiple-choice items, whereas AP exams are only half multiple choice (see also Kimball, 1989).

The content of the measures can also affect findings. For example, Vasta, Knott, and Gaze (1996) found that training eradicated sex differences on a water line problem when the outcome involved getting the correct answer but not when it involved

verbalization of the correct principle. Some variables are particularly sensitive: For example, in a meta-analysis, Voyer et al. (1995) reported that the effect size for mental rotation was dependent on the particular scoring method used. When the test was scored out of 20 (original method), sex differences were larger than when scored out of 40 ($d_s = 0.75\text{--}1.00$ vs. $0.50\text{--}0.74$, respectively); other scoring methods reduced the effect size further ($d = 0.10\text{--}0.19$). Although such differences may be reasonable in terms of measurement issues, the issue becomes problematic if researchers collapse across studies with different scoring methods. In their meta-analysis, Linn and Petersen (1985) showed that the effect size for gender depended on the particular rotation test used, with the four most common tests yielding significant inconsistencies.

Stereotype Threat

Many researchers have examined the impact of another subtle aspect of the testing situation by manipulating the mindset of the test taker with respect to cultural beliefs associated with gender. We focus here on stereotype threat, because it is the most actively researched topic in this category (see Ben-Zeev, Duncan, & Forbes, 2005) and is related to other influences on the testing situation (e.g., status characteristics theory; see Correll, 2004). Section 11 of the supplemental materials explains the reasoning behind stereotype threat, and Table 7 (also in the supplemental materials) summarizes this growing literature, which, although fairly consistent, raises intriguing questions. Below we merge the mean and right-tail studies rather than separate them because most fall on the borderline between mean and right tail, unlike other sections where there are often clearer differences.

Some researchers attribute the gender gap in mathematics, in part, to negative stereotypes that are activated when gender is salient (Lewis, 2005). Female test takers who marked the box corresponding to their gender after completing the SAT Advanced Calculus test scored significantly higher than those who checked it before starting. Identifying gender after the AP examination rather than before was thus predicted to add nearly 4,700 women eligible to begin college with advanced credit for calculus (Danaher & Crandall, 2008), presumably because directing attention to gender at the start of the examination causes anxiety that impedes working memory and hence performance (Beilock et al., 2007; Schmader & Johns, 2003).

In a recent study, Murphy, Steele, and Gross (2007) presented Stanford University undergraduate science majors videos that were either 50% male, 50% female or 75% male, 25% female. The latter resulted in the female viewers exhibiting identity threat, in which their heart rate and skin conductance were elevated vis-à-vis women who watched the 50%–50% video. Interestingly, the threatened women remembered more items related to the threat situation (other science paraphernalia) than did nonthreatened female peers or male viewers. Murphy et al. suggested that a threatening situation causes increased vigilance to threat-related objects. Thus, anxiety is not always associated with memory impediments. In addition to anxiety-related factors responsible for stereotype threat, especially the impediment to working memory (Beilock et al., 2007; Martens, Johns, Greenberg, & Schimel, 2006; Schmader & Johns, 2003), Correll (2004) has provided data on other mechanisms, contrasting different theories, such as human capital and status characteristics. However, more research is needed to explain

why in some studies the non-stereotype threat condition led not only to elevated female performance but also to a reduction in male performance and why in other studies the group advantaged by the stereotype experienced a boost over its non-stereotype level (e.g., Shih, Pittinsky, & Ambady, 1999).

Krendl et al. (2008) have shown that when women are in the stereotype threat condition, their underperformance in mathematics coincides with increased neural activity in part of the affective network involved in processing negative social information, the ventral anterior cingulate cortex (vACC). These authors showed that stereotype threat causes increased vACC activity that coincides with underrecruitment of brain regions involved in solving math problems, such as the angular gyrus, resulting in math underperformance (Krendl et al., 2008). Stricker and Ward (2004; see also Stricker, 2006), however, have argued that the performance effect is small and of no practical significance. They tested students taking the AP Calculus AB examination who were asked about gender and race either before or after the test. Using the standard small effect size accounting for 1% of variance ($d = .20$), Stricker and Ward failed to find effects of inquiring about ethnicity and gender on performance. In response, Danaher and Crandall (2008) reanalyzed these data, using slightly less stringent effect sizes and alpha levels and argued that such effect sizes are meaningful, with nearly 6% additional female students and 4.7% fewer male students achieving a passing score when sex and race were marked after the test rather than before. Danaher and Crandall underscored their prior claim that the result would be nearly 4,700 more female students starting college with AP-Calculus credit, a claim that was disputed by Stricker and Ward (2008) in a rejoinder. (This effect is represented in Figure 1 by the gatekeeper test results arrow.)

Stereotype threat findings do not imply that all sex differences in math performance can be eradicated by such manipulations. However, they do suggest that stereotype threat factors can undermine some women, perhaps especially mathematically talented ones. If cultural beliefs about male superiority are responsible for stereotype threat, and if such conditions are a major cause of women's underrepresentation in STEM (which we do not suggest), then male overrepresentation should be greater in countries not known for their egalitarian gender beliefs, such as Turkey and Korea, than in the United States and the United Kingdom, and this is generally found to be the case (see analyses by Guiso et al., 2008): The math gender gap disappears among 15-year-olds in countries viewed as highest on gender equality, in which respondents answer negatively to such questions as "When jobs are scarce, men should have more right to a job than women." Countries high in equality, such as Iceland, Sweden, and Norway, have virtually no math gap, even at the extreme right tail—scoring above the 99th percentile—whereas countries such as Turkey that rank very low on gender equality have math gaps in favor of males. However, seemingly incompatible anomalies exist, for example, the fact that there are proportionately twice as many female computer science majors in Turkey as in the United States (Charles & Bradley, 2006). Moreover, Penner (in press) shows that the odds ratio of being female at the right tail of the mathematics distribution is greater (by a factor of over 2) in countries not viewed as egalitarian, such as Hungary and Russia, than in egalitarian countries such as Norway, Denmark, and Sweden. Moreover, given that girls' performance in high school mathematics now matches that

of boys; that female students take and pass as many advanced courses in mathematics and science as boys; and that the gender gap at the right tail of the mathematics distribution appears to have been shrinking since 1975 (Hyde et al., 2008), it is difficult to know why female students still underperform in the United States (Correll, 2004) or why confusing problems take a greater toll on their performance than on that of male students. At what point will gender stereotypes fade, or have they already, as hinted by some evidence that the gender stereotypes on which stereotype threat is based are fading (Biek, 2006; Martens et al., 2006).

Further, it is not clear why, when asked how well they would need to do on a test to pursue coursework and careers in that area, male and female participants in the non-stereotype threat condition reported similar levels of test performance to each other as well as to male participants in the stereotype threat condition. The sole group reporting a need to score much higher was the group of female participants in the stereotype threat condition (e.g., Correll, 2004). Thus, being exposed to experimental feedback about alleged male superiority in a field does not lead males to underestimate how well they would need to do on a test to pursue future coursework or careers any more than it does males and females who were not exposed to such feedback, but it does lead females to overestimate the scores that they would need. One puzzling aspect of the stereotype threat findings is why the anxiety from self-evaluative threat in the presence of implicit negative stereotypes has not resulted in female students actually learning less mathematics than male students or doing poorly on school achievement tests; further, the question remains, why has male students' persistent underperformance on grades not led them to disidentify with math? As Davies, Spencer, Quinn, and Gerhardtstein (2002) pointed out, the defensive detachment resulting from stereotype threat should not undermine only women's short-term performance but also their long-term learning and aspirations. If a female student taking the SAT-M "has an extra worry to contend with that the boy does not, resulting in... a process that takes focused concentration and attentional resources" (D. M. Quinn & Spencer, 2001, p. 59), then why does this anxiety, and the diminished cognitive capacity that results, not lead to less mathematics learning in middle school when gender stereotypes become pronounced? How come female students are able to surmount such stereotypes to the point where they not only take as many advanced math courses as male students but receive better grades in them? Are their superior grades a result of being more compliant (doing homework, asking for help, being prepared), or is there a more complicated opponent process model to explain such perplexing issues? Moreover, why is it that at a young age, when they are first sensitive to stereotype threat (see Good, Aronson, & Harder, 2008), do female students not begin to "disidentify" with the domain of mathematics, that is, reconceptualize their values and identity to avoid stereotype threat by removing math-like activities as a basis for self-assessment (Inzlicht & Ben-Zeev, 2000)? Finally, if women and African Americans score higher when stereotype threat conditions are removed from the SAT-M, then they must have mastered the material prior to the administration of the SAT-M but suffered when the testing context interfered with the retrieval of this learning. If retrieval-time interference is the causal mechanism, then why does the SAT-M overpredict African American's college mathematics performance but underpredict female mathematics performance? Does the

mechanism operate differently when activated by gender versus racial stereotypes? Morgan and Mehta (2004) concluded that racial gaps in achievement are not easily explainable in terms of Black disidentification, and a similar argument could be made for sex differences.

Further investigation is needed to understand the mechanisms and extent of stereotype threat and how it can be alleviated. Assuming that students are aware of gender stereotypes about math (Good et al., 2008; cf. Biek, 2006; Martens et al., 2006; Raymond & Benbow, 1986), would female students attending single-sex schools be less affected by such stereotypes? Would same-sex testing sessions reduce the effect, as some suggest (e.g., McGuinness and Morley, 1991; Inzlicht & Ben-Zeev, 2000)? Meta-analysis of the stereotype threat literature is complicated by the variety of sample restraints and controls that are used: prior math achievement, math identification, gender identification, awareness of negative stereotypes, item difficulty, and combinations of these, to name a few. Although stereotype threat may play a role in explaining sex differences in performance on some sorts of tests, and therefore may slightly reduce the number of women qualifying for advanced study in STEM fields, it is unlikely to be a primary cause of sex differences in representation in careers in these fields. We consider it a minor, reasonably well documented, but poorly understood effect.

Evaluative Bias

At any given ability level, assessed performance may also be affected by cultural expectations in the form of bias, as depicted in Figure 1. Evidence of possible evaluative bias in some fields exists, although it is unclear whether this applies to math-intensive fields. Studies have provided evidence that career performance is sometimes evaluated differently in men and women, though it is often unclear whether this is unfair because controls are not possible to assess base rates (Trix & Psenka, 2003). Conducting such research is difficult, because of the private nature of evaluation, so there are fewer studies available (see Table 7 in the supplemental materials). All of the studies in this area concern the right tail, focusing on faculty, postdocs, or graduate students.

One striking piece of evidence comes from a Swedish study of peer review of postdoctoral fellowships in medical fields (Wenneras & Wold, 1997). The authors claimed profound discrimination against female postdoctoral applicants. Their conclusion was based on analyses of the scores by review committees, which were compared with objective data (e.g., total publications, first-authored articles, citations). Reviewers judged each application on three measures. Wenneras and Wold (1997) found that the translation of objective data into subjective scores was highly biased against women such that “a female applicant had to be 2.5 times more productive than the average male applicant to receive the same competence score” (p. 342). The authors showed that the most productive female applicants—those with 100 or more impact points (a measure of number of publications adjusted for citations)—composed the only group judged to be as competent as men, although only as competent as the least productive men, who had fewer than 20 total impact points. Wenneras and Wold’s (1997) study has garnered wide attention on account of its striking findings and the prestigious journal that published it (*Nature*). However, their conclusion of bias is open to alternative interpre-

tations. They analyzed 114 postdoctoral fellowship applications in 1995 to the Swedish Medical Research Council, 62 submitted by men and 52 by women. A total of 16 men were funded (25.8%) versus only 4 women (7.7%). With such small samples, tiny adjustments matter. For instance, what if the review panels were disproportionately composed of biomedical (not nursing or basic science) professionals and more of the male applicants came from medical or biomedical backgrounds? Only 27% of the female applicants were from medical backgrounds, whereas 60% of male applicants were from such backgrounds. One need not posit sex discrimination if reviewers preferred medical journals to basic science or nursing journals, even if the latter were more highly cited; reviewers were not given the impact ratings of the journals. If such preferences were operative, then the greater proportion of female applicants from nursing (12% vs. 3% male applicants) could have tilted the odds against them. Along these same lines, perhaps reviewers weighted factors such as sole-authored papers more than first-authored papers, and perhaps male applicants had more of these. Finally, the regression models described in Wenneras and Wold’s article entered each productivity variable alone rather than allowing for multiple variables to enter, as might be expected in the real world, a point noted by Sommers (2008). Efforts to obtain the authors data for reanalysis have been unsuccessful.

Two large-scale analyses run counter to the conclusions of Wenneras and Wold (1997). The first of these was a study by RAND Corporation commissioned to assess gender bias in grant awards at the National Science Foundation, the NIH, and the U.S. Department of Agriculture; the second analysis was commissioned by the Australian Science Foundation. The RAND study (RAND, 2005) concluded that there was no gender bias in the awarding of grants at the three federal agencies (as well as in two surveys), with the exception that men received more money for their grants. However, owing to a lack of data for possible controls, it is not clear what this single exception means. As far as the percentage of grants funded, there were no gender gaps at any agency.

The Australian Research Council processes over 3,000 grant applications annually in all areas of science, each reviewed by 4.3 reviewers, on average, resulting in 6,233 external reviewers, many of whom review multiple applications. Jayasinghe, Marsh, and Bond (2003, 2006) published several analyses of these data, using sophisticated measurement frameworks. This is an excellent basis upon which to examine bias against women because not only can one examine the fate of female applications but one can do so as a function of many potentially confounding variables, including the gender and field of the reviewers.

Marsh, Jayasinghe, and Bond (2008) reported that although only 15.3% of the applicants for grants were female, their success was proportional (15.2%). When gender of only the first-named investigator was considered, the success rate was 21% for both men and women. Detailed analyses on second- and third-named researchers also indicated an absence of sex differences in the success rate. (Supplemental analyses based on the mean of external ratings and the final panel committee ratings showed similar results; Jayasinghe et al., 2003.) Furthermore, the insignificant effect of gender was true for all nine social science, humanities, and science disciplines. Finally, Marsh et al. found no evidence of a gender bias as a function of the sex of the reviewers or the sex of the applicants. Similarly, Leboy’s (2007) analysis of success rates at

NIH for men and women revealed identical success rates for new submissions (both 18%) and similar rates for competing renewals (33% vs. 34%, favoring men). Recent evidence also suggests that the promotion process of women in academic science departments is now similar to that of men (with equal promotion rates; Ginther & Kahn, 2006; Mason & Goulden, 2004).

A study of search committee recommendations for hiring of psychology assistant professors by Steinpreis, Anders, and Ritzke (1999) gets closer to the conclusions of Wenneras and Wold (1997). Steinpreis et al. asked 238 psychologists to review fictitious assistant professor candidates and more advanced job seekers who were eligible for tenure. They used the same curriculum vitae (CV) but varied the sex of the applicant and found that reviewers favored the male assistant professor CV, although they did not favor men for the advanced job eligible for tenure. Thus, even though the male and female CVs were identical, assistant CVs labeled "male" were rated higher. Interestingly, the female reviewers also preferred the fictitious male CVs. Thus, bias in the system did not appear to be gender-specific to rater.

A similar finding emerged from a study in the field of music that found that nonblind auditions (those in which the jury could see the applicant) for positions in orchestras discriminated against women (Goldin & Rouse, 2000). In most cases studied, more women were selected when auditions were blind (a screen was placed between the jury and applicants), even with the same sample of players. Although this study took place in a nonacademic field, it provides further support for bias in supposedly gender-neutral evaluation. Research has also shown that women working as part of a business team receive less credit than men even for identical work. For tasks that are seen as stereotypically male (Heilman & Haynes, 2005), if there is ambiguity about the true quality of the woman's contribution to a joint task, it is downplayed. Both male and female judges rated a hypothetical worker's performance worse when they thought the worker was female, even though the description of the task, performance, and so forth was identical. This study used a business context but might easily apply to evaluation of team contributions in other fields, such as laboratory meetings in which ideas for experiments are developed. Although these studies showed that women can be rated lower than men who perform equally (downgraded by female raters as much as by male raters; see Rhode, 1997, for additional evidence), they did not concern the hiring and promotion of women in STEM fields, raising the hope that such biases are not as prevalent in these fields, clearly a hope in need of testing. Finally, Budden and her associates (Budden, Lortie, et al., 2008; Budden, Tregenza, et al., 2008) analyzed publication acceptance rates for women for the journal *Behavioral Ecology* following the start of blind peer review, noting that acceptances of female first-authored papers went up 7.9% in the 4 years following blind review compared with the 4 years prior to its onset. Webb, O'Hara, and Freckleton (2008), however, argued that the increase in women's acceptances was observed in the decade prior to blind reviewing and in other journals that never initiated it.

In sum, there is some suggestion of bias against female job candidates in psychology, but no direct evidence in the math-intensive fields where women are most underrepresented, and conflicting evidence regarding bias against female grant applicants. Also, Ginther and Kahn's (2006) data suggested that to whatever extent there was bias against female job and tenure

candidates in the past, there appears to be none today, with similar promotion and tenure rates.

Conclusion and Discussion

Returning to the framework in Figure 1 and the hypotheses represented by Figures A1, A2, and A3 in online Section 1, where do we stand regarding this large, often inconsistent, and at times diametrically contradictory literature? In this final section we evaluate the various alternative models and modify the framework in light of the strength of the evidence for each factor, boldfacing nodes for which there is evidence of a substantial effect and varying the width and darkness of each arrow to be congruent with the importance of each link and the convincingness of the evidence (Figure 5). This modification inevitably has a subjective element, so we base our judgment on the consistency of evidence, as well as the statistical significance, sample size, and effect sizes of individual studies.

There is considerable evidence supporting a causal role for most of the circle of variables. Broad contextual expectations and resources (consistent with the sociocultural model in Figure A3 in the supplemental materials) clearly affect the performance of males and females differently, presumably largely because of differences in abilities that develop as a result of aspects of motivation/interests and activities that are driven by proximal processes in the immediate environment. Evidence for the causal role of broad cultural expectations comes from the significant variability in sex differences between countries, social groups, and cohorts. A direct effect of biological sex on brain development, and hence mathematical ability (see biological models, Figures A2 and A3 in the supplemental materials), would not show such variability; there would be no compelling explanation for why, if the cause of sex differences in math was biological, two countries with similar gene pools would exhibit such different patterns of sex differences (e.g., Guiso et al., 2008). The proximal sociocultural variables in online Figure A3 that mediate this effect are, however, unclear, as studies of teacher and parental sex discrimination are inconclusive, and egalitarian attitudes are not always associated with narrowing of sex differences and are sometimes oppositely associated with narrowing (Charles & Bradley, 2006; Penner, in press). Notwithstanding this contextual variability, male test takers currently score at the right tail more often on many math tests in many cultures. A small part of this may be due to stereotype threat and bias on some kinds of evaluations, but the remainder probably reflects ability differences. These ability differences may, in part, depend on differences in experience such as a (biologically or environmentally driven) male preference for certain types of play and activities with objects compared with a female preference for people-oriented ones (see online Figures A2 and A3) and/or they may be due to males with high math abilities having fewer non-mathematics skills, hence devoting more effort to mathematical endeavors (see online Figures A2 and A3), or they may be due to hormonal or other biological differences (see online Figures A2 and A3). Evidence for all of these hypotheses is currently inconclusive, although this does not gainsay the evidence that more males score at the extreme right tail, despite narrowing over time that has sometimes but not always been observed.

Evidence is strong for sex-differentiated motivation to devote time and energy to a STEM career, that is, the life choices of

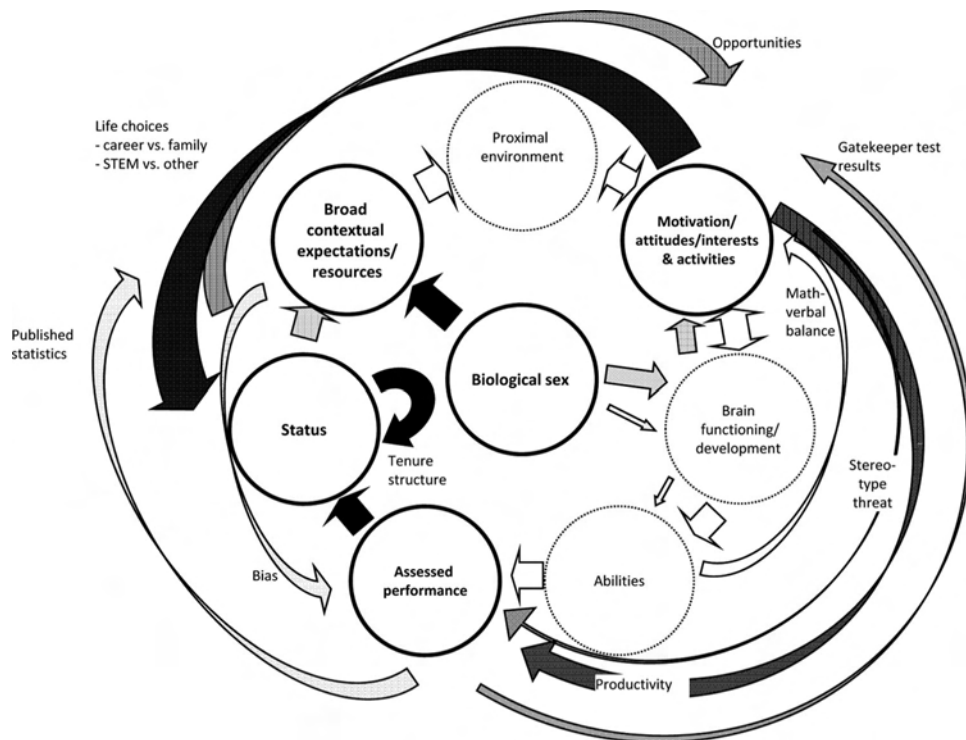


Figure 5. Evidence-based causal model. Nodes and links with stronger evidence of role in explaining sex differences in STEM (science, technology, engineering, and mathematics) professions are presented in boldface. Links are broad/narrow depending on importance.

women versus men (see online Figures A2 and A3). A number of surveys indicate that women with children devote somewhat less time than men to their career because they are expected to devote more time than men to family matters, and career progress suffers accordingly while children are young. Women with high math ability also choose non-STEM fields more often than men with high math ability, and they also drop out of STEM fields—especially math and physical science—at higher rates than men (Strenta et al., 1994), particularly as they advance (Preston, 2004). Further, although it applies equally to non-math-intensive careers, Hakim’s (2006) survey suggests that even very educated women are more likely than men to favor home-centered lifestyles and adaptive lifestyles, wherein family and home are paramount and work is adapted to fit around this choice. Because of a combination of these factors, fewer women than men enter STEM fields and remain in them long enough to reach the top. Thus, institutional barriers and stereotypes, both of which are real, do not appear likely to account for most of the sex differences, nor does outright discrimination against women in hiring and remuneration. To the extent that such barriers and biases operate to decrease the entry and retention of women in math-intensive fields, there is no compelling evidence that removal of these barriers would result in equalization of sex ratios, given the evidence that women’s lifestyle choices, societal expectations associated with child rearing, and career preferences tilt toward other careers, such as medicine, teaching, law, and veterinary medicine, over engineering and physics.

Evidence presented supports each prong of this argument. For example, the effect of contextual expectations and resources in

mediating the effect of biological sex is indicated by the sensitivity of sex differences to country, cohort, and social group: The magnitude of the male advantages at the right tail fluctuates around the world, including a number of countries where females are superior. It is also likely that the career status—the prevalence of each sex in different careers and at different levels—affects cultural expectations, although the evidence for this is not presented here. Cultural expectations are also likely to be affected by the published statistics regarding male and female performance (e.g., Correll, 2004; Dweck, 2007). Conversely, cultural expectations may also bias some assessments, for example, job candidate evaluations. However, such bias has not been demonstrated in math-intensive fields. Notwithstanding some evidence that women may be the victims of unfair evaluation in hiring, grant proposal awards, and salary and promotion reviews, the best evidence in this area indicates that grant applications are not influenced by gender (Jayasinghe et al., 2003, 2006; Leboy, 2007), and the best evidence on salary and hiring (e.g., Ginther & Kahn, 2006) indicates that gender differences are small (or nonexistent) among younger faculty and that hiring biases are not obvious. Much of the evidence of discrimination in the proximal environment is dated or anecdotal and not compelling as an explanation of why women are underrepresented in math fields, and especially of why fewer high school and college women express interest in these fields.

There are multiple pathways in Figure 5 between motivation/interests/activities and assessed performance. Activities such as publishing are reflected in productivity differences, which affect assessed performance, as they should (although there remains the question of whether there are gender differences in aspects such as

the relationship between research contribution and authorship status). The other connection from interests/activities to assessed performance—stereotype threat—is undesirable. Evidence for a small negative effect on test performance of awareness of a stereotype, coupled with identification with the domain, exists for some types of tests. This may adversely affect how abilities translate into assessed performance, although it is unclear how stereotypes diminish girls' performance on the SAT-M but not on measures contributing to high school and college math grades. An observer from another planet might wonder why boys do not acquire negative stereotypes about their math ability after years of witnessing girls outperform them in class. Further, it is not obvious that more recent cohorts endorse these stereotypes about gender and math. Finally, if women were found in STEM professions commensurate with their scores on SAT-M, there would be many more women in such professions; for example, Hyde et al. (2008) reported a ratio of boys to girls at the top 1% of mathematics performance of 2.06, that is, roughly two thirds of the top 1% are men. Thus, if engineering programs required the top 1% of math scorers, one would expect one third of engineers to be women, but only 15% actually are.

In the opposite direction is the effect of test performance on interests/activities, that is, gatekeeper test results. This arrow in Figure 5 refers to the fact that performance on some tests is used to limit opportunities to pursue further academic avenues. For example, SAT scores partially determine access to colleges, and GRE-Q scores influence entrance into STEM fields. Test performance also affects interest in math and science and is the most important cognitive factor predicting women's defection from undergraduate STEM fields, specifically, poor grades in science courses and lower SAT-M scores (Strenta et al., 1994).

Reciprocal arrows are also depicted in Figure 5 between motivation/interests/activities and career status. The life choices pathway, moving from interests/activities to career status, indicates the effect of women's life choices on their career status; for example, not applying for tenure-track jobs or looking for part-time work out of a desire to manage the conflicting demands of work and family. This effect is both powerful and well documented (e.g., J. A. Jacobs & Winslow, 2004; Mason & Goulden, 2004), and as seen, the penalty associated with having children early in one's career is greater for women than for men. A great deal of survey data accord with Scheibinger's (1987) view that some marital patterns discourage women from remaining in STEM careers (Stanford University, 2005, p. 1). Recently, Rudd, Morrison, Picciano, and Nerad (2008) analyzed those with PhD degrees in the social sciences 5 years out at Carnegie doctorate-granting institutions and found that 34% of female academics had a partner who also had earned a PhD degree, which is double the rate for male academics. This resulted in 25% of women changing jobs because of their partners' job moves versus only 14% of men changing jobs for their partners' job moves. Additionally, a survey of top graduate students revealed that three times more women desire to work part-time for a limited period of time, with 31% responding that this was "important" or "extremely important," compared with 9% of men (Lubinski et al., 2001). Moreover, as Preston (2004) has shown, women are more likely than men to exit STEM careers, especially as they advance. To some, this state of affairs suggests a problem that can be countered only through societal engineering of outcomes such that career success is no longer defined in terms

of long workweeks and conference travel that are incompatible with family needs (see Sommers, 2008). However, tampering with the current laborious work life of academics could have negative consequences for productivity and national needs and would require careful and sustained research before enactment.

The opportunities arrow, in the reverse direction, reflects a similar effect to the gatekeeper test results arrow, namely, illustrating that career status affects access to research opportunities. Although expected, this could unnecessarily reduce women in STEM fields inasmuch as a small, initial sex difference in career status permanently precludes activities after the reason for it has passed.

Finally, the tenure structure loop links the career status bubble to itself, reflecting the impact of the academic career structure: the tenure process. Academia has a rigid career path: The way to become a full professor is to be an associate professor, and the way to an associate professor is to be an assistant professor, so there is usually only one opportunity to enter the pipeline. Switching from a non-tenure-track post to a tenure-track one is seldom possible, irrespective of productivity, yet the tenure track is the gatekeeper for allocation of resources and opportunities. Greater flexibility might open opportunities for women and men unwilling to adhere to this traditional model. In a recent editorial, Hamel et al. (2006) argued that the changing gender composition of the faculty pipeline may, in the future, force a relaxation of the tight schedules for tenure and promotion reviews, noting that promotion criteria and timelines

require academic productivity unattainable without devotion of most waking hours to career activities, leaving little time for family and other priorities. This approach may prove untenable in the future, as women make up an increasing portion of the physician pool and as many male physicians take on more responsibility for child rearing and want more time for personal life. (p. 303)

These arguments apply equally to nonmedical fields.

Although biological sex could theoretically affect many of the variables in the outer circle, the evidence speaks primarily to its direct effects on cultural expectations and on the brain (leading to effects on motivation/interests/activities and abilities). That cultural expectations are influenced by biological sex (Figure A3 in the supplemental materials) is beyond dispute in many parts of the world. Whether the direct effects of biological sex on the brain, and hence on motivation/interests/activities and abilities, are a major contribution to the dearth of women in math-intensive fields, however, is unclear, hence, the paleness of the arrows marking these pathways. For example, if women are more home-centered and innately more interested in raising children (Hakim, 2006), then biological sex feeds directly into brain development/functioning and hence into motivation/interests/activities, which in turn affect career status through life choices (Figure A3 in the supplemental materials). It is possible, however, that any direct biological impact is small, with the bulk of the gender-interests relationship indirect, mediated by factors such as cultural expectations (Figure A3 in the supplemental materials). With regard to the link between biological sex and abilities (Figure A2 in the supplemental materials), also mediated by an effect on brain functioning and development, findings such as sex differences in white-gray matter ratios are promising but have not been causally linked to male advantage in three-dimensional mental rotation,

thought to underlie male superiority in advanced mathematics. Similarly, the relationship between testosterone and spatial abilities is unclear, with much evidence suggesting a causal role but other evidence calling this into question (see Puts, et al., 2008), hence the lightly shaded arrows connecting biological sex, brain development, and abilities. While not dismissing the possible causal role of biological factors in sex differences in STEM careers, we find it worth noting that in the past, biological hypotheses have often been believed out of proportion to the evidence supporting them (see Section 12 in the supplemental materials for examples). We hope that future research will converge on a clearer hormonal influence (or lack thereof), but as of this writing a strong causal role cannot be justified.

Thus, the data are not consistent enough to claim that the dearth of women in STEM careers has been shown to be primarily a result of direct consequence of biological sex differences (e.g., genes, hormones) impeding women's aptitude at math or spatial cognition, which, in turn, preclude their entry into STEM careers (Figure A2). There is too much inconsistency across studies, tests, cultures, and occasionally across epochs to justify a strong direct role of biological sex (though see Section 7 for more nuanced argument).

Putting aside the inconsistent and sometimes contradictory demonstrations of neural and hormonal influences on mental rotation, and the absence of studies focused on the right tail, we note that the putative role of math and spatial aptitude in the dearth of women in STEM is also problematic. Granted, there are substantial sex asymmetries at the extreme right tail on math and spatial aptitude, at least as indexed by timed (often multiple-choice) tests such as the SAT-M, and even though the magnitude of these gaps has shrunk dramatically over time in some analyses, they are still sizable by adolescence, despite the absence of consistent sex differences among pre-adolescent children (e.g., Lachance & Mazzocco, 2006). However, it is unclear what the inconsistency, along with the lack of reliable sex differences among young children, means. An assumption is that spatial ability underlies advanced mathematics (e.g., Fias & Fischer, 2005); that girls' presumed deficits in rotating three-dimensional figures constrain their ability to do advanced math and engineering. If true, however, better evidence is needed to causally link these two phenomena. Extensions of Sorby (2001, 2005) and others are needed in all math-intensive fields to determine whether females' difficulties with spatial cognition undermine their career progress.

There are no consistent sex differences among young children in other types of spatial cognition, such as embedded figures, spatial location, and spatial memory (Grimshaw et al., 1995; Lachance & Mazzocco, 2006). The tests demonstrating sex differences (e.g., SAT-M) are atheoretical, and adding or removing items may change their validity in unknown ways and alter sex differences. It is possible that scoring in the top 1% or top 0.1%, or even in the top 0.01% is helpful in becoming a successful scientist, but the evidence (Nuttall, Casey, & Pezaris, 2005) is indirect, although longitudinal data show that STEM scientists hail overwhelmingly from the right tail, often the extreme right tail.

Recall that Nuttall et al. (2005) reported that sex differences in various ability groups' SAT-M scores disappeared when mental rotation ability was covaried, but the mental-rotation ability remained an important predictor after SAT-M scores were controlled. This strongly suggests that spatial cognition plays an

important role in sex differences in math aptitude, but the question remains whether there is a threshold SAT-M score needed to be a successful STEM scientist. It would be helpful to have direct longitudinal evidence that SAT-M scores in the top 1% or 0.1% represent a threshold for success. Doubtless some STEM scholars will be shown to have lower SAT-M scores, but how aberrant are they? As a rough approximation of the amount of quantitative reasoning ability it might take to become a STEM leader (a tenured scientist at one of the top 50 research universities), Park et al. (2007) analyzed the 25+-year longitudinal data from SMPY's first three cohorts ($N = 2,409$); all participants were in the top 1% of quantitative reasoning ability by age 13 years. The mean SAT-M score of the participants who subsequently secured tenure-track positions at the top 50 U.S. universities was 100 points higher than the SAT-M scores of peers who became professors in the humanities, an underestimate because of ceiling scores of several of the adolescents who had maximum SAT-M scores of 800. Park et al. (2008) also showed that the odds ratio of publishing in STEM journals as a graduate student increased substantially with SAT-M scores, such that those in the top quarter of the top 1% at age 13 years exceeded those in the bottom quarter of the top 1% 25 years later. These authors have noted that it takes much more than exceptional quantitative reasoning ability to be a STEM leader, but exceptional quantitative reasoning ability is essential, and it continues to matter even as one goes far out on the right tail, hence our skepticism about Shalala et al.'s (2007) claim that low scores are common among STEM professionals (see Wise et al., 1979; Section 10 of the supplemental materials). Nevertheless, female students are doing very well in advanced math courses, including geometry and calculus, even if their SAT-M scores lag behind those of male students at the right tail. Moreover, the superiority of females in some other countries over U.S. and Canadian males is sometimes larger than the domestic gender gap, suggesting that a closer look at the right tail in these cultures is warranted. Relatedly, women from other countries occasionally represent higher proportions of the STEM workforce than do American women.

The conflict between work and family experienced by women in math-intensive fields is also evident in non-math-intensive fields, even ones in which women constitute the majority. If true, then the combination of sex differences in career aspirations that result in fewer women in the math pipelines (preferring to enter the pipelines for medicine, law, social science, biology, and humanities, whereas men are preferring math, engineering, physics, computer sciences, and chemistry), coupled with rigid schedules and family-unfriendly work conditions that result in some women opting out of all pipelines, can account for the dearth of women in STEM. Factors such as mental rotation, hormones, and discrimination may account for a small portion of the dearth, but the literature raises important reservations about the magnitude of each (e.g., inconsistencies, unrepresentative samples, no focus on the right tail, transnational reversals, female superiority in grades, occasionally shrinking gender ratios and large cohort effects, women succeeding in PhD programs but dropping out of tenure-track jobs, similar grant-funding rates, hiring and remuneration rates).

Past rates of potential female talent diverted from STEM fields could change if flexible options allowed more women to remain in these fields. Women's choices regarding family and career are highly constrained by the family division of labor. Universities and institutes could accommodate family responsibilities by giving

faculty part-time work that segues to full-time tenure track. Other family-friendly work policies could also be implemented to make it easier for mothers to remain in STEM careers and to juggle work and family demands. Flexibility within reasonable constraints should exist for pediatric visits and sick days, and maternity leave should allow women the possibility of short-term continuation of research with relief of teaching. These kinds of options can make the difference in making it past tenure. However, this is a different matter from encouraging women to switch from favoring family or non-STEM fields to math fields. A less rigid career structure would provide a choice that able women may or may not choose.

However, small initial differences in cognitive, social, or biological domains could potentially snowball into larger sex differences. Thus, it is conceivable that a small advantage in one domain (e.g., mental rotation ability or discrimination) could result in a cascade of experiences that eventuate in much larger differences later (see Section 13 of the supplemental materials for details).

In closing, a fundamental question plaguing this area of research is the causal relation between biology, spatial cognition, mathematics, and success in math-intensive STEM fields. The available evidence is inferential, but it lies at the heart of the matter. The construct validation process needed to establish male advantage in STEM fields as a function of superior spatial ability (possibly because of its role in advanced mathematics) is littered with loopholes. Nothing close to a tightly reasoned and supported argument currently exists. The closest and best, which was mentioned earlier, is represented in the work of Casey, Nuttall, Benbow, Lubinski, and their colleagues (see Park et al., 2008) and the animal work on hormones and spatial behaviors. Much more research will be needed to fortify a linkage between spatial ability and STEM career success. Are fewer women in these fields because they lack spatial skills that form the basis of higher mathematics, which, in turn, is critical to success in STEM fields, as some suggest? And is this due in part to hormones? When one measure of spatial cognition (e.g., two-dimensional mental rotation) does not predict mathematics, the inclination has been to focus on another measure (three-dimensional rotation) that does predict mathematics. When spatial reasoning fails to predict sex differences in early mathematics proficiency, the tendency has been to argue that perhaps such a prediction should not reveal itself until puberty when hormones surge or mathematics gets complex (e.g., Gouchie & Kimura, 1991). When nonheterosexual women outperform heterosexual women on spatial tests, despite not differing in circulating testosterone levels, the inference is made that they probably differed in their exposure to prenatal androgens, and that prenatal androgens are related to adult spatial abilities (van Anders & Hampson, 2005). Further, when two different measures of prenatal androgen exposure (digit ratio of 2nd to 4th finger and CAH girls) are differentially related to later spatial skill, the default is to suggest that the earlier timing of the former (first trimester) probably predates the development of the spatial systems influenced by androgen (2nd or 3rd trimester). Such post hoc arguments, albeit reasonable conjectures, are endlessly unfalsifiable. What is needed is a stronger theoretical basis for generating and testing highly specific hypotheses a priori. We hope that the first step in crafting such a theory is the critical examination, sifting, and connection of findings from the diverse disciplines and domains represented by the framework in Figure 5.

On the basis of this review of more than 400 published articles and chapters (including online materials), a confluence of factors predict the underrepresentation of women in math-intensive fields, including the observation that math-proficient women often prefer fields that are less math-intensive (e.g., biology, medicine, dentistry, veterinary medicine), and when they do enter math-intensive careers, they are more likely to drop out as they advance; more men score in the extreme math-proficient range on entrance tests such as the GRE-Q, thus gaining admission more frequently; women who are highly competent in math are more likely than men to also have high verbal competence, thus allowing the option of going into the humanities or law; and in some math-intensive fields, women with children are penalized as far as promotion. Of these factors, personal lifestyle choices, career preferences, and social pressures probably account for the largest portion of variance. This does not mean that math ability plays no role, because there is evidence from a number of studies indicating that among highly talented individuals, math ability is a significant predictor of scientific accomplishments and grades (Park et al., 2007). Recall that being in the top quartile of the top 1% in mathematics was a stronger predictor of publishing in STEM journals and getting patents by age 33 years than being in the bottom quartile of the top 1% (Park et al., 2008), and the bulk of scientists come from the top high school mathematics talent (Wise et al., 1979). But math ability does not appear to trump other factors in accounting for the underrepresentation of women in math-intensive fields: Benbow, Lubinski, and colleagues have demonstrated differences among men and women who are highly talented in mathematics in terms of who majors in STEM fields and who remain in these majors: Fewer math-talented women major in engineering and physics, and more major in medicine and biology, reflecting the importance of preferences (Webb et al., 2002). Relatedly, the math-verbal split is important in predicting later STEM accomplishments (Park et al., 2007). To complicate matters, the addition of spatial ability adds incremental validity over and above math ability to the prediction of which talented 13-year-olds become scientists, engineers, and mathematicians versus lawyers, social scientists, and humanists (Shea, Lubinski, & Benbow, 2001; see also Humphreys, Lubinski, & Yao, 1993). Therefore, preferences, ability, and relative strength of the math-to-verbal profile all contribute to the prediction of career trajectories.

The one research finding related to the underrepresentation of women in all academic careers, not just those that are math-intensive, that is robust, incontrovertible, and based on up-to-date information, is that women's fertility choices, and the timing of when to have children, are powerful predictors of career success, as are sex differences in lifestyle preferences (Hakim, 2006, 2007) and career choices. Several sources of evidence make clear the penalty associated with having children pretenure. Available evidence points to a special fertility penalty for mothers of young children in getting promoted in some math-intensive fields (e.g., Ginther & Kahn, 2006, reported that women were 9.6% less likely than comparable men to get promoted to full professor in the physical sciences). Encouragingly, the "childbearing penalty" is probably the most malleable of the factors we considered: If society deemed it desirable to increase the representation of women, various strategies could be implemented (see Shalala et al., 2007), such as deferred start-up of tenure-track positions and part-time work that segues to full-time tenure track work when

children are no longer in need of intensive care, designed flexibly to keep these women from opting out entirely, although one cannot assume that such changes will have the desired result if women disproportionately prefer home-centered and adaptive work lifestyles rather than single-minded commitment to career lifestyles (Hakim, 2006, 2007). Moreover, the viability of such a policy may vary from field to field, corresponding to the rate of knowledge obsolescence.

To summarize our conclusions regarding the underrepresentation of women in math-intensive fields, we note that a powerful explanatory factor is that mathematics-capable women disproportionately choose non-mathematics fields and that such preferences are apparent among math-competent girls during adolescence. Of women who enter STEM fields, approximately twice as many leave them as do men (Preston, 2004).

We found that evidence for a direct effect of innate hormonal differences on math and spatial ability (the basis for the intrinsic-ability-differences biological model in Figure A2) is contradictory and inconclusive, with scant data on right-tail samples. Despite the failure to link sex differences in mathematical and spatial ability to prenatal and postnatal hormones, the fact is that there are persistent sex differences in spatial reasoning and mathematical ability at the right tail, on the order of approximately 2 to 1 on various gatekeeper tests such as SAT-M and GRE-Q (e.g., Hyde et al., 2008), which may reflect purely sociocultural factors, purely biological factors, or some combination. On the basis of transnational data showing very inconsistent sex differences at the right tail, including countries where they are absent or even reversed (e.g., Guiso et al., 2008; Penner, in press) and U.S. data showing a narrowing of the sex gap at the right tail over time (Gates, 2006b), we conclude that the bases of mathematical and spatial differences are almost certainly not purely biological but rather must include a strong sociocultural component. The presence of fewer women at the right tail in mathematical and spatial ability renders fewer available for some math-intensive graduate programs on account of their GRE-Q scores. Having stated this, we note that data linking math and spatial abilities to STEM success are indirect, although strongly suggestive of playing a causal role in women's underrepresentation (Humphreys et al., 1993; Shea et al., 2001, for spatial support). However, if each sex's representation were solely a function of math ability, there would be roughly double the number of women in math-intensive careers as now exists, because assuming a 2:1 male-to-female ratio at the top 1% of math ability, women would be expected to comprise 33% of the professorships in math-intensive fields. In actuality, they comprise far less—as little as 10% of faculty in physics (Gates, 2006a; Ivie & Ray, 2005). Clearly, preferences must be a strongly causal factor in their opting to enter other careers.

We found that cultural and discriminatory causal pathways may be less important today than in the past, and unequal representation in STEM careers is not uniquely impeded by inequality in child-rearing responsibilities between the sexes because such inequality, although omnipresent, leads women with children to have less time for all careers, not just STEM ones. This effect is magnified by the coincidence of tenure decisions with childbearing in all fields. The tenure structure in academe demands that women who have children make their greatest intellectual achievements contemporaneously with their greatest physical and emotional achievements, a feat fathers are never expected to accomplish. When women opt

out of careers (or segue to part-time work in them) to have children, this is a choice men are not required to make. The reasons women opt out of math-intensive fields—either when making initial career selections or after they have begun a career—are complex. Reasons for preferring non-mathematics fields may include both free and coerced choices, which can be influenced by biological and sociocultural factors that either enable or limit women. Conclusions about women's underrepresentation are also thwarted because the cognitive underpinnings for success in mathematical fields are poorly understood, despite our knowledge of the importance of math and spatial skills (Park et al., 2007, 2008).

References

References marked with an asterisk indicate studies incorporated into the overall analysis but described in detail in the supplemental materials online.

- Achter, J. A., Lubinski, D., Benbow, C., & Eftekhari-Sanjani, H. (1999). Assessing vocational preferences among gifted adolescents adds incremental validity to abilities. *Journal of Educational Psychology, 91*, 777–786.
- Ackerman, P. L. (2006). Cognitive sex differences and mathematics and science achievement. *American Psychologist, 61*, 722–723.
- Adkins-Regan, E. (2005). *Hormones and animal social behavior*. Princeton, NJ: Princeton University Press.
- American Association of University Women. (1998). *Gender gaps: Where schools still fail our children*. Washington DC: American Association of University Women Educational Foundation.
- Ankney, C. D. (1992). Sex differences in relative brain size: The mismeasure of woman, too? *Intelligence, 16*, 329–336.
- Arden, R., & Plomin, R. (2006). Sex differences in variance of intelligence across childhood. *Personality and Individual Differences, 41*, 39–48.
- Astur, R. S., Ortiz, C. L., & Sutherland, R. J. (1998). A characterization of performance of men and women in a virtual Morris water task. *Behavioural Brain Research, 93*, 185–190.
- *Azurmendi, A., Braza, F., Sorozabal, S., García, A., Braza, P., Carreras, M. R., et al. (2005). Cognitive abilities, androgen levels, and body mass index in 5-year-old children. *Hormones and Behavior, 48*, 187–195.
- *Bachevalier, J., Brickson, M., Hagger, C., & Mishkin, M. (1990). Age and sex differences in the effect of temporal lobe lesions on the formation of visual discrimination in monkeys. *Behavioral Neuroscience, 104*, 885–889.
- *Bachevalier, J., Hagger, C., & Bercu, B. (1989). Gender differences in visual habit formation in 3-month-old rhesus monkeys. *Developmental Psychology, 22*, 585–599.
- Baenninger, M. A., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20*, 327–344.
- Baenninger, M. A., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences, 7*, 363–379.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*, 612–637.
- Baron-Cohen, S. (2007). Sex differences in mind: Keeping science distinct from social policy (pp. 159–172). In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence*. Washington, DC: American Psychological Association.
- *Baron-Cohen, S., Knickmeyer, R. C., & Belmonte, (2005). Sex differences in the brain: Implications for explaining autism. *Science, 310*, 819–823.
- Baron-Cohen, S., Lutchmaya, S., & Knickmeyer, R. C. (2004). *Prenatal testosterone in mind: Amniotic fluid studies*. MIT Press.
- Battista, M. T. (1990). Spatial visualization and gender differences in high

- school geometry. *Journal of Research in Mathematics Education*, 21, 47–60.
- Beaman, R., Wheldall, K., & Kemp, C. (2006). Differential teacher attention to boys and girls in the classroom. *Educational Review*, 58, 339–366.
- Becker, B. J., & Hedges, L. V. (1984). Meta-analysis of cognitive gender differences: A comment on an analysis by Rosenthal and Rubin. *Journal of Educational Psychology*, 76, 583–587.
- Becker, J. R. (1981). Differential treatment of females and males in mathematics classes. *Journal for Research in Mathematics Education*, 12, 40–53.
- *Beilock, S., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology*, 136, 256–276.
- Belkin, L. (2008, May 15). Diversity isn't rocket science, is it? *The New York Times*, p. A-2.
- Beller, M. & Gafni, N. (1996). The 1991 international assessment of educational progress in mathematics and sciences: The gender differences perspective. *Journal of Educational Psychology*, 88, 365–377.
- Ben-Zeev, T., Duncan, S., & Forbes, C. (2005). Stereotypes and mathematics performance. In J. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 235–249). New York: Psychology Press.
- Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral & Brain Sciences*, 11, 169–182.
- Benbow, C. P. (1992). Academic achievement in math and science of students between 13 and 23: Are there differences among students in the top 1% of math ability? *Journal of Educational Psychology*, 84, 51–61.
- Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability: Their status 20 years later. *Psychological Science*, 11, 474–480.
- Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, 210, 1262–1264.
- Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability. *Science*, 222, 1029–1030.
- *Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science*, 3, 203–206.
- Berenbaum, S., & Resnick, S. (2007). The seeds of career choices: Prenatal sex hormone effects on psychological sex differences. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 147–158). Washington, DC: American Psychological Association.
- Bhasin, S., Woodhouse, L., Casaburi, R., Singh, A. B., Bhasin, D., Berman, N., et al. (2001). Testosterone dose–response relationships in healthy young men. *American Journal of Physiology—Endocrinology and Metabolism*, 281, 1172–1181.
- Bhasin, S., Woodhouse, L., Casaburi, R., Singh, A. B., Mac. R. P., Lee, M., et al. (2005). Older men are as responsive as young men to the anabolic effect of graded doses of testosterone. *Journal of Clinical Endocrinology and Metabolism*, 90, 678–688.
- Biek, D. M. (2006). *Stereotype threat and domain identification*. Unpublished doctoral dissertation, Cornell University.
- Bornstein, M. H., Haynes, O., Pascual, L., Painter, K., & Galperin, C. (1999). Play in two societies: Pervasiveness of process, specificity of structure. *Child Development*, 70, 317–331.
- Brandon, P. R., Newton, B. J., & Hammond, O. W. (1987). Children's mathematics achievement in Hawaii: Sex differences favoring girls. *American Educational Research Journal*, 24, 437–461.
- Bridgeman, B., & Lewis, C. (1996). Gender differences in college mathematics grades and SAT–M scores. *Journal of Educational Measurement*, 33, 257–270.
- Bronfenbrenner, U., & Ceci, S. J. (1994). Nature–nurture in developmental perspective: A bioecological theory. *Psychological Review*, 101, 568–586.
- Brosnan, M. (1998). Spatial ability in children's play with Lego blocks. *Perceptual and Motor Skills*, 87, 19–28.
- Brosnan, M. (2006). Digit ratio and academia: Implications for the relationship between prenatal testosterone and academia. *British Journal of Psychology*, 97, 455–466.
- Brouwers, S. A., Van de Vijver, F. J. R., & Van Hemert, D. A. (in press). Variation in Raven's Progressive Matrices scores across time and place. *Learning and Individual Differences*. (Sex difference analysis available from S. A. Brouwers: s.a.brouwers@uvt.nl)
- Budden, A., Lortie, C., Tregenza, T., Aarssen, L., Koricheva, J., & Leimu, R. (2008). Response to Webb et al: Double-blind review: Accept with minor revisions, *Trends in Ecology and Evolution*, 23, 353–354.
- Budden, A., Tregenza, T., Aarssen, L., Koricheva, J., Leimu, R., & Lortie, C. (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology and Evolution*, 23, 4–6.
- Burnett, S. A., Lane, D. L., & Dratt, L. M. (1979). Spatial visualization and sex differences in quantitative ability. *Intelligence*, 3, 345–354.
- *Burt, S. A. (2008). Genes and popularity: Evidence of an evocative gene–environment correlation. *Psychological Science*, 19, 112–113.
- Byrnes, J. P., & Takahira, S. (1993). Explaining gender differences on SAT–Mathematics Items. *Developmental Psychology*, 29, 805–810.
- *Caldera, Y. M., Culp, A. M., O'Brien, M., Truglio, R. T., Alvarez, M., & Huston, A. C. (1999). Children's play preferences, construction play with blocks, and visual–spatial skills: Are they related? *International Journal of Behavioral Development*, 23, 855–872.
- Cambridge University. (2005, August 26). Student numbers. *Cambridge University Reporter*, 135(Special Issue No. 19).
- Caplan, P. J., McPherson, G. M., & Tobin, P. (1985). Do sex-related differences in spatial ability exist? *American Psychologist*, 40, 786–799.
- Carter, C. S., LaRussa, M. A., & Bodner, G. M. (1987). A study of two measures of spatial ability as predictors of success in different levels of general chemistry. *Journal of Research in Science Teaching*, 24, 645–657.
- Casey, M. B., Andrews, N., Schindler, H., Kersh, J. E., & Samper, A. (2008). The development of spatial skills through interventions involving block building activities. *Cognition and Instruction*, 26, 269–309.
- Casey, M. B., & Brabeck, M. M. (1989). Exceptions to the male advantage on a spatial task: Family handedness and college major as a factor identifying women who excel. *Neuropsychologia*, 27, 689–696.
- Casey, M. B., & Brabeck, M. M. (1990). Women who excel on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition*, 12, 73–84.
- Casey, M. B., Nuttall, R., Benbow, C. P., & Pezaris, E. (1995). Influence of spatial ability on gender differences in mathematics college entrance scores across diverse samples. *Developmental Psychology*, 31, 697–705.
- Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology*, 33, 669–680.
- Casey, M. B., Nuttall, R. L., & Pezaris, E. (1999). Evidence in support of a model that predicts how biological and environmental factors interact to influence spatial skills. *Developmental Psychology*, 35, 1237–1247.
- Casey, M. B., Nuttall, R. L., & Pezaris, E. (2001). Spatial–mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education*, 32, 28–57.
- Catsambis, S. (1994). The path to mathematics: Gender and racial–ethnic differences in mathematics participation from middle school to high school. *Sociology of Education*, 67, 199–215.
- Catsambis, S. (2005). The gender gap in mathematics: Merely a step function? In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differ-*

- ences in mathematics: An integrative psychological approach* (pp. 222–245). Cambridge, United Kingdom: Cambridge University Press.
- Cavallaro, M., Hansen, S., & Wenner, M. (2007, Fall). Bridging the gender gap. *Barnard Magazine*.
- Ceci, S. J. (1996). *On intelligence: A bioecological treatise on intellectual development*. Cambridge, MA: Harvard University Press.
- *Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap-closing: When the “have-nots” gain, but the “haves” gain even more. *American Psychologist*, *60*, 149–160.
- Ceci, S. J., & Williams, W. M. (2007). *Why aren't more women in science? Top researchers debate the evidence*. Washington, DC: American Psychological Association.
- Charles, M., & Bradley, K. (2006). A matter of degrees: Female underrepresentation in computer science programs cross-nationally. In J. McGrath Cohoon & B. Aspray (Eds.), *Women and information technology: Research on the reasons for underrepresentation* (pp. 183–203). Cambridge, MA: MIT Press.
- *Cherrier, M., Asthana, S., Plymate, S., Baker, L., Matsumoto, A. M., Peskind, E., et al. (2001). Testosterone supplementation improves spatial and verbal memory in healthy older men. *Neurology*, *57*, 80–88.
- *Christiansen, K. (1993). Sex hormone-related variations of cognitive performance in pKung San hunter-gatherers of Namibia. *Neuropsychobiology*, *27*, 97–107.
- *Christiansen, K., & Knussman, R. (1987). Sex hormones and cognitive functioning in men. *Neuropsychobiology*, *18*, 27–36.
- *Collaer, M. L., Geffner, M. E., Kaufman, F. R., Buckingham, B., & Hines, M. (2002). Cognitive and behavioral characteristics of Turner syndrome: Exploring a role for ovarian hormones in female sexual differentiation. *Hormones and Behavior*, *41*, 139–155.
- College Entrance Examination Board. (1939). *CEEB Special Aptitude Test in Spatial Relations*. New York: Author.
- Connellan, J., Baron-Cohen, S., Wheelwright, S., Ba'tki, A., & Ahluwalia, J. (2001). Sex differences in human neonatal social perception. *Infant Behavior and Development*, *23*, 113–118.
- *Connor, J. M., Serbin, L. A., & Schackman, M. (1977). Sex differences in children's response to training on a visual-spatial test. *Developmental Psychology*, *13*, 293–294.
- Correll, S. J. (2004). Constraints into preferences: Gender, status, and emerging career aspirations. *American Sociological Review*, *69*, 93–114.
- Cronin, V. (1967). Mirror-image reversal discrimination in kindergarten and first-grade children. *Journal of Experimental Child Psychology*, *5*, 577–585.
- *Cullen, M., Hardison, C., & Sackett, P. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology*, *89*, 220–230.
- Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology*, *38*, 1639–1655.
- *Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's mathematics performance. *Science*, *314*, 435.
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardtstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, *28*, 1615–1628.
- Davis, G. (in press). Improving the postdoctoral experience: An empirical approach. In R. Freeman & D. Goroff (Eds.), *The science and engineering workforce in the United States*. Chicago: NBER/University of Chicago Press.
- *Davis-Kean, P., Jacobs, J., Bleeker, M., Eccles, J. S., & Melanchuk, O. (2007, June 25). How dads influence their daughters' interest in mathematics [University of Michigan]. *ScienceDaily*. Retrieved February 9, 2008, from <http://www.sciencedaily.com/releases/2007/06/070624143002.htm>
- Davison, K., & Susman, E. (2001). Are hormone levels and cognitive ability related during early adolescence? *International Journal of Behavioral Development*, *25*, 416–428.
- De Vries, G., & Simerley, R. B. (2002). Anatomy, development, and function of sexually dimorphic neural circuits in the mammalian brain. In D. W. Pfaff, A. Arnold, A. Etgen, E. Fahrbach, R. Moss, & R. Rubin (Eds.), *Hormones, brain and behaviour: Vol. IV. Development of hormone-dependent neuronal systems* (pp. 137–191). San Diego, CA: Academic Press.
- Dearman, N. B., & Plisko, V. W. (1979). *The condition of education*. Washington, DC: National Center for Educational Statistics.
- Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish mental survey 1932. *Intelligence*, *31*, 533–542.
- Deary, I. J., Whalley, L., Lemmon, H., Crawford, J. R., & Starr, J. M. (1999). The stability of individual differences in mental ability from childhood to old age. *Intelligence*, *28*, 49–55.
- Delgado, A., & Prieto, G. (2004). Cognitive mediators and sex differences in math. *Intelligence*, *32*, 25–32.
- Department for Education and Skills. (2002). Autumn Package 2002 [data file] Retrieved November 5, 2008, from (http://www.standards.dfes.gov.uk/performance/2002KS4_whole_revised.pdf?version=1)
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Changes in grey matter induced by training. *Nature*, *427*, 311–312.
- Dweck, C. S. (2007). Is mathematics a gift? Beliefs that put females at risk. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 47–56). Washington, DC: American Psychological Association.
- *Dweck, C. S., & Repucci, N. D. (1973). Learned helplessness and reinforcement responsibility in children. *Journal of Personality and Social Psychology*, *25*, 109–116.
- Eals, M., & Silverman, I. (1994). The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology & Sociobiology*, *15*, 95–105.
- Eccles, J. S. (2007a). Families, schools, and developing achievement-related motivations and engagement. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization* (pp. 665–691). New York: Guilford Press.
- Eccles, J. S. (2007b). Where are all of the women? In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 199–210). Washington, DC: American Psychological Association.
- Eccles, J. S., Barber, B., & Jozefowicz, D. (1999). Linking gender to educational, occupational, and recreational choices: Applying the Eccles et al. model of achievement-related choices. In W. B. Swann, J. Langlois, & L. Gilbert (Eds.), *Sexism and stereotypes in modern society* (pp. 153–192). Washington DC: American Psychological Association.
- Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.
- Entwisle, D., R., Alexander, K., & Olson, L. S. (1994). The gender gap in mathematics: Its possible origins in neighborhood effects. *American Sociological Review*, *59*, 822–838.
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, *43*, 95–103.
- Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, *62*, 61–84.
- Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex Roles*, *30*, 81–92.
- Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*, 850–855.
- Fennema, E., Carpenter, T., Jacobs, V., Franke, M., & Levi, L. (1998).

- Longitudinal Study of Gender Differences in Young Children's Mathematical Thinking. *Educational Researcher*, 27, 6–11.
- Ferguson, R. C. (2006). *Teacher's guide: Research and benefits of chess*. New Windsor, NY: United States Chess Federation. Retrieved January 3, 2007, from http://www.quadcitychess.com/benefits_of_chess.html
- Fias, W., & Fischer, M. H. (2005). Spatial representation of numbers. In J. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 43–54). London: Routledge.
- Finegan, J. K., Niccols, G. A., & Sitarenios, G. (1992). Relations between prenatal testosterone levels and cognitive abilities at 4 years. *Developmental Psychology*, 28, 1075–1089.
- Fink, B., Brookes, H., Neave, N., Manning, J. T., & Geary, D. C. (2006). Second to fourth digit ratio and numerical competence in children. *Brain and Cognition*, 61, 211–218.
- Fischer, W. (2006). *The educational value of chess*. New Horizons for Learning. Retrieved December 9, 2006, from <http://www.newhorizons.org/strategies/thinking/fischer.htm>
- Flynn, J. R., & Rossi-Case, L. (2008). *Raven's Progressive Matrices: Female genes good—Massive IQ gains in Argentina*. Unpublished manuscript. Available from jim.flynn@stonebow.otago.ac.nz
- Fox, L. H., & Cohn, S. J. (1980). Sex differences in the development of precocious mathematical talent. In L. H. Fox, L. Brody, & D. Tobin (Eds.), *Women and the mathematical mystique* (pp. 94–112). Baltimore, MD: Johns Hopkins Press.
- Freeman, C. E. (2004). *Trends in educational equity of girls and women: 2004*. Washington, DC: U.S. Department of Education, National Center for Educational Statistics.
- Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in quantitative tasks. *Review of Educational Research*, 59, 185–213.
- Friedman, L. (1995). The space factor in mathematics: Gender differences. *Review of Educational Research*, 65, 22–50.
- Frome, P., & Eccles, J. S. (1998). Parental effects on adolescents' academic self-perceptions and interests. *Journal of Personality and Social Psychology*, 74, 435–452.
- Gallagher, A. M. (1992). *Sex differences in problem-solving strategies used by high-scoring examinees on the SAT-M* (ETS Research Report No. 92–33/College Board Report No. 92–2). New York: College Board.
- Gallagher, A. (1998). Gender and antecedents of performance in mathematics testing. *Teachers College Record*, 100, 297–314.
- Gallagher, A., & DeLisi, R. (1994). Gender differences in Scholastic Aptitude Test—Mathematics problem solving among high-ability students. *Journal of Educational Psychology*, 86, 204–211.
- Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, 75, 165–190.
- Gallagher, A. M., & Kaufman, J. C. (2005). *Gender differences in mathematics: An integrative psychological approach*. Cambridge, United Kingdom: Cambridge University Press.
- Gallagher, A., Levin, J., & Cahalan, C. (2002). *Cognitive patterns of gender differences on mathematics admissions tests* (ETS Research Report 02–19 / GRE Board Professional Report No. 96–17P). Princeton, NJ: Educational Testing Service.
- Gates, E. (2006a, April). A scientific point of view. *Physics Today*, pp. 64–65.
- Gates, E. (2006b, December). Women in physics: Why and why not. *Physics Today*, pp. 14–15.
- Gazzaniga, M., Ivry, R., & Magnum, G. (1998). *Cognitive neuroscience: The biology of the mind*. New York: Norton.
- Geary, D. C. (1996). Sexual selection and sex differences in mathematics abilities. *Behavioral and Brain Sciences*, 19, 229–284.
- Geary, D. C. (1998). *Male, female: The evolution of human sex differences*. Washington, DC: American Psychological Association.
- Geary, D. C. (2002). Sexual selection and sex differences in social cognition. In A. V. McGillicuddy-DeLisi & R. DeLisi (Eds.), *Biology, society, and behavior: The development of sex differences in cognition* (pp. 23–54). Greenwich, CT: Ablex.
- Geary, D. C., & DeSoto, M. C. (2001). Sex differences in spatial abilities among adults in the United States and China. *Evolution & Cognition*, 7, 172–177.
- Gender Equity Committee of the University of Pennsylvania. (2001, December 4). "The gender equity report." *Almanac Supplement*, 48(14).
- Giedd, J. N., Vaituzis, A. C., Hamburger, S. D., Lange, N., Rajapakse, J. C., Kaysen, D., et al. (1996). Quantitative MRI of the temporal lobe, amygdala, and hippocampus in normal human development: Ages 4–18 years. *Journal of Comparative Neurology*, 366, 222–232.
- Gierl, M., Khaliq, S., & Boughton, K. (1999, June). Differential item functioning on Alberta Education mathematics achievement tests: Prevalence and policy implications. In X. Ma (chair), *Improving large-scale assessment in education research*. Symposium conducted at the annual meeting of the Canadian Society for the Study of Education Sherbrooke, Québec, Canada.
- Ginther, D. K. (2001, February). *Does science discriminate against women? Evidence from academia, 1973–97* (Working Paper No. 2001–02). Atlanta, GA: Federal Reserve Bank of Atlanta, Research Department.
- Ginther, D. K. (2004). Why women earn less than men: Economic explanations for the gender salary gap in science. *AWIS: Association for Women in Science*, 33, 1–5.
- Ginther, D. K., & Kahn, S. (2006). *Does science promote women? Evidence from academia 1973–2001*. Unpublished manuscript. Retrieved January 2, 2006, from http://www.nber.org/~sewp/Ginther_Kahn_revised8-06.pdf
- *Goldberg, C. (1999, March 23). MIT acknowledges bias against female professors. *The New York Times*, p. 1.
- Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, 90, 715–741.
- Goldstein, J. H. (1994). Sex differences in toy play and use of video games. In J. H. Goldstein (Ed.), *Toys, play, and child development* (pp. 110–129). New York: Cambridge University Press.
- Good, C., Aronson, J., & Harder, J. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level mathematics courses. *Journal of Applied Developmental Psychology*, 29, 17–28.
- *Good, C., Dweck, C. S., & Rattan, A. (2005a). *An incremental theory decreases vulnerability to stereotypes about mathematics ability in college females*. Unpublished raw data, Columbia University.
- *Good, C., Dweck, C. S., & Rattan, A. (2005b). *Portraying genius: How fixed vs. malleable portrayal of mathematics ability affects females' motivation and performance*. Unpublished raw data, Columbia University.
- *Gordon, H. W., Corbin, E. D., & Lee, P. A. (1986). Changes in specialized cognitive function following changes in hormone levels. *Cortex*, 22, 399–415.
- Gouchie, C., & Kimura, D. (1991). The relationship between testosterone levels and cognitive ability patterns. *Psychoneuroendocrinology*, 16, 323–334.
- Grabner, R. H., Neubauer, A. C., & Stern, E. (2006). Superior performance and neural efficiency: The impact of intelligence and expertise. *Brain Research Bulletin*, 69, 422–439.
- Grabner, R. H., Stern, E., & Neubauer, A. C. (2003). When intelligence loses its impact: Neural efficiency during reasoning in a familiar area. *International Journal of Psychophysiology*, 49, 89–98.
- Grimshaw, G., Sitarenios, G., & Finegan, J. (1995). Mental rotation at 7 years: Relations with prenatal testosterone levels and spatial play experiences. *Brain and Cognition*, 29, 85–100.
- Grossi, D., Orsini, A., Monetti, C., & De Michele, G. (1979). Sex differ-

- ences in children's spatial and verbal memory span. *Cortex*, 15, 667–670.
- Guay, R. (1977). *Purdue Spatial Visualization Test: Rotations*. West Lafayette, IN: Purdue Research Foundation.
- Guiso, L., Ferdinando, M., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, 320, 1164–1165.
- Guo, G., & Stearns, E. (2002). The social influences on the realization of genetic potential for intellectual development. *Social Forces*, 881–910.
- Gur, R. C., Alsop, D., Glahn, D., Petty, R., Swanson, C. L., Maldjian, J. A., et al. (2000). An fMRI study of sex differences in regional activation to a verbal and a spatial task. *Brain and Language*, 74, 157–170.
- Gur, R. C., & Gur, R. E. (2007). Neural substrates for sex differences in cognition. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 189–198). Washington, DC: American Psychological Association.
- Haier, R. (2007). Brains, bias, and biology: Follow the data. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 113–120). Washington, DC: American Psychological Association.
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2004). Structural brain variation and general intelligence. *NeuroImage*, 23, 425–433.
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2005). The neuroanatomy of general intelligence: Sex matters. *NeuroImage*, 25, 320–327.
- Hakim, C. (2000). *Work–lifestyle choices in the 21st century*. Oxford, England: Oxford University Press.
- Hakim, C. (2003). *Models of the family in modern societies: Ideals and realities*. Aldershot: Ashgate.
- Hakim, C. (2004). *Key issues in women's work: Female diversity and the polarization of women's employment* (2nd ed.). London: GlassHouse Press.
- Hakim, C. (2005). Sex differences in work–life balance goals. In D. Houston (Ed.), *Work–life balance in the twenty-first century* (pp. 55–79). London: Palgrave Macmillan.
- Hakim, C. (2006). Women, careers, and work–life preferences. *British Journal of Guidance and Counseling*, 34, 279–294.
- Hakim, C. (2007). Dancing with the devil? Essentialism and other feminist heresies. *British Journal of Sociology*, 58, 123–132.
- Halari, R., Sharma, T., Hines, M., Andrew, C., Simmons, A., & Kumari, V. (2006). Comparable fMRI activity with differential behavioural performance on mental rotation and overt verbal fluency tasks in healthy men and women. *Experimental Brain Research*, 169, 1–14.
- Halpern, D. F. (2007). Science, sex, and good sense: Why women are underrepresented in some areas of science and mathematics (pp. 121–130). In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence*. Washington, DC: American Psychological Association.
- Halpern, D. F., Benbow, C. P., Geary, D., Gur, R., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8, 1–51.
- Hamel, M., Ingelfinger, J., Phimister, E., & Solomon, C. (2006). Women in academic medicine—Progress and challenges. *The New England Journal of Medicine*, 355, 310.
- *Hampson, E. (1990). Estrogen-related variations in human spatial and articulatory-motor skills. *Psychoneuroendocrinology*, 15, 97–111.
- *Hampson, E., & Kimura, D. (1988). Reciprocal effects of hormonal fluctuations on human motor and perceptual–spatial skills. *Behavioral Neuroscience*, 102, 456–459.
- Hampson, E., & Moffat, S. D. (2005). The psychobiology of gender: The cognitive effects of reproductive hormones in the adult nervous system. In R. J. Sternberg, A. Eagley, & A. Beal (Eds.), *The psychology of gender* (2nd ed., pp. 38–64). New York: Guilford Press.
- *Hampson, E., Rovet, J. F., & Altmann, D. (1998). Spatial reasoning in children with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Developmental Neuropsychology*, 14, 299–320.
- Harris, A. M., & Carlton, S. (1993). Patterns of gender differences on mathematics items on the Scholastic Aptitude Test. *Applied Measurement in Education*, 6, 137–151.
- Harvard University. (2005). *Women at Harvard University*. Cambridge, MA: Author. Retrieved January 4, 2007, from <http://www.hno.harvard.edu/guide/underst/under4.html>
- Haslanger, S. (2007, September 10). Philosophy and sexism. *Inside Higher Education*. Retrieved October 11, 2007, from www.insidehighered.com/news/2007/09/10/philos
- *Hausman, P., & Steiger, J. H. (2001, February). Confession without guilt? MIT jumped the gun to avoid a sex-discrimination controversy, but shot itself in the foot. *Independent Women's Forum*, 1–12.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41–45.
- Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology*, 91, 684–689.
- Heilman, M. E., & Haynes, M. C. (2005). No credit where credit is due: Attributional rationalization of women's success in male–female teams. *Journal of Applied Psychology*, 90, 905–916.
- Hill, S. T., & Johnson, J. M. (2004). *Science and engineering degrees: 1966–2001*. Arlington, VA: National Science Foundation. Retrieved July, 1, 2007, from www.nsf.gov/statistics/NSF04311/
- Hines, M. (2007). Do sex differences in cognition cause the shortage of women in science? In S. Ceci, & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 101–112). Washington, DC: American Psychological Association.
- Hines, M., Fane, B. A., Pasterski, V. L., Mathews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality. *Psychoneuroendocrinology*, 28, 1010–1026.
- Hines, M., & Kaufman, F. R. (1994). Androgen and the development of human sex-typical behavior—Rough-and-tumble play and sex of preferred playmates in children with CAH. *Child Development*, 65, 1042–1053.
- Hoffer, T. B., & Grigorian, K. (2005, December). *All in a week's work: Average work weeks of doctoral scientists and engineers* (Science Resources Statistics Publication. No. 06–302). Arlington, VA: National Science Foundation.
- Hogervorst, E., Bandelow, S., & Moffat, S. D. (2005). Increasing testosterone levels and effects on cognitive functions in elderly men and women. *Drug Targets—CNS & Neurological Disorders*, 4, 531–540.
- Hong, E., O'Neil, H. F., & Feldon, D. (2005). Gender effects on mathematics achievement. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics* (pp. 163–189). Cambridge, United Kingdom: Cambridge University Press.
- Hrdy, S. B. (1999). *Mother nature: A history of mothers, infants and natural selection*. New York: Pantheon.
- Hugdahl, K., Thomsen, T., & Erslund, L. (2006). Sex differences in visuo-spatial processing: An fMRI study of mental rotation. *Neuropsychologia*, 44, 1575–1583.
- Humphreys, L., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, scientist, or artist. *Journal of Applied Psychology*, 78, 250–261.
- Hyde, J. S. (1981). How large are cognitive gender differences? *American Psychologist*, 36, 892–901.
- Hyde, J. S. (1990). Meta-analysis and the psychology of gender differences. *Signs: Journal of Women in Culture and Society*, 16, 55–73.
- Hyde, J. S. (2005). The gender similarity hypothesis. *American Psychologist*, 60, 581–592.
- Hyde, J. S. (2007). Women in science: Gender similarities in abilities and sociocultural forces. In S. J. Ceci, & W. M. Williams (Eds.), *Why aren't*

- more women in science? *Top researchers debate the evidence* (pp. 131–145). Washington, DC: American Psychological Association.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.
- Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics studies and effects: A meta-analysis. *Psychology of Women Quarterly, 14*, 299–324.
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321*, 494–495.
- Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science, 314*, 599–600.
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365–371.
- Irwing, P., & Lynn, R. (2005). Sex differences in means and variability on the Progressive Matrices in university students: A meta-analysis. *British Journal of Psychology, 96*, 505–524.
- Isgor, C., & Sengelaub, D. R. (2003). Effects of neonatal gonadal steroids on adult CA3 pyramidal neuron dendritic morphology and spatial memory. *Journal of Neurobiology, 55*, 179–190.
- Ivics, R., & Ray, K. (2005). *Women in physics and astronomy*. College Park, MD: American Institute of Physics.
- Jacobs, J. A., & Winslow, S. E. (2004). Overworked faculty: Job and stresses and family demands. *Annals of American Political and Social Scientist, 596*, 104–129.
- Jacobs, J. E., Davis-Kean, P., Bleeker, M., Eccles, J. S., & Malanchuk, O. (2005). “I can, but I don’t want to”: The impact of parents, interests, and activities on gender differences in mathematics. In A. Gallagher & J. Kaufman (Eds.), *Gender differences in mathematics* (pp. 246–263). Cambridge, United Kingdom: Cambridge University.
- Jacobs, J. E., & Eccles, J. S. (1985). Gender differences in mathematics ability: The impact of media reports on parents. *Educational Researcher, 14*, 20–25.
- *Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers’ gender-role stereotypic beliefs on mothers’ and children’s ability perceptions. *Journal of Personality and Social Psychology, 63*, 932–944.
- Jahoda, G. (1979). On the nature of difficulties in spatial-perceptual tasks: Ethnic and gender differences. *British Journal of Psychology, 70*, 351–363.
- Jardine, J., & Martin, N. G. (1983). Spatial ability and throwing accuracy. *Behavior Genetics, 13*, 331–340.
- Jayasinghe, U. W., Marsh, H. W., & Bond, N. (2003). A multilevel cross-classified modeling approach to peer-review of grant proposals. *Journal of the Royal Statistical Society, 166A*, 279–300.
- Jayasinghe, U. W., Marsh, H. W., & Bond, N. A. (2006). New reader trial approach to peer review in funding research grants: An Australian experiment. *Scientometrics, 69*, 591–606.
- Johnson, E. S., & Meade, A. C. (1987). Developmental patterns of spatial ability: An early sex difference. *Child Development, 58*, 725–740.
- *Jordan, T., Wurstenberg, K., Heinze, H., Peters, M., & Jancke, L. (2002). Women and men exhibit different cortical activation patterns during mental rotation tasks. *Neuropsychologia, 40*, 2397–2408.
- Keach, R. (1994). *Benefits of chess for youth*. In Y. Seirawan (Ed.), *Winning chess strategies*. Microsoft Press. Retrieved January 3, 2007, from <http://home1.gte.net/kaech/benefits.html>
- Kelly, A. (1988). Gender differences in pupil-teacher interactions: A meta-analytic review. *Research in Education, 39*, 1–24.
- Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in mathematics performance: The role of children’s approach to schoolwork. *Developmental Psychology, 42*, 11–26.
- Kersh, J., Casey, B. M., & Mercer Young, J. M. (2008). Research on spatial skills and block building in girls and boys: The relationship to later mathematics learning. In B. Spodak & O. Saracho (Eds.), *Mathematics, science and technology in early childhood education: Contemporary perspectives on mathematics in early childhood education* (pp. 233–253). Charlotte, NC: Information Age.
- Kimball, M. M. (1989). A new perspective on women’s mathematics achievement. *Psychological Bulletin, 105*, 198–214.
- Kimura, D. (1996). Sex, sexual orientation and sex hormones influence human cognitive function. *Current opinion in neurobiology, 6*, 259–263.
- Kimura, D. (2000). *Sex and cognition*. Cambridge, MA: MIT Press.
- Kimura, D. (2002). Sex hormones influence human cognitive pattern. *Neuroendocrinology Letters, 23*(Suppl. 4), 67–77.
- Kimura, D. (2007). “Under-representation” or misrepresentation? In S. Ceci & W. Williams (Eds.), *Why aren’t more women in science? Top researchers debate the evidence* (pp. 39–46). Washington, DC: American Psychological Association.
- Knickmeyer, R. C., & Baron-Cohen, S. (2006). Fetal testosterone and sex differences in typical social development and in autism. *Journal of Child Neurology, 21*, 825–845.
- Kolen, M. J., & Brennan, R. L. (1995). *Test equating*. New York: Springer Verlag.
- Krendl, A. C., Richeson, J. A., Kelley, W. M., & Heatherton, T. F. (2008). The negative consequences of threat: An fMRI investigation of the neural mechanisms underlying women’s underperformance in math. *Psychological Science, 19*, 168–175.
- *Kucian, K., Loenneker, T., Dietrich, T., Martin, E., & von Aster, M. (2005). Gender differences in brain activation patterns during mental rotation and number related cognitive tasks. *Psychological Science, 47*, 112–131.
- Kulp, M. T. (1999). Relationship between visual motor integration skill and academic performance in kindergarten through third grade. *Optometry and Vision Science, 76*, 159–163.
- Kuncel, N., & Hezlett, S. (2007). Standardized tests predict graduate student success. *Science, 315*, 1080–1081.
- Kurdek, L. A., & Sinclair, R. (2001). Predicting reading and mathematics achievement in 4th-grade children from kindergarten readiness scores. *Journal of Educational Psychology, 93*, 451–455.
- Lachance, J. A., & Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences, 16*, 195–216.
- Lacreuse, A., Kim, C. B., Moss, M. B., Rosene, D. L., Killiany, R. J., Moore, T. L., et al. (2005). *Behavioral Neuroscience, 119*, 118–126.
- Lamb (1996). *Completing school in Australia: Trends in the 1990s* (Research Report No. 1). Melbourne, Australia: Australian Council for Educational Research.
- Leahey, E., & Guo, G. (2001). Gender differences in math trajectories. *Social Forces, 80*, 713–732.
- Leboy, P. (2007, September 21). Why women leave academic medicine. *Inside Higher Education*. Retrieved June 2, 2008, from www.insidehighered.com/news/2007/09/21/women
- *Leinhardt, G., Seewald, A., & Engel, M. (1979). Learning what’s taught: Sex differences in instruction. *Journal of Educational Psychology, 71*, 432–439.
- Leslie, D. W. (2007, March 1). The reshaping of America’s academic workforce. *Research Dialogue, 87*. New York: TIAA-CREF. Retrieved May 20, 2008, from http://www.tiaa-crefinstitute.org/research/dialogue/The_Reshaping_of_America's_Academic_Workforce.html
- Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early spatial differences. *Developmental Psychology, 35*, 940–949.
- Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science, 16*, 841–845.
- Lewis, D. (2005, June 24). Mathematics: Probing performance gaps. *Science, 308*, 1871–1872.

- Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher*, 18, 17–27.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56, 1479–1498.
- Lippa, R. A. (1998). Gender-related individual differences and the structure of vocational interests on the people–things dimension. *Journal of Personality and Social Psychology*, 74, 996–1009.
- Lippa, R. A. (2005). Subdomains of gender-related occupational interests: Do they form a cohesive bipolar M-F dimension? *Journal of Personality*, 73, 693–730.
- Lippa, R. A. (2006). The gender reality hypothesis. *American Psychologist*, 61, 639–640.
- LoBello, S. G., & Gulgoz, S. (1991). Factor analysis of the Wechsler Preschool and Primary Scale of Intelligence—Revised. *Psychological Assessment*, 3, 130–132.
- Lohman, D., & Lakin, J. (in press). Consistencies in sex differences on the cognitive abilities test across countries, grades, test forms, and cohorts. *British Journal of Educational Psychology*.
- Long, J. (1992). Measures of sex differences in scientific productivity. *Social Forces*, 71, 159–178.
- Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "general intelligence," objectively determined and measured." *Journal of Personality and Social Psychology*, 86, 96–111.
- Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for math–science expertise. *Perspectives in Psychological Science*, 1, 316–345.
- Lubinski, D. S., & Benbow, C. P. (2007). Sex differences in personal attributes for the development of scientific expertise. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 79–100). Washington, DC: American Psychological Association.
- Lubinski, D., Benbow, C. P., Shea, D. L., Eftekhari-Sanjani, H., & Halvorson, B. J. (2001). Men and women at promise for scientific excellence: Similarity not dissimilarity. *Psychological Science*, 12, 309–317.
- Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human talent over two decades. *Psychological Science*, 17, 194–199.
- Lummis, M., & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology*, 26, 254–263.
- Lynn, R. (1991). Racial differences in intelligence: A global perspective. *The Mankind Quarterly*, 31, 255–296.
- Lynn, R., & Irwing, P. (2004). Sex differences on the Progressive Matrices: A meta-analysis. *Intelligence*, 32, 481–498.
- Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Macintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, 28, 559–571.
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Science*, 97, 4395–4403.
- Malas, M. A., Dogan, S., Hilal Evcil, E., & Desdicioglu, K. (2006). Fetal development of the hand, digits and digit ratio (2D:4D). *Early Human Development*, 82, 469–475.
- Malouf, M. A., Migeon, C. J., Carson, K. A., Petrucci, L., & Wisniewski, A. B. (2006). Cognitive outcome in adult women affected by CAH due to 21-hydroxylase deficiency. *Hormone Research*, 65, 142–150.
- Manning, J. T. (2002). *Digit ratio: A pointer to fertility, behaviour and health*. London: Rutgers University Press.
- Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist*, 63, 160–168.
- Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, 42, 236–243.
- Marulis, L., Warren, C., Uttal, D., & Newcombe, N. (2005, October). A meta-analysis: The effects of training on spatial cognition in children. Paper presented at the biennial meeting of the Cognitive Development Society, San Diego, CA.
- Mason, M. A., & Goulden, M. (2004). Marriage and baby blues: Redefining gender equity and the academy. *Annals of American Political and Social Scientist*, 596, 86–103.
- *Massachusetts Institute of Technology. (1999, March). A study on the status of women faculty in science at MIT. *MIT Faculty Newsletter*, 11(4). Retrieved June 10, 2008, from <http://web.mit.edu/fnl/women/women.html>
- Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, 23, 337–341.
- Mau, W.-C., & Lynn, R. (2000). Gender differences in homework and test scores in mathematics, reading and science at tenth and twelfth grade. *Psychology, Evolution, & Gender*, 2, 119–125.
- Mauleón, E., & Bordons, M. (2006). Productivity, impact, and publication habits by gender in the area of materials science. *Scientometrics*, 66, 199–218.
- McDowell, J. M. (1982). Obsolescence of knowledge and career publication profiles: Some evidence of differences among fields in costs of interrupted careers. *American Economics Review*, 72, 752–768.
- McFadden, D., Loehlin, J. C., Breedlove, S. M., Lippa, R. A., Manning, J. T., & Rahman, Q. (2005). A reanalysis of five studies on sexual orientation and the relative length of the 2nd and 4th fingers (the 2D:4D ratio). *Archives of Sexual Behavior*, 34, 341–356.
- McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, 86, 889–917.
- McGuinness, D., & Morley, C. (1991). Sex differences in the development of visuo-spatial ability in pre-school children. *Journal of Mental Imagery*, 15, 143–150.
- *McIntyre, R. B., Lord, C. G., Gresky, D. M., Ten Eyck, L. L., Frye, G. D. J., & Bond, C. F., Jr. (2005). A social impact trend in the effects of role models on alleviating women's mathematics stereotype threat. *Current Research in Social Psychology*, 10(9), 116–137.
- McKeever, W. F., Rich, D., Deyo, R., & Conner, R. (1987). Androgens and spatial ability: Failure to find a relationship between testosterone and ability measures. *Bulletin of the Psychonomic Society*, 25, 438–440.
- *McLaughlin, T. (2006, December 27). *Top Wall Street jobs still elude women, minorities*. Boston: Boston Globe. Retrieved September 19, 2008, from http://www.boston.com/news/education/higher/articles/2006/12/27/top_wall_street_jobs_still_elude_women_minorities/
- Mervis, J. (2007, March 16). Education research: U.S. math tests don't line up. *Science*, 315, 1485.
- *Moffat, S. D., & Hampson, E. (1996). A curvilinear relationship between testosterone and spatial cognition in humans. *Psychoneuroendocrinology*, 21, 323–337.
- Moffat, S. D., Zonderman, A., Metter, E., Blackman, M., Harman, S., & Resnick, S. M. (2002). Longitudinal assessment of serum free testosterone concentration predicts memory and cognitive status in men. *Journal of Clinical Endocrinology and Metabolism*, 87, 5001–5007.
- Monastersky, R. (2005, March 4). Editorial. *The Chronicle of Higher Education*. 51, p. A1.
- Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science*, 19, 1063–1066.
- Morgan, S. L., & Mehta, J. D. (2004). Beyond the laboratory: Evaluating the survey evidence for disidentification explanation of Black–White differences in achievement. *Sociology of Education*, 77, 82–101.

- Muller, C. (1998). Gender differences in parental involvement and adolescents' mathematics achievement. *Sociology of Education*, 71, 336–356.
- Mullis, I. V. S., Martin, M. O., Fierros, E. G., & Goldberg, A. L. (2000). *Gender differences in achievement: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College. International Study Center. Retrieved May 1, 2008, from [timss1995i/TIMSSPDF/T95_GChapter%201.pdf](http://timss.bc.edu/timss1995i/TIMSSPDF/T95_GChapter%201.pdf)
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2005). *TIMSS 2003 international report on achievement in the mathematics cognitive domains: Findings from a developmental project*. Boston: TIMSS & PIRLS International Study Center. Retrieved November 11, 2005, from http://timss.bc.edu/PDF/t03_download/T03MCOGDRPT.pdf
- Murphy, M. C., Steele, C., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in mathematics, science, and engineering settings. *Psychological Science*, 18, 879–885.
- National Science Foundation. (2005). *Science and engineering indicators 2004: Elementary and secondary education, mathematics and science coursework and student achievement*. Arlington, VA: National Science Foundation.
- National Science Foundation. (2006). *Science and engineering indicators 2006* (Vol. 1 NSB 06–01, Vol. 2, NSB 06–01A). Arlington, VA: National Science Foundation.
- National Science Foundation. (2007). *Women, minorities, and persons with disabilities in science and engineering: 2007* (NSF 07–315).
- Nations Report Card. (2007). *Twelfth-grade NAEP mathematics achievement-level results*. Washington, DC: National Center for Education Statistics. Retrieved February 22, 2007, from http://nationsreportcard.gov/reading_math_grade12_2005/s0307
- Neave, N., Menaged, M., & Weightman, D. (1999). Sex differences in cognition: The role of testosterone and sexual orientation. *Brain and Cognition*, 41, 245–262.
- Nelson, D., & Brammer, C. N. (2008). Women in science: A top-down approach. *Science*, 320, 1159–1160.
- Nettle, D. (2007). Empathizing and systemizing: What are they, and what do they contribute to our understanding of psychological sex differences? *British Journal of Psychology*, 98, 237–255.
- Newcombe, N. S. (2007). Taking science seriously: Straight thinking about spatial sex differences. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Eminent researchers debate the evidence on a key controversy of our time* (pp. 69–78). Washington, DC: American Psychological Association.
- Nie, N. H., & Golde, S. D. (2008, June/July). Does education really make you smarter? *Miller-McCune: Turning Research Into Solutions* 1(2), 56–65.
- Nordvik, H., & Amponsah, B. (1998). Gender differences in spatial abilities and spatial activity among university students in an egalitarian educational system. *Sex Roles*, 38, 1009–1023.
- Nuttall, R. L., Casey, M. B., & Pezaris, E. (2005). Spatial ability as a mediator of gender differences on mathematics tests. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 121–142). Cambridge, United Kingdom: Cambridge University Press.
- O'Boyle, M. W., Cunningham, R., Silk, T. J., Vaughan, D., Jackson, G., Syngieniotis, A., & Egan, G. F. (2005). Mathematically gifted male adolescents activate a unique brain network during mental rotation. *Cognitive Brain Research*, 25, 583–587.
- *O'Brien, T. (2006, March 19). Why do so few women reach the top of big law firms? *The New York Times*, p. B-1. Retrieved January 4, 2007, from <http://www.nytimes.com/2006/03/19/business/yourmoney/19law.html?ex=1300424400&en=f9b2756ce77b02bc&ei=5088&partner=rssnyt&emc=rss>
- Orsini, A., Schiappa, O., & Grossi, D. (1981). Sex and cultural differences in children's spatial and verbal memory span. *Perceptual and Motor Skills*, 53, 39–42.
- *Papierno, P. B., & Ceci, S. J. (2006). Promoting equity or inducing disparity: The costs and benefits of widening achievement gaps through universal interventions. *The Georgetown Public Policy Review*, 10, 1–15.
- Park, G., Lubinski, D., & Benbow, C. P. (2007). Contrasting intellectual patterns for creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science*, 18, 948–952.
- Park, G., Lubinski, D., & Benbow, C. P. (2008). Ability differences among people who have commensurate degrees matter for scientific creativity. *Psychological Science*, 19, 957–961.
- *Parsons, J. E., Adler, T. F., & Kaczala, C. M. (1982). Socialization of achievement attitudes and beliefs: Parental influences. *Child Development*, 53, 310–321.
- Penner, A. M. (2003). International gender-item difficulty interactions in mathematics and science achievement tests. *Journal of Educational Psychology*, 95, 650–655.
- Penner, A. M. (2005, August). *A quantile regression approach to gender differences in international mathematics*. Paper presented at the summer meeting of the International Sociological Association Research Committee on Social Stratification and Mobility, Los Angeles.
- Penner, A. M. (in press). Gender differences in extreme mathematics achievement: An international perspective on biological and social factors. *American Journal of Sociology*.
- Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, 37, 239–253.
- Peters, T., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotation Test: Different versions and factors that affect performance. *Brain and Cognition*, 28, 39–58.
- Pinker, S. (2002). *The blank slate*. New York: Viking.
- *Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance. *Journal of Educational Psychology*, 94, 396–404.
- Preston, A. (2004, Summer). Plugging leaks in the scientific workforce. *Issues in Science & Technology*, pp. 69–74.
- Pribyl, J. R., & Bodner, G. M. (1987). Spatial ability and its role in organic chemistry: A study of four organic courses. *Journal of Research in Science Teaching*, 24, 229–240.
- Puts, D. A., McDaniel, M. A., Jordan, C. L., & Breedlove, S. M. (2008). Spatial ability and prenatal androgens: Meta-analyses of congenital adrenal hyperplasia. *Archives of Sexual Behavior*, 37, 100–111.
- *Quaiser-Pohl, C., Geiser, C., & Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental-rotation ability. *Personality and Individual Differences*, 40, 609–619.
- Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, 57, 55–72.
- Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, 19, 1067–1070.
- RAND. (2005). *Is there gender bias in federal grant programs?* (RAND Infrastructure, Safety, and Environment Research Brief RB-9147-NSF). Retrieved March 1, 2008, from http://rand.org/pubs/research_briefs/RB9147/RAND_RB9147.pdf
- Rathbun, A., West, J., & Germino-Hausken, E. (2004). *From kindergarten through third grade: Children's beginning school experiences*. Washington, DC: National Center for Education Statistics.
- Ravitch, D. (1998, December 17). Girls are beneficiaries of gender gap. *The Wall Street Journal. Education*, p. A22.
- Raymond, C. L., & Benbow, C. P. (1986). Gender differences in mathematics: A function of parental support and student sex typing? *Developmental Psychology*, 22, 808–819.
- Reis, S. M., & Park, S. (2001). Gender differences in high-achieving

- students in math and science. *Journal for the Education of the Gifted*, 25, 52–73.
- Resnick, S. M., Berenbaum, S. A., Gottesman, I. I., & Bouchard, T. J. (1986). Early hormonal influences on cognitive functioning in congenital adrenal hyperplasia. *Developmental Psychology*, 22, 191–198.
- Rhoads, S. E. (2004). *Taking sex differences seriously*. San Francisco: Encounter Books.
- Rhode, D. L. (1997). *Speaking of sex: The denial of gender inequality*. Cambridge, MA: Harvard University Press.
- Ripa, C. P. L., Johannsen, T. H., Mortensen, E. L., & Muller, J. (2003). General cognitive functions, mental rotations ability, and handedness in adult females with CAH [Abstract]. *Hormones and Behavior*, 44, 72
- *Robert, M., & Chevrier, E. (2003). Does men's advantage in mental rotation persist when real three-dimensional objects are either felt or seen? *Memory & Cognition*, 31, 1136–1145.
- Robinson, N. M., Abbott, R. D., Berninger, V. W., & Busse, J. (1996). The structure of abilities in math-precocious young children. *Journal of Educational Psychology*, 88, 341–352.
- Robinson, N. M., Abbott, R. D., Berninger, V. W., Busse, J., & Mukhopadhyah, S. (1997). Developmental changes in mathematics precocious young children. *Gifted Child Quarterly*, 41, 145–158.
- Roof, R. L., Zhang, Q., Glasier, M. M., & Stein, D. G. (1993). Gender-specific impairment on Morris water maze task after entorhinal cortex lesion. *Behavioural Brain Research*, 57, 47–51.
- Royer, J. M., & Garofoli, L. M. (2005). Cognitive contributions to sex differences in mathematics performance. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics* (pp. 99–120). New York: Cambridge University Press.
- Rudd, E., Morrison, E., Picciano, J., & Nerad, M. (2008). Finally equal footing for women in social science careers? *CIRGE Spotlight on Doctoral Education #1*. Seattle, WA: University of Washington, Center for Innovation and Research in Graduate Education. Retrieved July 23, 2008 from www.cirge.washington.edu
- Rushton, J. P. (1992a). Cranial capacity related to sex, rank, and race in a stratified random sample of 6,325 U.S. military personnel. *Intelligence*, 16, 401–413.
- Rushton, J. P. (1992b). The brain size IQ debate. *Nature*, 360(6402), 292.
- Rushton, J. P., & Ankney, C. D. (2007). The evolution of brain size and intelligence. In S. M. Platek & T. K. Shackelford (Eds.), *Evolutionary cognitive neuroscience* (pp. 121–161). Cambridge, MA: MIT Press.
- *Rutter, M., & Silberg, J. (2002). Gene–environment interplay in relation to emotional and behavioral disturbance. *Annual Review of Psychology*, 53, 463–490.
- *Sackett, P., Hardison, C., & Cullen, M. (2004). On interpreting stereotype threat as accounting for African American–White differences on cognitive tests. *American Psychologist*, 59, 7–13.
- Sanders, G. (2007). Testing predictions from the hunter–gatherer hypothesis2: Sex differences in the visual processing of near and far space. *Evolutionary Psychology*, 5, 666–679.
- Sanders, G., Bereczkei, T., Csatho, A., & Manning, J. (2005). The ratio of the 2nd to 4th finger length predicts spatial ability in men but not women. *Cortex*, 41, 789–795.
- Sanders, G., Sjodin, M., & deChastelaine, M. (2002). On the elusive nature of sex differences in cognition: Hormonal influences contributing to within-sex variation. *Hormones and Behavior*, 31, 145–152.
- Schattman, L., & Sherwin, B. B. (2007). Effects of the pharmacologic manipulation of testosterone on cognitive functioning in women with polycystic ovary syndrome: A randomized, placebo-controlled treatment study. *Hormones and Behavior*, 51, 579–686.
- Scheibinger, L. (1987). The history and philosophy of women in science: A review essay. *Signs: Journal of Women in Culture and Society*, 12, 305–332.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440–452.
- Schratz, M. M. (1978). A developmental investigation of sex differences in spatial (visual–analytic) and mathematical skills in three ethnic groups. *Developmental Psychology*, 14, 263–267.
- *Seid, J. (2006). *10 best-paid executives: They're all men*. New York: CNN. Retrieved January 4, 2007, from <http://money.cnn.com/2006/10/03/news/newsmakers/mpwpay/index.htm>
- *Serbin, L. A., Zerkowicz, P., Doyle, A. B., Gold, D., & Wheaton, B. (1990). The socialization of sex-differentiated skills and academic performance: A mediational model. *Sex Roles*, 23, 613–628.
- Shalala, D. E., Agogino, A. M., Bailyn, L., Birgeneau, R., Cauce, A. M., Deangeles, C., et al. (2007). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*. Washington, DC: National Academies Press.
- Shayer, M., Ginsberg, D., & Coe, R. (2007). 30 years on—A large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975–2003. *British Journal of Educational Psychology*, 77, 25–41.
- Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93, 604–614.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703.
- Sherman, J. A. (1967). The problem of sex differences in space perception and aspects of intellectual functioning. *Psychological Review*, 74, 290–299.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science* 10, 80–83.
- Siegel, A. W., & Schadler, M. (1977). The development of young children's spatial representations of their classroom. *Child Development*, 48, 388–394.
- Slabbekoorn, D., van Goozen, S. H., Megens, J., Gooren, L. J. G., & Cohen-Kettenis, P. T. (1999). Activating effects of cross-sex hormones on cognitive functioning. *Psychoneuroendocrinology*, 24, 432–447.
- *Sleet, D. A. (1985). Differences in the social complexity of children's play choices. *Perceptual and Motor Skills*, 60, 283–287.
- Sommers, C. H. (2008, March/April). Why can't a woman be more like a man? *The American*, 2(2). Retrieved September 19, 2008, from www.american.com/archive/2008/march-april-magazine-contents/why-can2019t-a-woman-be-more-like-a-man
- Sorby, S. A. (2001). A course in spatial visualization and its impact on the retention of female engineering students. *Journal of Women and Minorities in Science and Engineering*, 7, 153–172.
- Sorby, S. A. (2005). Impact of changes in course methodologies on improving spatial skills. *Journal of Geometry and Graphics*, 9, 99–105.
- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, 60, 950–958.
- Spelke, E. S., & Grace, A. D. (2007). Sex, mathematics, and science. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Eminent researchers debate the evidence on a key controversy of our time* (pp. 57–68). Washington, DC: American Psychological Association.
- *Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's mathematics performance. *Journal of Experimental Social Psychology*, 35, 4–28.
- Spinath, F., Spinath, B., & Plomin, R. (2008). The nature and nurture of intelligence and motivation in the origins of sex differences in elementary school achievement. *European Journal of Personality*, 22, 211–229.
- *Sprafkin, C., Serbin, L. A., Denier, C., & Connor, J. M. (1983). Sex-differentiated play: Cognitive consequences and early interventions. In M. B. Liss (Ed.), *Social and cognitive skills: Sex roles and children's play* (pp. 167–192). New York: Academic Press.

- *Stafford, R. E. (1972). Hereditary and environmental components of quantitative reasoning. *Review of Educational Research*, 42, 183–201.
- Stanford University. (2005, February 9). No evidence of innate gender differences in mathematics and science, scholars assert. *Stanford Report*. Palo Alto, CA: Author.
- Stanley, J., Keating, D. P., & Fox, L. H. (1974). *Mathematical talent: Discovery, description, and development*. Baltimore, MD: The Johns Hopkins University Press.
- *Steele, C. M. (1997). A threat is in the air: How stereotypes shape "intellectual identity and performance." *American Psychologist*, 52, 613–629.
- *Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.
- Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The impact of gender on the review of the CVs of job applicants and tenure candidates: A national empirical study. *Sex Roles*, 41, 509–528.
- Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive abilities test scores: A UK National Picture. *British Journal of Educational Psychology*, 76, 463–480.
- Strenta, A. C., Elliott, R., Adair, R. Matier, M., & Scott, J. (1994). Choosing and leaving science in highly selective institutions. *Research in Higher Education*, 35, 513–547.
- Stricker, L. J. (2006). Stereotype threat: A clarification. *Science*, 312, 1310–1312.
- Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology*, 34, 665–693.
- Stricker, L. J., & Ward, W. C. (2008). Stereotype threat in applied settings re-examined: A reply. *Journal of Applied Social Psychology*, 38, 1656–1663.
- Stumpf, H., & Stanley, J. C. (1996). Gender-related differences on the College Board's advanced placement and achievement tests, 1982–1992. *Journal of Educational Psychology*, 88, 353–364.
- Stumpf, H., & Stanley, J. C. (1998). Stability and change in gender-related differences on the College Board advanced placement and achievement tests. *Current Directions in Psychological Science*, 7, 192–196.
- Summers, L. H. (2005, January 14). *Remarks on diversifying the science and engineering workforce*. Cambridge, MA: National Bureau of Economic Research. Retrieved July 29, 2006, from <http://www.president.harvard.edu/speeches/2005/nber.html>
- Swiatek, M. A., Lupkowski-Shopluk, A., & O'Donoghue, C. C. (2000). Gender differences in above-level EXPLORE scores of gifted third through sixth graders. *Journal of Educational Psychology*, 92, 718–723.
- Tartre, L. A. (1990). Spatial skills, gender, and mathematics. In E. H. Fennema & G. C. Leder (Eds.), *Mathematics and gender* (pp. 27–59). New York: Teachers College Press.
- Tendick, F., Downes, M., Goktekin, T., Cavusoglu, M., Feygin, D., Wu, X., et al. (2000). A virtual environment testbed for training laparoscopic surgical skills. *Presence: Teleoperators & Virtual Environments*, 9, 236–255.
- Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The role of computer and videogame usage to gender differences in mental rotation. *Sex Roles*, 53, 433–441.
- Terlecki, M. S., Newcombe, N. S., & Little, M. (2007). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology*, 22, 996–1013.
- Thilers, P., MacDonald, S. W. S., & Herlitz, A. (2006). The association between endogenous free testosterone and cognitive performance: A population-based study. *Psychoneuroendocrinology*, 31, 565–576.
- Thurstone, L., & Thurstone, T. (1962). *The primary mental abilities test*. Chicago: Science Research Associates.
- Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse & Society*, 14, 191–220.
- Uttal, D. H., Gregg, V. H., Tan, L. S., Chamberlain, M., & Sines, A. (2001). Connecting the dots: Children's use of a systematic figure to facilitate mapping and search. *Developmental Psychology*, 37, 338–350.
- Valian, V. (1998). *Why so slow? The advancement of women*. Cambridge, MA: MIT Press.
- Valian, V. (2007). Women at the top in science—and elsewhere. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 27–37). Washington, DC: American Psychological Association.
- van Anders, S., & Hampson, E. (2005). Testing the prenatal androgen hypothesis: Measuring digit ratios, sexual orientation, and spatial abilities in adults. *Hormones and Behavior*, 47, 92–98.
- Van Goozen, S. H. M., Cohen-Kettenis, P. T., Gooren, L. J. G., Frijda, N. H., & Van de Poll, N. E. (1994). Activating effects of androgens on cognitive performance. *Neuropsychologia*, 32, 1153–1157.
- Van Goozen, S. H. M., Cohen-Kettenis, P. T., Gooren, L. J. G., Frijda, N. H., & Van de Poll, N. E. (1995). Gender differences in behaviour: Activating effects of cross-sex hormones. *Psychoneuroendocrinology*, 20, 343–363.
- *Vandenberg, S. G. (1975). Sources of variation in performance of spatial tasks. In J. Eliot & N. J. Salkind (Eds.), *Children's spatial development* (pp. 141–163). Springfield, IL: Charles C Thomas.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599–604.
- Vasta, R., Knott, J. A., & Gaze, C. E. (1996). Can spatial training erase the gender differences on the water-level task? *Psychology of Women Quarterly*, 20, 549–567.
- Voyer, D., Boyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117, 250–270.
- *Waber, D. P. (1977). Sex differences in mental abilities, hemispheric lateralization, and rate of physical growth at adolescence. *Developmental Psychology*, 13, 29–38.
- Waber, D. P., DeMoor, C., Forbes, P., Almi, C., Botteron, G., Leonard, D., et al. (2007). The NIH MRI study of normal brain development: Performance of a population based sample of healthy children aged 6 to 18 years on a neuropsychological battery. *Journal of the International Neuropsychological Society*, 13, 729–746.
- Wai, J., Lubinski, D., & Benbow, C. P. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology*, 97, 484–492.
- Wai, J., Lubinski, D., & Benbow, C. P. (in press). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*.
- Wainer, H., & Steinberg, L. S. (1992). Sex differences in performance on the Mathematics section of the Scholastic Aptitude Test: A bidirectional validity study. *Harvard Educational Review*, 62, 323–336.
- Watson, N. V., & Kimura, D. (1991). Nontrivial sex differences in throwing and intercepting: Relation to psychometrically-defined spatial functions. *Personality and Individual Differences*, 12, 375–385.
- Webb, R. M., Lubinski, D., & Benbow, C. P. (2002). Mathematically facile adolescents with math-science aspirations: New perspectives on their educational and vocational development. *Journal of Educational Psychology*, 94, 785–794.
- Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, 99, 397–420.
- Webb, T. J., O'Hara, B., & Freckleton, R. (2008). Does double-blind review benefit female authors? *Trends in Ecology and Evolution*, 23, 351–353.

Weinberger, C. J. (2005, January). *Is the science and engineering workforce drawn from the far upper tail of the math ability distribution?* Paper presented at the National Bureau of Economic Research conference on diversifying the science & engineering workforce, Cambridge, MA.

Wenneras, C., & Wold, A. (1997). Nepotism and sexism in peer review. *Nature*, 387, 341-343.

*Williams, C. L., & Meck, W. H. (1991). The organizational effects of gonadal steroids on sexually dimorphic spatial ability. *Psychoneuroendocrinology*, 16, 155-176.

Williams, T. J., Pepitone, M. E., Christensen, S. E., Cooke, B. M., Huberman, A. D., Breedlove, N. J., et al. (2000). Finger-length ratios and sexual orientation. *Nature*, 404, 455-456.

Wise, L. L., Steel, L., & MacDonald, C. (1979). *Origins and career consequences of sex differences in high school mathematics achievement* (Report to the National Institute of Education on its Grant NIE-G-78-001). Palo Alto, CA: American Institutes for Research.

Wolfgang, C. H., Stannard, L. L., & Jones, I. (2001). Block play performance among preschoolers as a predictor of later school achievement in mathematics. *Journal of Research in Childhood Education*, 15, 173-180.

Xie, Y., & Shauman, K. A. (2003). *Women in science: Career processes and outcomes*. Cambridge, MA: Harvard University Press.

Young, J. W. (1991). Gender bias in predicting college academic performance: A new approach using item response theory. *Journal of Educational Measurement*, 28, 37-47.

Young, J. W. (1994). Differential prediction of college grades by gender and by ethnicity: A replication study. *Educational and Psychological Measurement*, 54, 1022-1029.

Received October 23, 2007
 Revision received July 29, 2008
 Accepted September 2, 2008 ■



**AMERICAN PSYCHOLOGICAL ASSOCIATION
 SUBSCRIPTION CLAIMS INFORMATION**

Today's Date: _____

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do **NOT** duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION	MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL)	
ADDRESS	DATE YOUR ORDER WAS MAILED (OR PHONED)	
CITY	<input type="checkbox"/> PREPAID <input type="checkbox"/> CHECK <input type="checkbox"/> CHARGE CHECK/CARD CLEARED DATE: _____	
STATE/COUNTRY	ZIP	
YOUR NAME AND PHONE NUMBER	ISSUES: <input type="checkbox"/> MISSING <input type="checkbox"/> DAMAGED	
TITLE	VOLUME OR YEAR	NUMBER OR MONTH

Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4-6 weeks.

(TO BE FILLED OUT BY APA STAFF)

DATE RECEIVED: _____	DATE OF ACTION: _____
ACTION TAKEN: _____	INV. NO. & DATE: _____
STAFF NAME: _____	LABEL NO. & DATE: _____

Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242

PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.