

Women in Academic Science: A Changing Landscape

Stephen J. Ceci¹, Donna K. Ginther², Shulamit Kahn³, and Wendy M. Williams¹

¹Department of Human Development, Cornell University; ²Department of Economics, University of Kansas; and ³School of Management, Boston University

Summary

Much has been written in the past two decades about women in academic science careers, but this literature is contradictory. Many analyses have revealed a level playing field, with men and women faring equally, whereas other analyses have suggested numerous areas in which the playing field is not level. The only widely-agreed-upon conclusion is that women are underrepresented in college majors, graduate school programs, and the professoriate in those fields that are the most mathematically intensive, such as geoscience, engineering, economics, mathematics/computer science, and the physical sciences. In other scientific fields (psychology, life science, social science), women are found in much higher percentages.

In this monograph, we undertake extensive life-course analyses comparing the trajectories of women and men in math-intensive fields with those of their counterparts in non-math-intensive fields in which women are close to parity with or even exceed the number of men. We begin by examining early-childhood differences in spatial processing and follow this through quantitative performance in middle childhood and adolescence, including high school coursework. We then focus on the transition of the sexes from high school to college major, then to graduate school, and, finally, to careers in academic science.

The results of our myriad analyses reveal that early sex differences in spatial and mathematical reasoning need not stem from biological bases, that the gap between average female and male math ability is narrowing (suggesting strong environmental influences), and that sex differences in math ability at the right tail show variation over time and across nationalities, ethnicities, and other factors, indicating that the ratio of males to females at the right tail can and does change. We find that gender differences in attitudes toward and expectations about math careers and ability (controlling for actual ability) are evident by kindergarten and increase thereafter, leading to lower female propensities to major in math-intensive subjects in college but higher female propensities to major in non-math-intensive sciences, with overall science, technology, engineering, and mathematics (STEM) majors at 50% female for more than a decade. Post-college, although men with majors in math-intensive subjects have historically chosen and completed PhDs in these fields more often than women, the gap has recently narrowed by two thirds; among non-math-intensive STEM majors, women are more likely than men to go into health and other people-related occupations instead of pursuing PhDs.

Importantly, of those who obtain doctorates in math-intensive fields, men and women entering the professoriate have equivalent access to tenure-track academic jobs in science, and they persist and are remunerated at comparable rates—with some caveats that we discuss. The transition from graduate programs to assistant professorships shows more pipeline leakage in the fields in which women are already very prevalent (psychology, life science, social science) than in the math-intensive fields in which they are underrepresented but in which the number of females holding assistant professorships is at least commensurate with (if not greater than) that of males. That is, invitations to interview for tenure-track positions in math-intensive fields—as well as actual employment offers—reveal that female PhD applicants fare at least as well as their male counterparts in math-intensive fields.

Corresponding Author:

Stephen J. Ceci, Department of Human Development, Martha Van Rensselaer Hall, Cornell University, Ithaca, NY 14853
E-mail: sjc9@cornell.edu

Along these same lines, our analyses reveal that manuscript reviewing and grant funding are gender neutral: Male and female authors and principal investigators are equally likely to have their manuscripts accepted by journal editors and their grants funded, with only very occasional exceptions. There are no compelling sex differences in hours worked or average citations per publication, but there is an overall male advantage in productivity. We attempt to reconcile these results amid the disparate claims made regarding their causes, examining sex differences in citations, hours worked, and interests.

We conclude by suggesting that although in the past, gender discrimination was an important cause of women's underrepresentation in scientific academic careers, this claim has continued to be invoked after it has ceased being a valid cause of women's underrepresentation in math-intensive fields. Consequently, current barriers to women's full participation in mathematically intensive academic science fields are rooted in pre-college factors and the subsequent likelihood of majoring in these fields, and future research should focus on these barriers rather than misdirecting attention toward historical barriers that no longer account for women's underrepresentation in academic science.

Keywords

pipeline leakage, STEM, gender bias, cognitive sex differences, salary gap, productivity, citations, opting out, tenure, work-life balance

Introduction

We present a comprehensive synthesis of the empirical findings and logical analyses informing the question of why women are underrepresented in certain academic fields of science. Our emphasis is on those fields that are spatially and mathematically intensive—the ones in which women are most underrepresented—such as geoscience, engineering, economics, mathematics/computer science, and the physical sciences, including chemistry and physics (which we abbreviate as GEEMP).¹ In this article, we compare math-intensive fields to non-math-intensive fields of science, including life science, psychology, and social science (which we abbreviate as LPS), in which women are often at parity with or exceed the number of men. A burgeoning literature bears on this topic, produced by scholars from diverse disciplines (e.g., psychology, economics, sociology, endocrinology, mathematics, philosophy, bibliometrics, and education), and there are many disagreements and confusions among researchers, the lay public, and policymakers.

These disagreements reside at multiple levels: (a) disagreements about the design and interpretation of studies (e.g., why 3-D mental rotation tasks show a male advantage whereas paper-folding tasks that entail similar processes do not; why some ethnic groups display gender² differences in math but others do not, or even show a reverse gender trend); (b) disagreements over which fields should be emphasized in studying female underrepresentation (e.g., although the GEEMP fields show large sex differences, so do some humanities fields, such as philosophy); (c) disagreements regarding which types of institutions exhibit the largest sex differences (teaching-intensive vs. research-intensive, or R1, institutions); (d) disagreements about which dependent variables should be the focus of analysis (e.g., matriculation and graduation rates, transition rates, rates of obtaining

tenure-track positions, interviewing outcomes, hiring outcomes, promotion rates, salary, job satisfaction, work-product evaluation, job persistence, grant-getting success, awards, number of publications, leaving science); (e) disagreements over the cause of pipeline leakage from high school to college to graduate school to career; (f) disagreements concerning how gender bias should be defined; (g) differences across disciplines in how the problem is framed (e.g., in terms of market forces vs. implicit processing vs. discrimination); and (h) disagreements concerning how productivity and excellence should be conceptualized.

We examine all of these variables in the following pages. Assembling and comparing these variables in different ways leads to interesting strata for analysis, such as salary gaps for Type of Field \times Type of Institution, or causes of attrition of female college majors in each field at the point of graduate-school entry.

There has been a tendency in the literature to conflate historical findings with current findings, thus obscuring both trends over time and the current state of the field. (This is particularly likely to happen when discussing hiring, persistence, and remuneration rates that may have changed in recent decades or even in recent years.) In fact, the results we present below show such a dramatic increase in the number of women in science at all levels over the past 40 years that research based on data prior to the 1990s may have little bearing on the current circumstances women encounter. As a result, we present the most recent data available, and our synthesis of the literature emphasizes recent studies—those done since 2000—augmented by our own analyses of recent data in order to shed light on the current situation for women in science, rather than on what was once historically true.

Our attempt to resolve arguments and inconsistencies in the literature and media relies on a life-course framework. We examine causal factors beginning in early

childhood (e.g., sex differences in cognitive processing during infancy that may combine with gendered socialization experiences during the preschool period). These early-childhood factors are then linked to differences in mathematics achievement and math identity in high school (e.g., gaps in Advanced Placement, or AP, math and science coursework and subsequent expectations for college majors) and college (e.g., majors and graduate-school or career aspirations). This set of differences is then followed with an examination of gender differences in the completion of graduate programs and subsequent application for tenure-track academic positions, and, ultimately, with an investigation of the role of gender in tenure-track hiring in the academy (or, alternatively, outside it), the acquisition of tenure, and promotion to full professor. We provide new data on the attrition of women at some of these critical points and evaluate factors claimed to be responsible. To foreshadow our conclusions, we find that although women are underrepresented in GEEMP fields, the overall state of the academy (collapsing across the many hundreds of between-sex contrasts involving salary, promotion, type of institution, type of field, and transition points) is largely one of gender neutrality—with some notable exceptions that should be of interest to members of specific fields.

Two of us come from psychology (Ceci and Williams) and two of us from economics (Ginther and Kahn). In the past, we worked on different aspects of this research, using different methods and paradigms. Although we had read each other's research, this *Psychological Science in the Public Interest* report marks the first time we have worked together. We hope that the integration of our different orientations has resulted in a novel perspective on the major arguments related to why women are underrepresented in certain fields. For example, whether a chilly climate drives professional women out of science, whether biased hiring is a culprit, or why experimental findings often have not been confirmed by naturalistic analyses are examples of questions on which our perspectives converged and informed each other, as a result of our reading each other's literatures and engaging in continual discussions.

For regular readers of *Psychological Science in the Public Interest*, this report only minimally overlaps with the journal's highly influential 2007 report by Halpern, Benbow, Geary, Gur, Hyde, and Gernsbacher (cited approximately 350 times as of this writing). The primary overlap is on the topic of sex differences in mathematical aptitude, which we have updated to cover research that has emerged since their report was published. However, this is a small part of the present report, which focuses instead on the factors responsible for the dearth of women professionals in math-intensive academic fields—including why so many women who are PhD recipients

opt not to pursue academic careers by not competing for tenure-track positions at R1 institutions, or accept such posts only to drop out or become stalled. Indeed, most of the evidence we draw on not only addresses a set of questions different from that of the Halpern et al. (2007) report (i.e., concerning salary, persistence, and promotion) but also addresses evidence that was not available at the time of their report, as most of the analyses reported herein are based on recent data, some available only in late 2013 and early 2014.

En route to answering questions about the dearth of women in math-intensive fields, we will examine earlier transition points in males' and females' lives, from primary school to college, from college to graduate school, from receipt of PhD to postdoctoral positions and assistant professorships, and upward to senior ranks. We will examine the sources of attrition of females at each point, the different effects of family aspirations and the presence of young children on the careers of women and men, and sex differences in productivity, remuneration, and promotion.³ We have thus attempted to provide a comprehensive report on the state of the art and science on the question of women's underrepresentation in math-intensive fields of academic science. Again, we have labeled these underrepresented fields GEEMP (for geoscience, engineering, economics, mathematics/computer science, and the physical sciences) to distinguish them from other science, technology, engineering, and mathematics (STEM) fields—life science, psychology, and social science (LPS)—in which women are well represented.

An Overview of the Problem

"Contradictory" is the word that best characterizes the literature on women's underrepresentation in academic science. There is agreement that women are underrepresented in all math-intensive fields in the academy. In all GEEMP fields in 2010, for example, women comprised only 25% to 44% of tenure-track assistant professors and 7% to 16% of full professors (our calculations here are based on the National Science Foundation's, or NSF's, 2010 Survey of Doctorate Recipients, or SDR: <http://ncesdata.nsf.gov/doctoratework/>). But there is heated debate over why women are so conspicuously absent in these fields compared with LPS fields. In the LPS fields, the comparable figures show that women hold 66% of the tenure-track assistant professorships in psychology, 45% in social science (excluding economics), and 38% in life science; for full professorships, the figures are 35%, 23%, and 24%, respectively.

In this section, we present the educational milestones leading to an academic science career by juxtaposing the percentage of women who complete each educational level (e.g., baccalaureate in a GEEMP field) with those

who complete the next level (PhD in a GEEMP field): Contrasts between rates of high school graduation, receipt of bachelor's degrees, receipt of PhDs, and assistant professorships will be used to frame the arguments in later sections of this article. Along the way, we note the junctures at which women are less—or more—likely to proceed to the next level, using the latest data available to perform these analyses. In the rest of this section, we disaggregate scientific fields to show the very different trends in female participation by field, before addressing potential explanations regarding female representation in scientific careers.

Educational milestones for women and men

Half of all 24- to 25-year-olds with at least a high school diploma are women, but women have represented more than half of bachelor's-degree recipients since the mid-1980s, and made up 57% of bachelor's-degree holders as of 2010 (Fig. 1a).

As mentioned previously, women are equally represented or overrepresented in some STEM fields (the LPS fields) and underrepresented in others (the GEEMP fields). Figure 1a shows that by 2011, the proportion of females among the bachelor's-degree-holding STEM majors was only a few percentage points below the proportion of females among all majors (averaging 6.5 percentage points), and was essentially the same as the proportion of females among high school graduates—with all exceeding 50%.

However, the contrast between GEEMP and LPS fields is stark. Women received only 25% of GEEMP bachelor's degrees in 2011, a more-than-30-percentage-point difference from the overall percentage of females among bachelor's-degree recipients. Moreover, after growing in the 1990s, the percentage of women in these majors has become increasingly smaller since 2002. In contrast, women are significantly overrepresented in LPS fields, receiving almost 70% of these bachelor's degrees. As in the GEEMP fields, the number of female baccalaureates in LPS fields grew in the 1990s; however, it has not fallen during the past decade. Thus, combining all STEM fields masks important differences in degree trends between GEEMP and LPS fields.

Figure 1b compares the percentage of female LPS and GEEMP bachelors, PhDs, and assistant professors from 1994 to 2011. Within the LPS fields, the percentage of female PhDs granted increased from 46.1% in 1994 to 57.9% in 2011. GEEMP female PhDs, which started at a much lower 16.8% in 1994, had increased to 26.3% by 2011. This growth was greater than the growth in bachelors 7 years earlier (the approximate interval between receipt of a bachelor's degree and a PhD in science).

Figure 2 directly compares cohort sizes of GEEMP and LPS bachelors with GEEMP and LPS PhDs 7 years later, for PhDs in the early 1990s and in 2007 through 2011, in essence creating artificial cohorts. In both periods, a smaller percentage of females than males proceeded from a STEM undergraduate major to a STEM PhD. This gender gap was particularly large for GEEMP fields in the early 1990s, yet by 2007 to 2011, the GEEMP gap had fallen by two thirds. There was a smaller gender gap in the likelihood of proceeding from undergraduate major to PhD in LPS fields than there was in GEEMP fields in the early 1990s. However, since then, the gender gap for LPS fields has narrowed by a much smaller amount than that for GEEMP fields, and this has mostly been due to fewer males getting LPS PhDs rather than to more females getting PhDs.

Figure 3 disaggregates the trends in female representation among STEM bachelors and STEM PhDs into more detailed fields and over a longer period of time. Figure 3a shows that since the mid-1990s (and, in some cases, before then), the majority of bachelor's degrees in psychology (>70%), life science (70%), and social science (>70%) have been awarded to women. The number of bachelor's degrees awarded to women in geoscience, the physical sciences, and engineering has more than doubled since the 1970s. In contrast, in mathematics/computer science, the number of bachelor's degrees awarded to women has dropped significantly. This has resulted disproportionately from women decreasing their numbers in one field: computer science.⁴

Figure 3b shows the percentage of PhDs by disaggregated field. With the exception of mathematics/computer science, patterns similar to those for bachelor's degrees emerge in all fields. In the most recent decade (and before that, for psychology), more than half of the PhDs in psychology and life science were awarded to women. By 2011, nearly half of PhDs in social science (excluding economics) were awarded to women. Geoscience, the physical sciences, mathematics/computer science, and engineering have shown tremendous growth in the numbers of PhDs awarded to women, from very small numbers (10% or less) in the 1970s. Both economics and engineering possess a higher percentage of female doctorates than they do of female bachelor's-degree holders.

As we follow scientists through the pipeline, we can see that women have increased their representation as tenure-track assistant professors in LPS fields, rising from 27.5% to 32.3%, and in GEEMP fields, rising much more markedly, from 14.3% to 22.7% (Fig. 1b). In Figure 4a, we show more disaggregated fields over a longer period. Within LPS fields, psychology had the highest percentage of female assistant professors throughout the period, followed by social science. The percentage of female

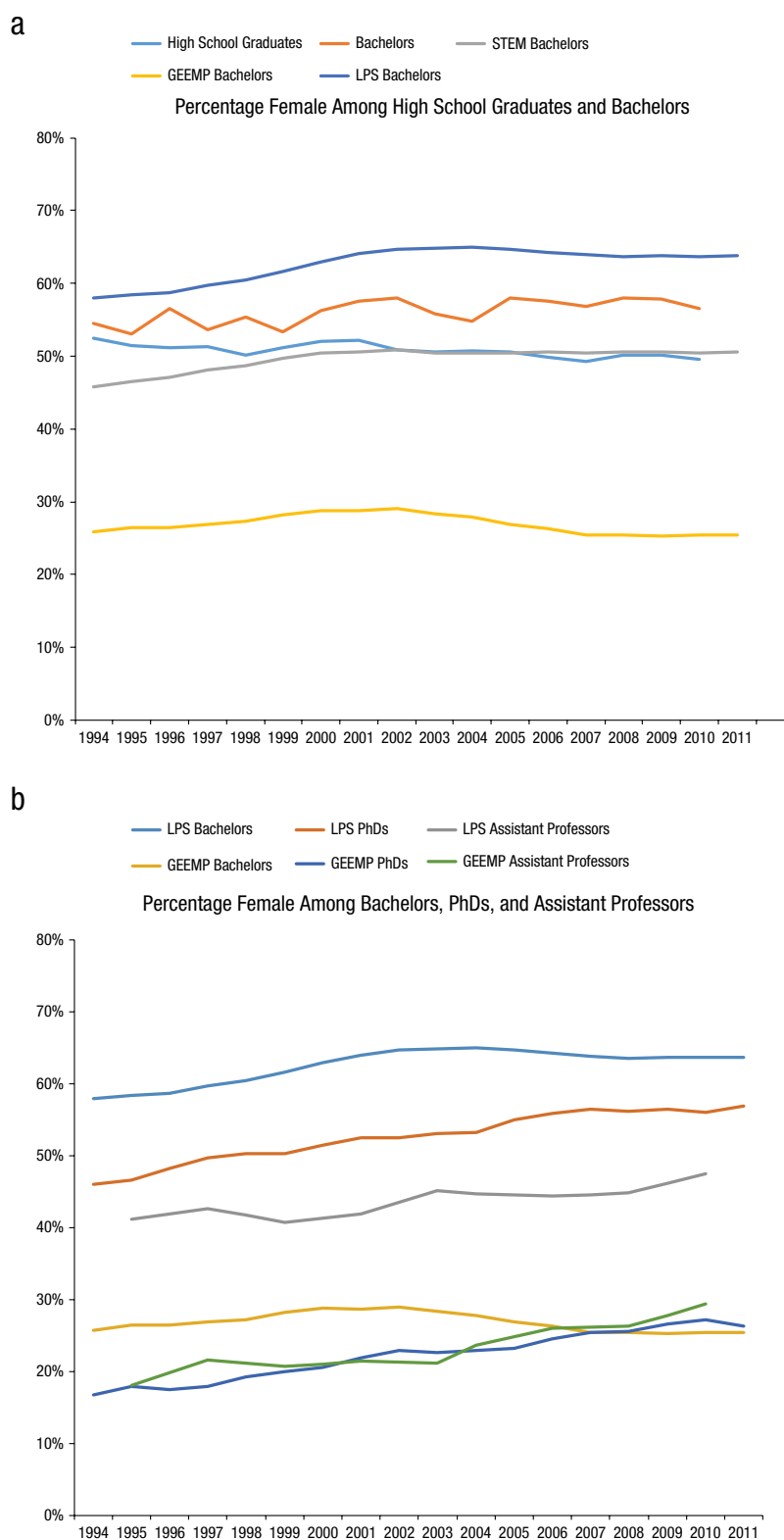


Fig. 1. Percentage female among high school graduates and bachelor's-degree holders (a) and among bachelor's-degree holders, PhDs, and tenure-track assistant professors (b) from 1994 to 2011, as a function of field (science, technology, engineering, and mathematics = STEM; life science, psychology, and social science = LPS; geoscience, engineering, economics, mathematics/computer science, and the physical sciences = GEEMP). The graph in (a) shows data for 24- and 25-year-olds drawn from the Current Population Survey Outgoing Rotations Data (<http://www.bls.gov/cps/#data>); data shown in the graph in (b) are drawn from the National Science Foundation's WebCASPAR database (<https://ncsesdata.nsf.gov/webcaspar/>).

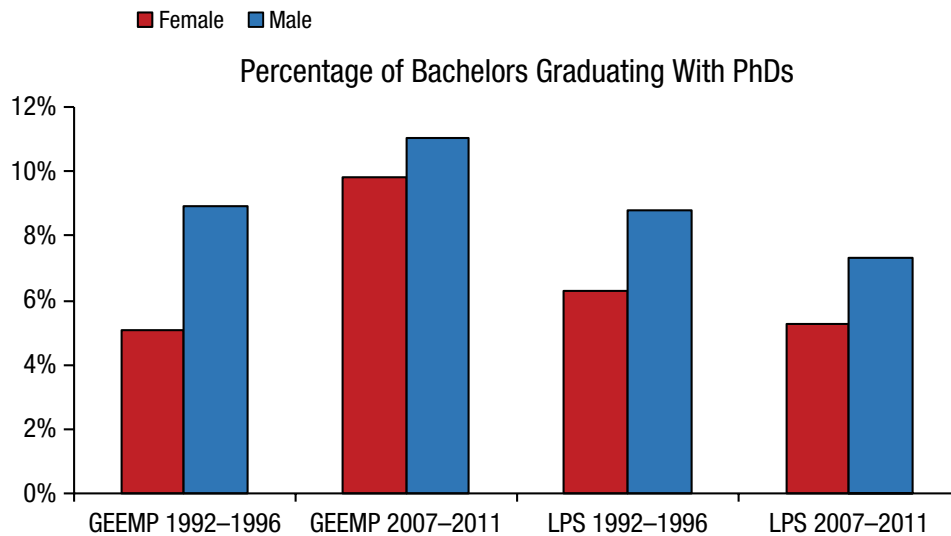


Fig. 2. Percentage of bachelor's-degree holders who attained PhDs as a function of field (geoscience, engineering, economics, mathematics/computer science, and physical science = GEEMP; life science, psychology, and social science = LPS), time period, and gender. Data shown here are drawn from the National Science Foundation's WebCASPAR database (<https://ncesdata.nsf.gov/webcaspar/>).

assistant professors in life science increased over the '70s, '80s, and early '90s, but has hovered between 30% and 40% in the 15 years since, despite the fact that women made up a continually increasing proportion of life-science PhDs. Within the GEEMP fields, engineering has shown the most remarkable growth, going from nearly 0% female in 1973 to 30% in 2010.

Transition from PhD to tenure track

To study the transition from PhDs to tenure-track assistant professorships, we again created artificial cohorts and compared the percentage of female PhDs to the percentage of female assistant professors 5 to 6 years later. This analysis indicated that in LPS fields, fewer women than men proceed to assistant professorship. Thus, the percentage of female assistant professors in LPS fields from 1993 to 1995 was 28.4%, compared with 41.6% of women in the corresponding PhD years—a gap of 13 percentage points, or almost one third of the women not progressing from PhD to assistant professor (Fig. 5). In recent years, this gap between the percentage of PhDs granted to women and the percentage of women with doctorates who subsequently assume assistant professorships has widened rather than narrowed: The percentage of female assistant professors from 2008 to 2010 was 31.6%, whereas the percentage of women in the corresponding PhD years was 53.2%—a gap of 22 percentage points. The overall gender gap for PhDs in biological sciences in moving from PhD to tenure-track assistant professorships during the earlier period was also documented in Ginther and Kahn (2009), even after controls for demographics, degree characteristics, and fields were added.

In contrast, in the GEEMP fields, women progressed from PhD to assistant professorships in approximately the same ratio as did men—a finding again similar to Ginther and Kahn's (2009) results. Thus, the percentage of women among assistant professors from 1993 to 1995 was 7% higher than the percentage of women in the corresponding PhD years. Among assistant professors from 2008 to 2010, the percentage of women was 5% lower than the percentage of women in the corresponding PhD years, a small drop (see Fig. 5).

Viewing all ranks together, women as a percentage of the total faculty (tenure track and tenured combined) grew from very low numbers in the early 1970s to approximately one third of faculty in geoscience and life science in 2010. In contrast, as Figure 4b shows, women still make up 20% or less of the total faculty in the mathematically intensive GEEMP fields.

Moreover, women make up a significantly larger portion of assistant professors—with rates in most GEEMP and LPS fields 10 percentage points higher than for all tenured and tenure-track faculty combined. This higher proportion of females among assistant professors in 2010 suggests that in the near future, female representation in both GEEMP and LPS fields will rise at tenured ranks, if promotion and retention is similar for female and male faculty. This is a big "if," and one we consider below.

Academic careers can and do evolve over time, and faculty often transition into administrative positions, becoming department chairs, deans, provosts, and university presidents. To examine this, we first analyzed the SDR, combining all fields, to compare the fractions of women and men who were in administrative positions at R1 institutions, and found no significant sex differences

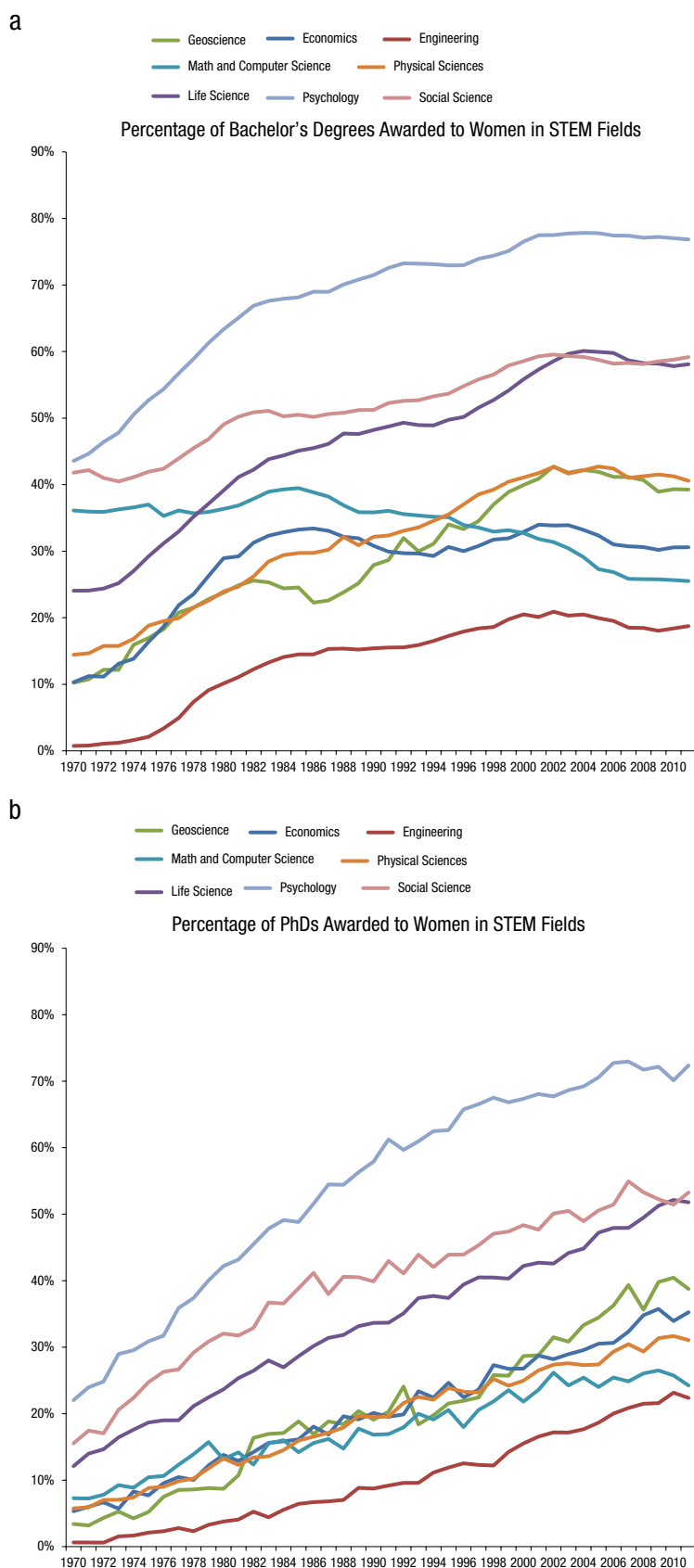


Fig. 3. Percentage of bachelor's degrees (a) and PhDs (b) awarded to women as a function of science, technology, engineering, and mathematics (STEM) discipline. Data shown here are drawn from the National Science Foundation's WebCASPAS database (<https://ncesdata.nsf.gov/webcaspar/>).

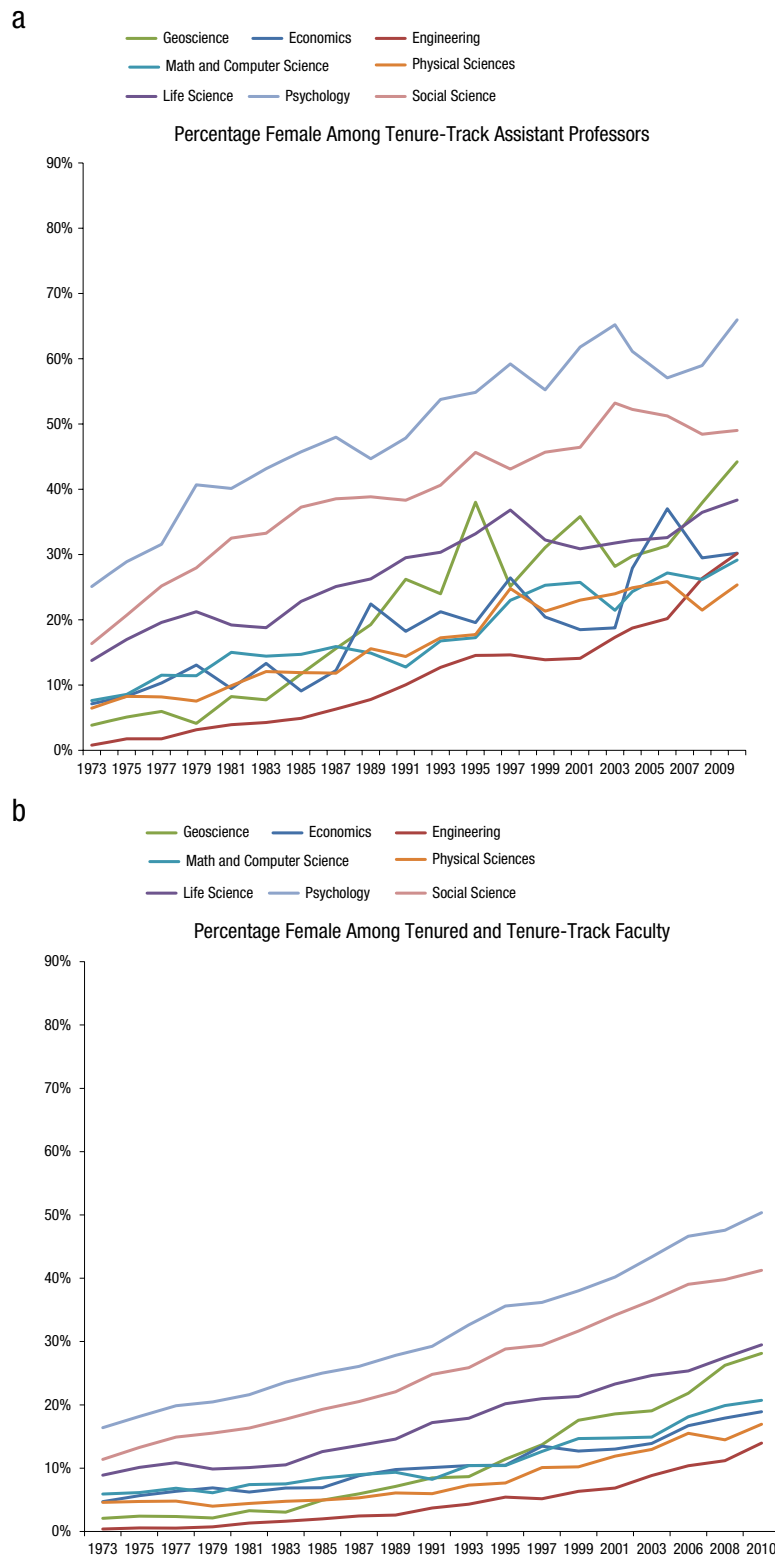


Fig. 4. Percentage female among tenure-track assistant professors (a) and among tenured or tenure-track faculty (b) from 1973 to 2010 as a function of field. Values shown are weighted percentages. Data shown here are drawn from the National Science Foundation's Survey of Doctorate Recipients (www.nsf.gov/statistics/srvdoctoratework).

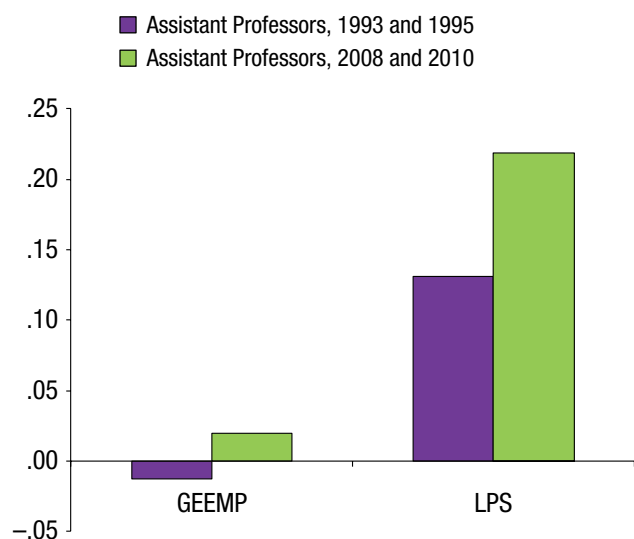


Fig. 5. Differences (in proportions) between the percentage of female PhDs and the percentage of female assistant professors 5 to 6 years post-PhD as a function of broad discipline (geoscience, engineering, economics, mathematics/computer science, and physical science, or GEEMP, vs. life science, psychology, and social science, or LPS) and cohort (1990s vs. 2000s). Positive values indicate that lower percentages of female than male PhDs entered tenure-track jobs. Calculated from WebCASPAP data and the National Science Foundation's Survey of Doctorate Recipients (www.nsf.gov/statistics/srvydoctoratework).

at any administrative level. We then determined whether there were sex differences in university administrative positions when all types of institutions were combined, and found some slight but significant sex differences: Women were less likely than men to be deans, directors, or department chairs (12.1% vs. 15.1%; $p < .01$) but were equally likely to be presidents, provosts, and chancellors (1.2% vs. 1.2%).

Thus, the points of leakage from the STEM pipeline depend on the broad discipline being entered—LPS or GEEMP. By graduation from college, women are over-represented in LPS majors but far underrepresented in GEEMP fields. In GEEMP fields, by 2011, there was very little difference in women's and men's likelihood to advance from a baccalaureate degree to a PhD and then, in turn, to advance to a tenure-track assistant professorship. Another way to think of this is that far fewer women are interested in (or perhaps capable in, as we discuss below) GEEMP fields to begin with, but once women are within GEEMP fields, their progress resembles that of male GEEMP majors. In contrast, whereas far more women than men major in LPS fields, in 2011, the gender difference in the probability of advancing from an LPS baccalaureate degree to a PhD was not trivial, and the gap in the probability of advancing from PhD to assistant professorship was particularly large, with fewer women than men advancing.

Evidence on Potential Explanations for Women's Underrepresentation in Academic GEEMP Careers

Although women's underrepresentation in math-intensive fields is not in doubt, its cause is hotly disputed. The disciplines of economics and psychology differ in their approach to and modeling of gender differences in career outcomes. In general, economists focus on *comparative advantage*, whereby individuals choose to work in areas where they are relatively more productive—weighing the costs and benefits of alternative careers and nonmarket activities—and on market forces balancing supply (based on comparative advantage) and demand (based on productivity); when these explanations are not supported by the data, economists try to understand why discrimination can be a self-reinforcing equilibrium. In contrast, psychology has tended to focus on early socialization practices, implicit and explicit biases, stereotypes, and biological sex differences in explaining this gap. In the face of the evidence just shown—indicating that the sources of the underrepresentation in GEEMP fields can be seen early, many years before college—the economists among us agree with the psychologists that early socialization and possibly even biological differences can lead to differences in comparative advantage. Moreover, they emphasize that anticipated gender differences in future career opportunities lead to behaviors and choices that reinforce early socialization, a position that psychologists also endorse. And the psychologists among us agree with the economists that productivity differences play an important role in explaining later persistence, promotion, and salary in some fields, as we describe later.

In LPS fields, the issues are different. Women drop off the academic ladder post-bachelor's. Here, too, the economists and psychologists on our team started by emphasizing different possible avenues of post-baccalaureate gender differences, with economists emphasizing rational choices, where the opportunity cost of balancing work and family is associated with not pursuing academic science, and psychologists emphasizing people-versus-things preferences that result in many STEM females opting for medicine, law, and veterinary science over GEEMP fields. Psychologists have charted large sex differences in occupational interests, with women preferring so-called "people-oriented" (or "organic," or natural-science) fields and men preferring "things" (people- and thing-oriented individuals are also termed "empathizers" and "systematizers," respectively; e.g., Auyeung, Lombardo, & Baron-Cohen, 2013). This people-versus-things construct goes back to Thorndike (1911) and is one of the salient dimensions running through vocational interests; it also represents a difference of 1 standard

deviation between men and women in vocational interests. Lippa has repeatedly documented very large sex differences in occupational interests, including in transnational surveys, with men more interested in “thing”-oriented activities and occupations, such as engineering and mechanics, and women more interested in people-oriented occupations, such as nursing, counseling, and elementary school teaching (e.g., Lippa, 1998, 2001, 2010).⁵ And in a very extensive meta-analysis of over half a million people, Su, Rounds, and Armstrong (2009) reported a sex difference on this dimension of a full standard deviation.⁶ However, despite differences between us at the start, over time our respective views on women’s migration from LPS appear to have converged, as readers will see.

Framed against the drop-offs depicted in Figures 1 through 5, we have organized our discussion of potential explanations for the underrepresentation of women in GEEMP careers by following the life course, beginning with prenatal hormones that are thought to influence cognition; continuing with sex differences in childhood socialization, cognitive aptitude, and achievement; and stretching through to sex differences in the academic-career outcomes of productivity, pay, and promotion. As will be seen, at each stage, we evaluate the evidence supporting each potential explanation for the ultimate underrepresentation of women in academic science careers and carry this evaluation forward to the next stage. We begin with the earliest developmental period for which empirical data are available.

Potential explanation #1: Mathematical- and spatial-ability differences from in utero through high school

Here, we consider argument and evidence in support of the claim that the shortage of women in math-intensive fields results in part from spatial- and quantitative-ability differences favoring males, some of which are alleged to emanate from early in life. We segue from this argument to a discussion of sex differences in later quantitative ability, including at the elite level of mathematical aptitude from which many academics in GEEMP hail.

Is there a prenatal basis for males’ early spatial and mathematical aptitude? It is routinely argued that behaviors that show sex differences in spatial and mathematical performance are influenced by androgen levels, especially those in the prenatal period for most mammals (Finegan, Niccols, & Sitarenios, 1992; Hines et al., 2003). Prenatal androgens are associated with a certain cognitive profile later in life, including mental

rotation ability (for reviews, see Auyeung et al., 2013; Falter, Arroyo, & Davis, 2006; Hines et al., 2003; Valla & Ceci, 2011). Male hormones are postulated to organize brain systems for a range of mechanisms for information processing.

However, a straightforward hormonal account of spatial and mathematical performance has been limited for several reasons. First, girls with exceptionally high levels of prenatal androgens (those afflicted with congenital adrenal hyperplasia, or CAH) do not consistently perform on later math and spatial-aptitude batteries as would be expected if male hormones organized the developing brain for optimized spatial processing. For example, Hines et al. (2003) showed that females with CAH did not perform better than unaffected females on mental rotations despite being exposed to much higher levels of male hormones prenatally (for a review, see Valla & Ceci, 2011).⁷ Second, Auyeung and her colleagues (2013) have shown that the hormonal influence is time- and dose-dependent and that a critical window exists beyond which the effect of hormones is greatly attenuated. Finally, in an ambitious analysis that connected structural and behavioral measures in a large sample of youth, Ingallhalikar et al. (2014) demonstrated sex differences in inter- versus intra-regional connectivity density. In men, there are more dense connections within regions and within hemispheres than between them, which optimizes doing one thing at a time very well, and greater focal intrahemispheric activation in males is an asset in spatial tasks. In women, by contrast, there are more long-range connections, which are especially suited for language and similar processing.

Vuoksima and her colleagues (2010) examined mental rotation performance as a function of twinship. The reasoning was that when a male baby is gestated, he is exposed to an androgen bath around the end of the first trimester, which has been associated with postnatal cognitive and personality outcomes, including 3-D mental rotation ability. The quantity of prenatal androgen is limited in the case of a female fetus except when her co-twin is a male. In this case, the concentration of male hormones in the female co-twin’s prenatal environment is larger than it would have been if her co-twin were female or if she had no twin at all. Vuoksima et al. (2010) reported that later in life, those females who had male twins (and thus were exposed to higher prenatal concentrations of androgen) performed better on the mental rotation test (MRT) than did females who shared their prenatal environment with female co-twins:

The superior MRT performance of female twins from opposite-sex pairs compared with female twins from same-sex pairs remained statistically

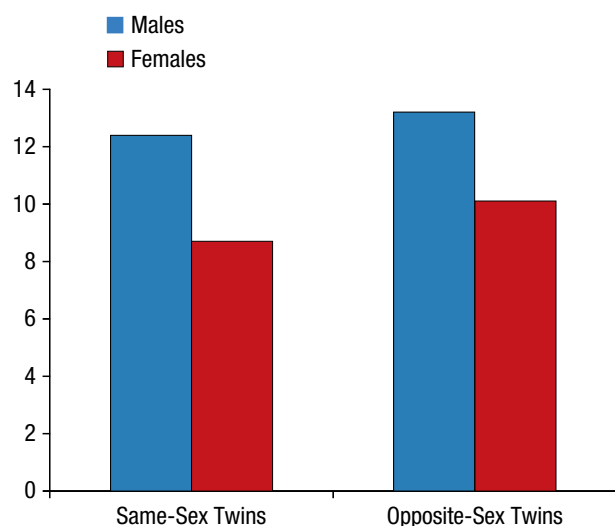


Fig. 6. Number of correct mental rotations on a mental rotation test as a function of gender of co-twin (Vuoksima et al., 2010).

significant, $b = 1.31$, $p = .006$, after controlling for age, birth weight, gestational age, mother's age at the twins' birth, and computer-game experience. (p. 1070)

The p value of the same comparison for males was .11, neither significant nor convincingly absent. These differences are graphed in Figure 6.

Thus, Vuoksima et al. suggested that the females' superior spatial ability was due to their sharing a prenatal environment with males. However, one can imagine substantial postnatal environmental differences in gross motor play (e.g., with block-building) when a girl's sibling is a boy compared with when it is another girl. A female twin whose sibling is a male might also be exposed to similar spatial language that is used by parents primarily with their male preschoolers (e.g., when they are playing with puzzles) and that has been associated with spatial and mathematical skills 6 months later (Levine, Ratliff, Huttenlocher, & Cannon, 2012). However, Miller and Halpern (2014, p. 40) pointed out that girls who had a brother close to them in age had no advantage. Also, note in Figure 6 that males who shared their prenatal setting with other males did not benefit from the increased androgens and even appeared to be slightly worse (albeit insignificantly) than boys who shared their prenatal (and postnatal) environment with females. Obviously, there is no linear relationship between prenatal androgens and later spatial ability, because although androgen levels are highest among two male twins, they do not manifest higher MRT scores. Work with older individuals has often suggested that the optimal level of androgen and, specifically, testosterone is not very high (Valla & Ceci, 2011).

Early spatial differences between boys and girls. As will be seen later, the literature on early sex differences in both mathematical and spatial ability is largely based on studies of mean performance rather than on comparisons between the sexes at the right tail of the distribution, where most GEEMP academics are found. Meta-analytic studies have documented large sex differences in 3-D mental rotation ability favoring males. However, there are only small differences in cognitively-less-demanding 2-D spatial rotation tasks among children and teenagers (Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995), and no systematic sex differences or even a female advantage in other types of spatial ability, such as that needed in spatial-memory and object-location tasks.

Over 100 studies have examined sex differences on these types of tests. The research has uniformly reported large effect sizes favoring male superiority in 3-D spatial processing, with effect sizes often greater than 0.80 SD (e.g., see Linn & Petersen, 1985; Voyer et al., 1995).

Interestingly, some spatial tasks show a male advantage when they are framed as geometry problems but a female advantage when they are framed as an art task (Huguet & Regner, 2009). Various interventions for teaching spatial processing have demonstrated that the gap between the sexes can be narrowed, though not usually fully closed, at least within the confines of the training durations, which typically have been one semester or less (Ceci, Williams, & Barnett, 2009).

Boys' spatial-ability advantage could be dismissed if it were clearly the result of practice playing dynamic video games or building with Lincoln Logs, Erector Sets, Legos, and so on, because if that were the basis, it could be contravened by exposing girls to such activities. Some research has shown that engagement in spatial activities (e.g., shaping clay, drawing and cutting 2-D figures) between the ages of 2 and 4 years predicts mathematical skills at age 4.5 years. In one recent experiment (Grissmer et al., in press), an intervention based on transforming spatial materials by making 2-D copies of 3-D designs and vice versa (using paper, pattern blocks, clay, etc.) elevated disadvantaged children's ranking on the Woodcock-Johnson Applied Problems and KeyMath-3 Numeration tests from around the 32nd to the 48th percentile; the children's visual-spatial ability was likewise elevated as a function of the play activity, from 33% to 47%.

However, sex differences have been observed in infant spatial performance, long before play activities would have an impact. Four studies have shown that young male infants outperform their female counterparts on mental rotation tasks. This outcome depends on methodological features of the experiments,⁸ a point we shall return to later in our discussion of the meaning of early gaps in spatial ability and its role in sex differences in

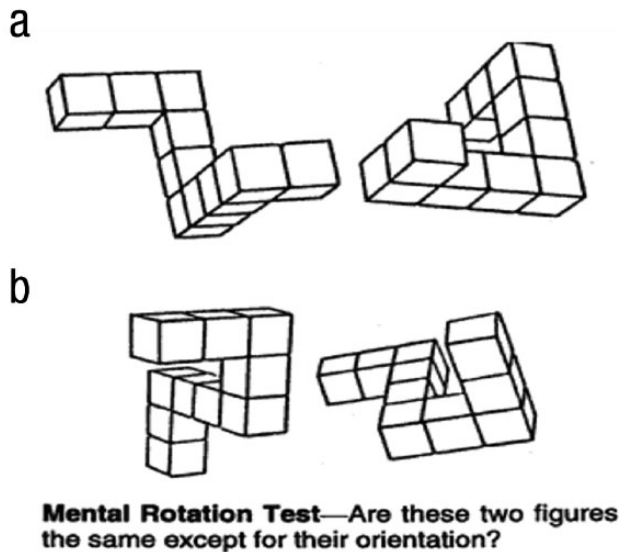


Fig. 7. Example display from Shepard-Meltzer (1970). Adapted from "Mental Rotation of Three-Dimensional Objects," by R. N. Shepard and J. Metzler, 1971, *Science*, 171, p. 702. Copyright 1970 by the American Association for the Advancement of Science. Adapted with permission.

math achievement. First, we describe the evidence for early spatial differences that seems to favor male babies.

Using a simple habituation paradigm with a Shepard and Metzler (1971) spatial-display task (Fig. 7), researchers have shown that infant boys appear to recognize spatial rotations earlier than their female peers. Following habituation to an object, when infants are shown it in a new perspective in alternating turns with its mirror image, 3-month-old boys demonstrate a novelty preference (looking at the novel display instead of the rotated version of the familiar display), but girls of this age do not. Researchers found that female 3-month-olds looked at the familiar and novel objects for similar durations, whereas male 3-month-olds looked significantly longer at the novel, rather than the familiar, object, which implies that they mentally rotated its image (Moore & Johnson, 2011). This suggests that boys' spatial cognition is running in advance of girls'. Quinn and Liben (2008) found a similar result for 3- to 4-month-olds with 2-D rotation.

Recently, Miller and Halpern (2014) reviewed the evidence for these early male advantages and concluded that there were notable inconsistencies:

More dramatically, four studies have found male advantages in mental rotation tasks among infants as young as 3 months of age. However, many other infant studies did not detect these differences when alternate mental rotation tasks were used, including tasks that closely matched those used in prior studies. Similar male advantages in rotation tasks are sometimes detected among preschoolers and

kindergarteners but sometimes not. Causes for these nuanced differences across studies and tasks are currently unclear. Although often interpreted as reflecting innate brain differences, early-emerging sex differences do not necessarily establish biological or environmental causation. For instance, sex differences in high mathematics test performance are reversed (female advantage) among Latino kindergarteners, indicating the early emerging effects of family and culture. (p. 39)

To further complicate the interpretation of the early male advantage in spatial processing as an index of direct biological unfolding, researchers in Germany have shown that 3-D mental rotation is linked to seemingly small differences favoring infants who engage in early crawling and manual manipulation. In Figure 8, Schwarzer, Freitag, and Schum (2013) presented 9-month-olds with a 3-D rotation task. Half of these infants had been crawling for 9 weeks, and some of them were prone to manually exploring objects that had been presented to them. The infants were habituated to a video of an object rotating back and forth through a 240° angle around its longitudinal axis. When tested by being shown the same object rotating through the unseen 120° angle and a mirror image of the display, the crawlers looked significantly longer at the novel (mirror) object than at the familiar object, and this was true regardless of their manual-exploration scores. In contrast, the noncrawling infants' mental rotation was influenced by their manual exploration. These results, along with others reviewed by Miller and Halpern (2014) that are discussed above, suggest that subtle environmental differences, such as early crawling and object manipulation, can influence spatial cognition. Although this does not rule out a biological basis to the male advantage (perhaps early crawling is biologically determined and occurs earlier for male babies), it presents an environmental hypothesis that could be tested with an intervention that induced infant girls to manipulate objects and crawl, even though there were no sex differences in these activities by 9 months of age.

Sex differences in quantitative ability. The literature on early sex differences in both math and spatial ability is largely based on studies of mean performance rather than on work isolating the top tail of the distributions. We will begin with some broad generalizations about adult sex differences in mathematics performance and then work backward to examine their developmental antecedents. Figure 9 relates mathematical-ability differences across STEM fields to female representation in those fields. Despite large increases in the numbers of women earning bachelor's degrees and PhDs in STEM disciplines, there is a strong negative association between

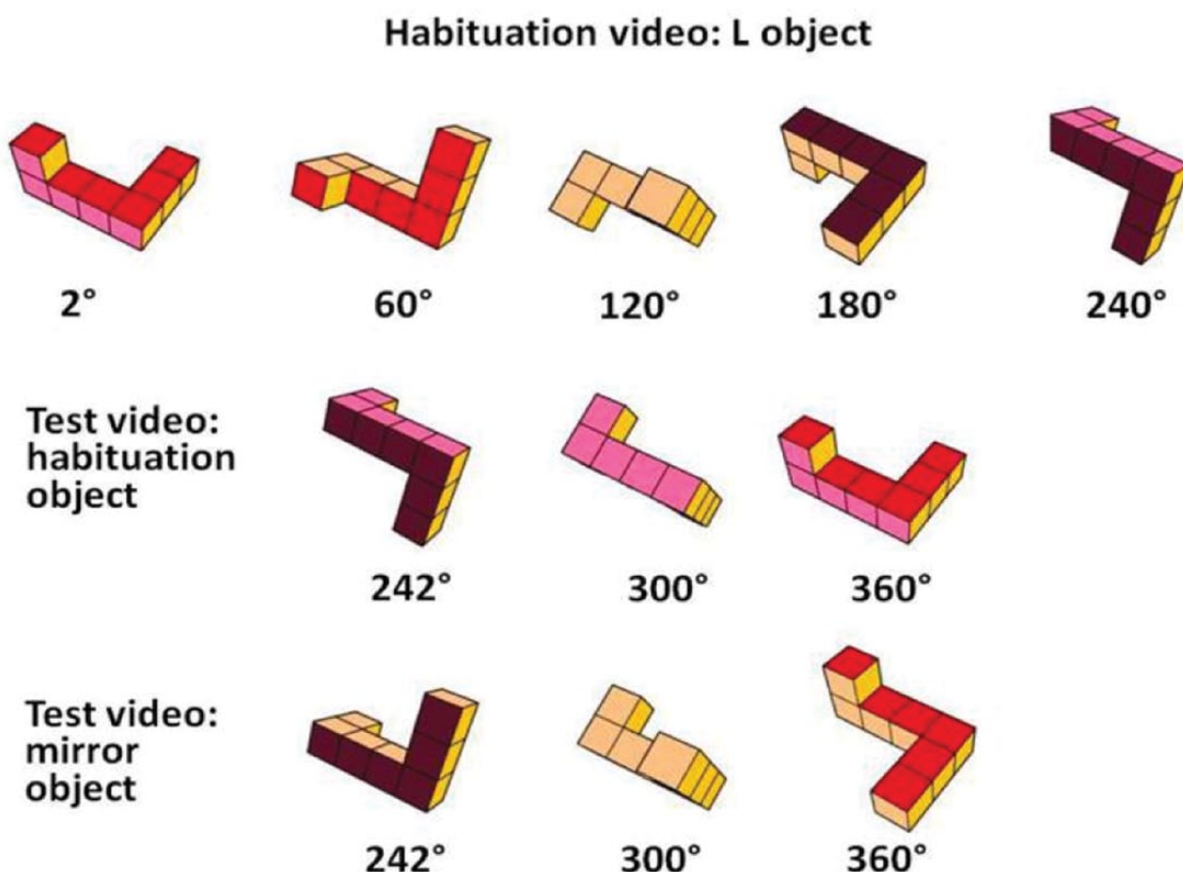


Fig. 8. Examples of the images presented in habituation and test videos in Schwarzer, Freitag, and Schum (2013). The images in the habituation video rotated back and forth through a 240° angle (2° to 240°). The images in the test videos rotated back and forth through a previously unseen 120° angle (242° to 360°). Reprinted from “How Crawling and Manual Object Exploration are Related to the Mental Rotation Abilities of 9-Month-Old Infants,” by G. Schwarzer, C. Freitag, and N. Schum, 2013, *Frontiers in Psychology*, 4. Copyright 2013 by G. Schwarzer, C. Freitag, and N. Schum. Reprinted with permission.

the mathematical content of the PhD—as measured by the average GRE quantitative scores of applicants, with 170 being the top score—and the percentage of women receiving advanced degrees in those fields. Simply put, the more math, the fewer women.

Are these sex differences in PhDs the result of sex differences in mathematical aptitude and spatial ability that are manifested early in life? According to this “gender essentialist” claim (i.e., that the sexes are fundamentally different, probably because of biological differences that interact with the environment), early spatial and quantitative differences cascade into later gaps between the sexes, with females scoring less often in the very top of the math distribution, thus impeding their admission into PhD programs in math-based fields. Much of the evidence for sex differences in math aptitude has been reviewed elsewhere and found wanting as *the* primary causal factor in women’s underrepresentation (e.g., Andreescu, Gallian, Kane, & Mertz, 2008; Ceci & Williams, 2007, 2010a, 2010b; Hyde & Mertz, 2009; Miller & Halpern,

2014). By this we do not mean that spatial and quantitative differences between the sexes are not real or instrumental in the dearth of women applying to GEEMP fields. Rather, as we explain below, the fact that sex differences in aptitude vary by cohort, nation, within-national ethnic groups, and the form of test used means that they are malleable, and there are other causal factors that may be more important in accounting for the lack of women in GEEMP careers. Moreover, mathematics is heterogeneous, comprising many different cognitive skills, some more important in, say, geometry and others more important in algebra or calculus.

None of this means that sex differences in quantitative and spatial ability play no role, but it cautions against assuming that early sex differences translate directly into later sex segregation in career outcomes. For example, the above-documented sex differences favoring men on spatial 3-D tasks do not translate into reliable sex differences in geometry but depend on whether one measures grades or scores on standardized tests that do not cover

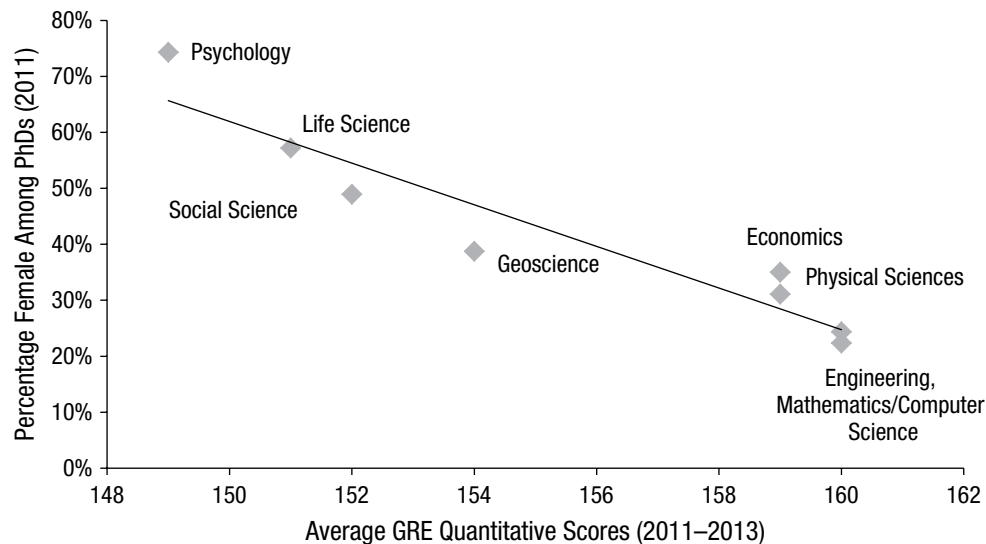


Fig. 9. Percentage of PhDs awarded to females across STEM fields as a function of fields' average GRE Quantitative Reasoning scores. Data shown here are drawn from Educational Testing Service (http://www.ets.org/s/gre/pdf/gre_guide.pdf) and the WebCASPAS database (<https://ncesdata.nsf.gov/webcaspar/>).

materials directly taught in classrooms (Else-Quest, Hyde, & Linn, 2010; Lindberg, Hyde, & Petersen, 2010). For that matter, sex differences in mathematics scores do not translate into grades in math classes, including complex math in college (Ceci et al., 2009), and 40% to 48% of baccalaureates in mathematics have been awarded to women for at least two decades without exception (see Table A1 in the Appendix). Again, none of this means that biological sex differences play no role in the shortage of women in GEEMP fields, but it does mean that care must be taken in linking these data to women's underrepresentation in science, let alone touting sex differences as *the* primary causal factor.

As already noted, much of the literature on sex differences in mathematical and spatial ability comes from studies focusing on the average mathematical ability of the sexes. Janet Hyde and her colleagues have analyzed the sex gap in average mathematics ability a number of times, sometimes using national probability samples involving millions of school-aged children (Hyde, Fennema, & Lamon, 1990; Hyde, Lindberg, Linn, Ellis, & Williams, 2008). Repeatedly, she has found small to non-existent gaps at the center of the math distribution, with boys' and girls' average performance almost entirely overlapping (d s = 0.05–0.26, favoring boys).⁹ Hyde et al.'s (1990) meta-analysis included 100 studies (involving 3 million children) and found no sex differences for children at any age or for any type of problem—from the simplest types of math (fact recall) to more complex types of problem solving—with the exception of complex math problems for high-school-aged students, where there was a small male advantage, $d = 0.29$. In

addition, using national probability samples, Hedges and Nowell (1995) found a comparably sized male advantage among high school students. Most studies have reported no differences in algebra skills, which have a high verbal component that plays to females' strength (Halpern, 2012), but superior performance by boys on 3-D solid geometry (Kimball, 1989).

However, by the beginning of the 21st century, girls had reached parity with boys—including on the hardest problems on the National Assessment of Educational Progress (NAEP) for high school students. As Hyde and Mertz (2009) concluded:

Items from 12th-grade data categorized by NAEP as hard and by the researchers as requiring complex problem solving were analyzed for gender differences; effect sizes were found to average $d = 0.07$, a trivial difference. These findings provide further evidence that the average U.S. girl has now reached parity with the average boy, even in high school, and even for measures requiring complex problem solving. (p. 8802)

This parity was most likely the result of increased mathematics-course-taking by girls (Blair, Gamson, Thorne, & Baker, 2005) that by this time had closed the course gap, which had been sizable through the 1980s.

Sex differences at the right tail. It is one thing to say that the sex gap at the center of the math distribution closed, but what about the gap at the extreme right tail—the top 5%, 1%, or even .01% (1 in 10,000)? After all, if we

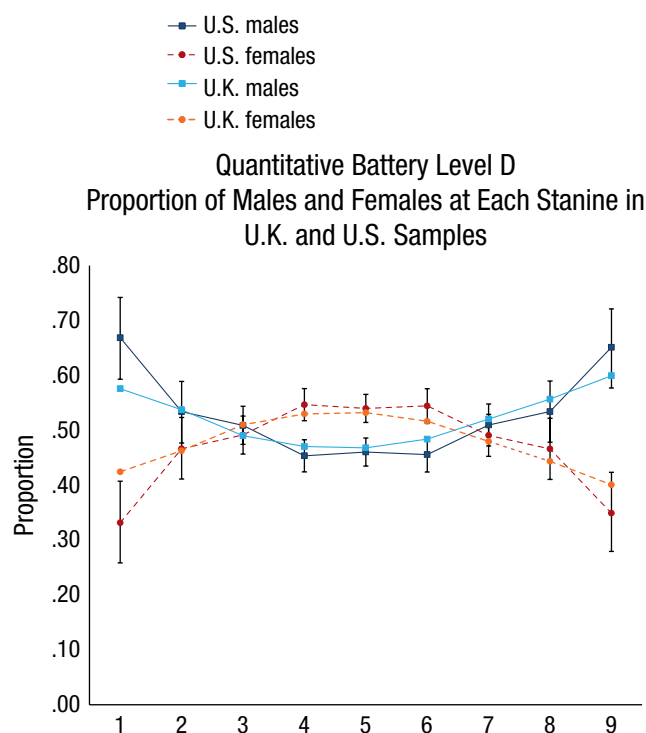


Fig. 10. Male-to-female proportions in the distribution of scores on the Quantitative Battery Level D (mean stanine = 5, $SD = 2$) as a function of stanine and gender. Results shown here are based on data from 318,599 students in the United States and 320,000 10- to 11-year-olds in the United Kingdom. Adapted from "Consistencies in Sex Differences on the Cognitive Abilities Test Across Countries, Grades, Test Forms, and Cohorts," by D. F. Lohman and J. M. Lakin, 2009, *British Journal of Educational Psychology*, 79, p. 399. Copyright 2009 by the British Psychological Society. Adapted with permission.

are concerned about the underrepresentation of women in math-intensive fields in the academy, then we are probably talking not about individuals with average mathematical ability but rather about those with high ability, as the GRE Quantitative Reasoning scores in Figure 9 suggest. What is known about sex differences among children at the right tail? To answer this question, we will augment the earlier reviews with data published more recently, as well as with analyses that we have run in the process of preparing this report.

A number of studies have reported male advantages on standardized math tests as early as kindergarten. In the largest of these studies, Penner and Paret (2008) analyzed a large, nationally representative sample of 5-year-olds entering kindergarten in the 1998-through-1999 cohort and found a small advantage for boys at the right tail on a standardized test. Notwithstanding this demonstration of male superiority at the right tail of young children's standardized-math-test distribution, sex differences in mathematics are not stable until adolescence (Ceci et al., 2009), and they vary according to whether math aptitude is indexed by classroom grades or by

standardized tests that do not directly assess what is taught in classrooms, with the biggest sex differences on the latter. Below is a summary of this literature.

Lohman and Lakin (2009) analyzed 318,599 American 3rd to 11th graders and found that a higher proportion of boys were in the highest stanine (top 4%) of the math distribution of high scorers, and this overrepresentation was relatively constant across national samples from 1984 to 2000. Strand, Deary, and Smith (2006) reported a similar finding of male overrepresentation at the right tail of this same test for a national sample of 320,000 11-year-olds from the United Kingdom. As can be seen in Figure 10, boys are significantly more likely to score in the top 4% (1.75 SD above the mean) as well as in the bottom 4% in quantitative ability, and this pattern has been fairly stable over more than a 16-year period. Hedges and Nowell's (1995) analyses of six national data sets also showed consistency in the sex ratios at the top tail over a 32-year period.

Using a huge data set, Wai and his colleagues (Wai, Cacchio, Putallaz, & Makel, 2010; Wai & Putallaz, 2011) have reported two analyses of sex differences at the right tail, as have Hyde et al. (2008), Ellison and Swanson (2010), and Andreescu et al. (2008). The latter reported large sex differences in the number of students achieving top rankings on the most challenging mathematics competitions, such as the 9-hour Mathematical Olympiad test or the even more challenging William Lowell Putnam Mathematical Competition. However, the magnitude of these differences fluctuates across epochs, countries, and even schools within countries, which indicates that sociocultural factors are driving some of the sex differences at the right tail.

Wai et al.'s (2010) findings are based on a stability analysis of right-tail ratios over 30 years and involved 1.6 million 7th graders (who were all highly able intellectually). They show that the large sex gap reported in the early 1980s (a 13.5:1 ratio of males to females among those 7th graders scoring in the top 0.01% on the SAT-Mathematics) had shrunk by the early 1990s to 3.8:1 and has remained near that ratio since. These data are based on large, nonrandom samples of talent-search adolescents who were administered the SAT at age 13, so the cause of the gap-closing is unknown—for example, perhaps more Asian and Asian-American females took the test toward the end of this period than at the beginning and, as a group, Asians and Asian-Americans excel in mathematics and have a lower male-to-female ratio at the right tail (Halpern et al., 2007), or perhaps the increasing number of math courses taken by females over this period was the primary cause of this closing, or perhaps changes to the composition of the test itself tilted it in favor of females (see Spelke, 2005, for examples of item changes that favor each sex).

However, even in recent representative samples, there have continued to exist large sex differences at the

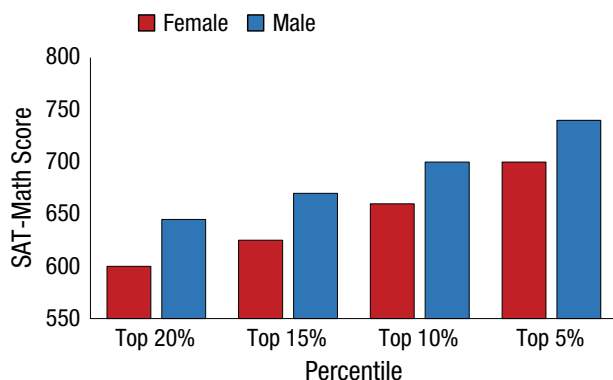


Fig. 11. 2013 SAT-Math scores as a function of sex and percentile. Data shown here are from the College Board (<http://research.collegeboard.org/content/sat-data-tables>).

extreme right tail of math performance. Among nearly two million 6th and 7th graders (equal numbers of males and females) who were administered the SAT-Mathematics or ACT Mathematics tests as part of a talent screening, the male-to-female ratio of those scoring at the top was just over 2-to-1. Hyde et al. (2008) also found a 2.09:1 ratio among the top 1% of math scorers on the NAEP. Mullis, Martin, Fierros, Goldberg, and Stemler (2000) reported small and inconsistent sex differences on Trends in International Mathematics and Science Study assessments at Grade 8, but consistent male superiority by Grade 12, particularly among the highest quartile of mathematics scorers. This underscores the earlier point about the heterogeneity of math, as the context of what is termed “math” (e.g., geometry vs. algebra) changes across grades. Relatedly, Stoet and Geary (2013) reported their analysis of the Programme for International Student Assessment data set for the 33 countries that provided data in all waves from 2000 to 2009. They, too, found large sex differences at the right tail: 1.7:1 to 1.9:1 favoring males at the top 5% and 2.3:1 to 2.7:1 favoring males at the top 1%. Thus, a number of very-large-scale analyses converged on the conclusion that there are sizable sex differences at the right tail of the math distribution.

Figure 1 in Park, Lubinski, and Benbow (2007, p. 950), based on over 2,400 mathematically talented participants assessed on the SAT at age 12 and followed for 25 years (to examine their educational, occupational, and creative outcomes), provides further evidence of the importance of quantitative ability in succeeding in a GEEMP career. It also has an added feature: Of the 18 participants in this group who secured tenure in a STEM field at a top-50 U.S. university, their age-12 mean SAT-Mathematics score was 697 (the lowest scorer in the group was 580). Getting a score of 697 at age 12 is very rare, but that is the point. This study affords a purchase on the level of mathematical reasoning associated with occupying such positions.

It also highlights why the 30-year analysis by Wai et al. (2010) described above is so germane. A second piece in this same series of analyses controlled for educational credentials at elite schools (Park, Lubinski, & Benbow, 2008), and age-12 SAT-Mathematics assessments still revealed impressive validities for predicting creative outcomes such as publishing in refereed STEM outlets and securing patents.

Wai and Putallaz (2011) opined that a sex difference of the observed magnitude at the right tail is probably part of the larger explanation for female underrepresentation, although only part. In their words:

Swiatek, Lupkowski-Shoplik, and O'Donoghue (2000) also examined perfect scores on the [EXPLORE Math Test] and [EXPLORE Science Test] for participants from 1997 to 1999 and found a male-female math ratio of 2.27 to 1 and a science ratio of 1.74 to 1. We replicate these findings using an independent sample from 1995 to 2000 (3.03 to 1 for the math subtest and 1.85 to 1 for the science subtest). We also extend these findings to demonstrate that the male-female math and science ratios for 5th- and 6th-grade students, in addition to 7th-grade students, have been fairly stable for the last 16 or more years . . . These findings provide more evidence in addition to Wai et al. (2010) that male-female math and science reasoning differences are still likely part of the equation explaining the underrepresentation of women in high-level science, technology, engineering, and mathematics (STEM) careers. (p. 450)

Sex differences favoring males in the most recent SAT-Mathematics data for college-bound seniors from the College Board (<http://research.collegeboard.org/content/sat-data-tables>) resemble sex differences in all previous SAT-Mathematics data. As can be seen in Figure 11, the same score that gets a girl into the top 5% of the female mathematics distribution gets a boy into only the top 10% of the male distribution, and the same score that gets a girl into the top 10% gets a boy into only the top 20%.

And yet, even if the population of potentially eligible PhD candidates was tilted 2-to-1 in favor of males, that could not explain gaps of up to 6-to-1 among senior professors in GEEMP fields such as mathematics, computer science, physics, and engineering. As seen in Figure 9, the average quantitative scores of PhD candidates in the most math-intensive fields hover around the 75th percentile, a region where the sex gap is considerably less than 2-to-1. Moreover, a much higher proportion of females than males take the SAT and GRE (2011 GRE test takers were 55% female, 41% male, and 4% unknown), which means that their scores are statistically depressed by more

lower-scoring females, whereas the smaller fraction of males results in higher mean scores (Nie, Golde, & Butler, 2009). Controlling for this would not close the mean gap entirely, but it would narrow it somewhat. Note that this is unlikely to explain the observed sex differences at the *extreme* right tail of the math distribution.

In their analysis of international mathematical competition (International Mathematical Olympiad and William Lowell Putnam Mathematical Competition) awards, Andreescu, Gallian, Kane, and Mertz (2008) found that females comprised many more of the potentially profoundly mathematically elite than they actually comprised in these competitions: These researchers calculated that females constituted 1 in 3 to 8 of all potentially capable candidates, depending on the national conditions, whereas females comprised only roughly 1 in 30 participants in the competitions. In this regard, the economists Ellison and Swanson (2010) also surmised that the vast talent potential of females was being untapped because of national differences in curricula and high school culture that result in failure to identify and foster high-math-ability females, the majority of whom presently come from a very small percentage of American high schools. (If all girls attended these high schools, there would be many more females scoring in the elite category, according to these authors' estimate.)

Finally, several analyses of transnational mathematics data (from Trends in International Mathematics and Science Study assessments and Programme for International Student Assessment surveys) have documented variability in male-to-female ratios and in variance ratios (the male variance divided by the female variance—see, e.g., Else-Quest et al., 2010; Penner, 2008). Else-Quest et al. (2010) cited evidence of variance ratios not significantly different from 1:1.19 in the United States, 1:1.06 in the United Kingdom, 1:0.99 in Denmark, and 1:0.95 in Indonesia. These transnational differences suggest that something more than raw mathematical potential could be driving the variance ratio of males to females. We address what these other drivers may be below, but for now we acknowledge that variance ratios as well as sex differences at the extreme right tail are real (most nationally representative samples indicate greater variance for boys' mathematics scores, on the order of 8% to 45%, with a mean that we estimate to be around 15%), and that there is a 2-to-1 ratio favoring males among the top 1% of math scorers. Halpern (2012) pointed out that variance ratios may underestimate true variance in the population because many more males are developmentally delayed and never make it into these assessments.

Even if they are highly mutable, such sex differences could be instrumental in the lower number of females applying for and/or admitted into and/or achieving at high levels in GEEMP fields, notwithstanding our claim

that this factor cannot account for all or possibly even most of the sex difference—and, obviously, notwithstanding our claim that these differences can and do change, which we briefly describe next.

Having documented the rather pronounced sex differences at the right tail of the mathematics distribution, it is important to add a caveat: There are a number of examples of inconsistency in the gender ratios at the right tail, and these divergences are also based on large national samples and meta-analyses (Becker & Hedges, 1984; Friedman, 1989; Hyde et al., 1990; Hyde & Linn, 2006; Hyde & Mertz, 2009). We have reviewed much of this literature elsewhere (Ceci et al., 2009), pointing out countries in which females are at parity with or even excel over males at the right tail (e.g., in Iceland, Singapore, and Indonesia, more girls than boys scored at the top 1% at certain ages). Hyde and Mertz's (2009) review revealed that the magnitude of the male advantage at the right tail has been decreasing, more in some countries than in others, and the greater male variance in math scores is not ubiquitous, as we have seen above. In Lohman and Lakin's (2009) 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 Non-Verbal Battery over the same time period: 9th-stanine proportions of females to males changed from 0.72 in 1984, to 0.83 in 1992, to 0.87 in 2000. Relatedly, the male-to-female ratio at the top 4% is larger in the United States (2:1) than it is in the United Kingdom (roughly 3:2), which further illustrates the influence of cultural factors on the ratio.

Pope and Sydnor (2010) found wide variations across U.S. states in the male-to-female ratio of NAEP test scores at the 95th percentile, with sex differences in some states less than half the size of sex differences in others. Moreover, states with more gender-equal math and science NAEP scores also have more gender-equal NAEP reading scores at the top tail, although girls have the higher rate in reading whereas boys have the higher rates in math and science. For instance, the ratio of males to females with NAEP scores in the 95th percentile in math/science is approximately 1.4-to-1 in the New England states but 1.8-to-1 in the East South Central states. However, the ratio of females to males with NAEP scores in the 95th percentile in reading is approximately 2.1-to-1 in the New England states and 2.6-to-1 in the East South Central states. Although Pope and Sydnor's work did not identify causal relationships, their results suggest that gender norms strongly influence mathematic achievement at the top tail.

In short, temporal data, ethnic data, and trans-state/transnational data all indicate that the ratio of males to females at the right tail is not carved in stone; these ratios can and do change, and they differ for ethnic groups (as

noted above, Hyde & Mertz, 2009, found sizable gaps favoring white males at the extreme right tail but found the opposite pattern for Asian-Americans, with more females at the right tail) and time periods. Resolving 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 elsewhere (Ceci et al., 2009). 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 a higher proportion of one sex takes a test, the scores of that sex as a whole go down (and there have been increases in the proportion of female students taking some tests, e.g., the SAT; Nie et al., 2009); (c) changes in analytic approaches—for example, the use of extreme-tail-sensitive approaches such as quantile regression results in smaller sex differences at the tail (see Penner, 2008); (d) changes in the type and number of math courses each sex has taken (Hyde et al., 2008); and (e) differences in school culture (Ellison & Swanson, 2010; also see Stumpf & Stanley, 1998, for additional factors that may affect sex differences).

Stereotype threat and competition. In some situations, stereotype threat has been shown to lower females' math performance. Steele and his associates have shown that the awareness that others expect members of a social group to do poorly on math, even when this belief is not endorsed by the group's members, is sufficient to create anxiety and poorer performance among them (e.g., Spencer, Steele, & Quinn, 1999; Steele, 1997). Correll (2001) has shown that males estimate their math ability to be higher than comparable females' ability, and that females not only underestimate their math ability but also overestimate how much math ability is necessary to succeed at higher levels of math (Correll, 2004).

Even subtle priming of sex can reduce females' math performance. For example, female test takers who marked the gender box *after* completing the SAT Advanced Calculus test scored higher than female peers who checked the gender box *before* starting the test, and this seemingly inconsequential order effect has been estimated to result in as many as 4,700 extra females being eligible to start college with advanced credit for calculus had they not been asked to think about their gender before completing the test (Danaher & Crandall, 2008; cf. Stricker & Ward, 2004, 2008, for criticism about data assumptions made by Danaher and Crandall). Good, Aronson, and Harder (2008) conducted a field experiment in an upper-level college calculus course and found that women in a stereotype-nullifying treatment condition

outperformed men. The sensitization to gender prior to starting the test presumably causes females anxiety resulting from doubts about their math ability, and this anxiety reduces working memory and lowers performance (Beilock, Rydell, & McConnell, 2007; Schmader & Johns, 2003). Krendl, Richeson, Kelley, and Hamilton (2008) have shown that when women are in a 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.

Niederle and Vesterlund (2010) have argued that girls' performance on math tests and their willingness to compete in high-stakes testing environments are influenced by the gender differences of the other competitors and test takers. They found that girls performed better in competitions against other girls and worse in competitions in which boys outnumbered girls. Thus, girls' lower average test scores in subjects where boys are the majority of test takers may reflect gender differences in attitudes toward competition. Cotton, McIntyre, and Price (2013) examined gender differences in repeated math competitions for elementary-school-aged children. They found a significant male advantage in math performance in the first round of the math contest. However, in subsequent rounds, girls outperformed boys. In addition, the male advantage dissipated once time pressure was removed. Furthermore, Cotton et al. (2013) found no male advantage in any period when language-arts questions were used. These results may reconcile the finding that girls get better grades (e.g., higher grade point averages, or GPAs, in math) when they are repeatedly examined but perform poorly relative to boys in high-stakes, one-shot tests such as the SAT or AP tests. These results also suggest that girls may shy away from competition in areas where the gender norm is that girls underperform relative to boys; in other words, girls do not compete in the presence of stereotype threat. Although these stereotype-threat studies all measured mean performance, the psychological factor is likely to apply all along the ability distribution.

Bottom line. The literature on gender and math ability is based largely on sex differences in mean performance. Similarities and differences in *average* math aptitude are unlikely to be the major cause of the dearth of women in underrepresented GEEMP fields because there are no consistent sex differences at the midpoint of the math distribution and GEEMP faculty do not hail from the middle of the distribution, as some have claimed (however, see Hill, Corbett, & St. Rose, 2010, p. 21, for the claim that GEEMP professions do not require high math aptitude).

Even though there is a sex gap at the right tail of the math-ability distribution, as Hyde et al. (2008), Lohman and Lakin (2009), Stoet and Geary (2013), and Wai et al.

(2010) have amply documented, the extent of this gap is mutable, changing over time and across ethnicities, states, and nations, and dependent on environments and the salience of stereotypes.

Moreover, there are still many more math-talented females than women receiving PhDs and going on to professorships. Since the early 1990s, females have received 40% to 48% of the bachelor's degrees in mathematics (Andreescu et al., 2008; see Section B in the Appendix), and Daverman (2011) reported that for the 10-year period spanning 2001 through 2010, roughly 30% of the PhDs in mathematics were awarded to U.S.-born women. Thus, mathematical- and spatial-ability differences at the right tail are not inevitable (recall the earlier findings of variability across epochs and nations in the male-to-female math ratios at the right tail—Hyde & Mertz, 2009; Penner, 2008), nor can they explain why females take a number of complex mathematics courses comparable to that of their male peers, and tend to get slightly better grades in them. Finally, even if the 2-to-1 male-to-female ratio among the top 1% of math scorers is part of the explanation for the lower presence of women in math-intensive fields, it is unlikely to be the largest part given the greater degree of underrepresentation among women professors than would be expected on the basis of a 2-to-1 ratio favoring males, and the unevenness of women across the GEEMP fields, with the proportion of women in some fields, such as mathematics, being relatively higher than in others, such as physics and computer science.

Thus, the claim that the early male advantage in 3-D spatial processing cascades into a subsequent male advantage in spatial and mathematical performance is fraught with interpretive snarls. Although some correlations have been reported showing that early spatial performance sometimes predicts later math scores (e.g., Grissmer et al., in press; Levine et al., 2012), there are significant lacunae in the evidence that preclude strong causal conclusions. One missing link concerns whether the spatial ability seen early in males is achieved by females at a slightly older age and, if so, whether by this time have males moved on to a higher level of spatial processing, in an unending spiral of catch-up in which the sexes never meet. Or do the two sexes proceed in tandem once females achieve the earlier male level? Moreover, the predictive correlations between early spatial performance and later mathematical performance are neither large nor consistent, and, as we have shown, there are notable examples of reversed sex differences as a function of type of mathematics, country, social class, historical era, and ethnic group. Finally, it bears noting that any decrement in 3-D female infant spatial processing has not thwarted legions of females from achieving at high levels in mathematics, earning nearly half of bachelor's degrees and 30% of PhDs.

Potential explanation #2: Sex differences in high school STEM interest and attitudes

In addition to math- and spatial-aptitude differences between the sexes, there are sex differences in orientation, identity, and attitudes toward STEM careers that emerge early in childhood and are already pronounced by middle school and high school. Adolescent surveys have revealed that sex segregation within STEM “preferred occupations” is already fairly large (e.g., biomedical professions are favored by females, and engineering professions are favored by males), and this segregation eventually mimics that in the careers men and women enter following graduate training. Perez-Felkner and her colleagues reviewed some of this research (Perez-Felkner, McDonald, Schneider, & Grogan, 2012), showing early emergence of gendered differences in subjective orientations toward mathematics and science. This research revealed that by age 5, girls receive the message that math is for boys; the process of identification with math and science is underway well before high school (although it is worth pointing out that this message is not effective in dissuading the nearly half of college math majors who are female). By middle school, boys are more than twice as likely as girls to expect to work in science or engineering (9.5% vs. 4.1%; Legewie & DiPrete, 2012a).

Surveys report that when boys and girls are asked whether they are interested in becoming engineers or computer scientists, about a quarter of boys indicate that they are whereas fewer than 5% of girls express interest—greatly preferring careers as physicians, veterinarians, teachers, nurses, and so on. For example, one poll of 8- to 17-year-olds showed that 24% of boys were interested in engineering versus only 5% of girls; a survey of 13- to 17-year-olds showed that 74% of boys were interested in computer science versus only 32% of girls (see Hill et al., 2010, p. 38, for citations). Legewie and DiPrete (2012a) argued that these early preferences are not stable until high school and do not predict later gender segregation in college majors, but they nevertheless reveal pronounced early gender stereotyping and career choices. Sex differences in math and science orientation/identification begins to solidify after middle school. Among high school students, we observe differences in interest in STEM courses and in expectations of STEM majors.

The College Board reports the numbers of male and female students choosing to take math and science Advanced Placement (AP) courses (Fig. 12). Although overall, the female AP students outnumber male AP students (55% to 45%), girls take the majority of science AP tests in only two fields—biology (58%) and environmental science (55%). Only 19% of the AP Computer Science

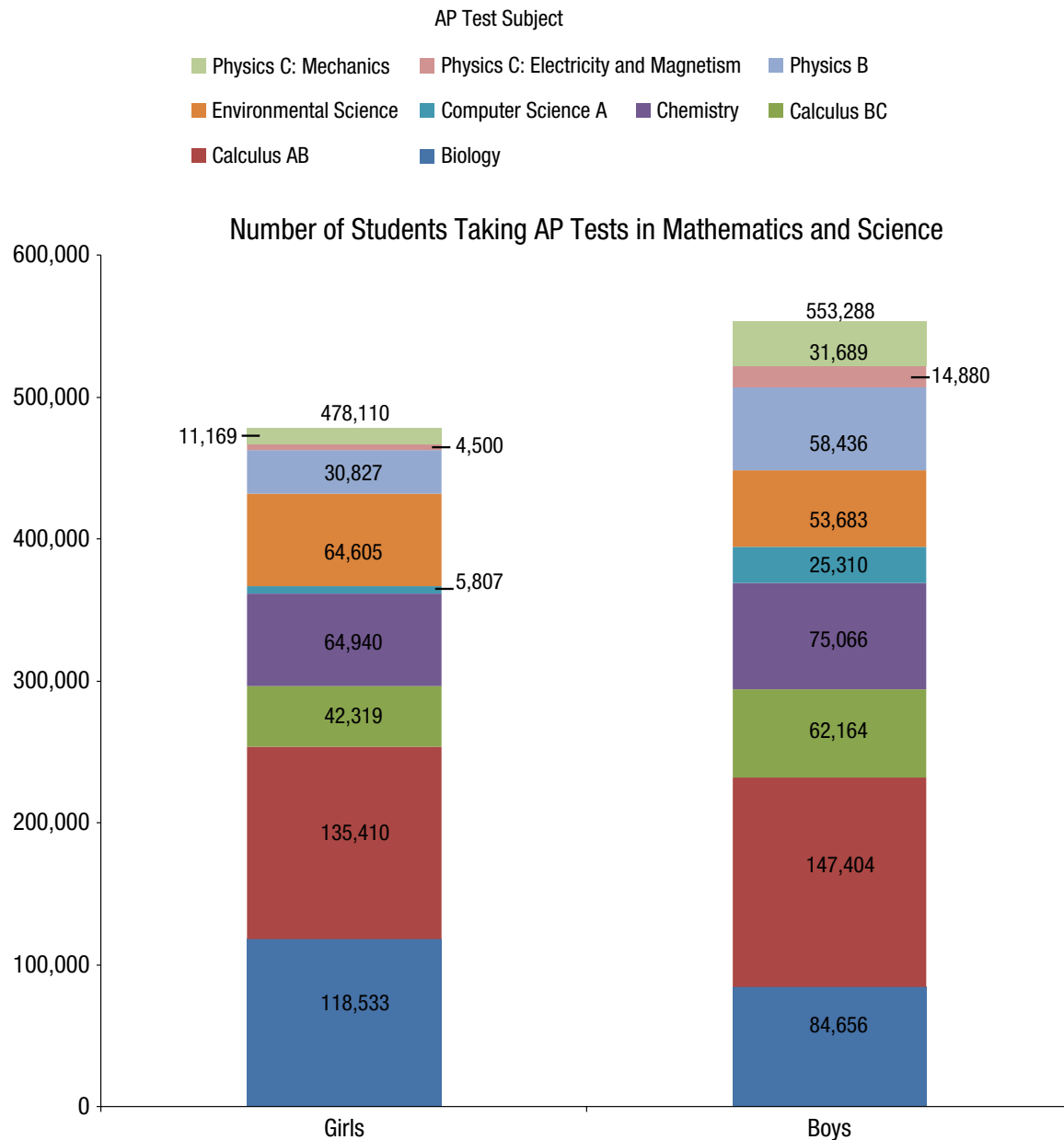


Fig. 12. Number of students taking Advanced Placement (AP) tests in mathematics and science subjects in 2013 as a function of student gender and test subject. Data shown here are from the College Board (<http://media.collegeboard.com/digitalServices/pdf/research/2013/Program-Summary-Report-2013.pdf>).

and 23% of the AP Physics C: Electricity and Magnetism exams are taken by girls.

These large differences are mirrored in high school students' expectations about their college majors. Xie and Shauman (2003) found that among high school seniors expecting to attend college, the percentage of females expecting to major in science and engineering¹⁰ was less than a third that of males. This difference was not explained by family income, mothers' and fathers' educational attainment, parents' expectations for their

children's educational attainment, family computer ownership, the students' expectations of future family roles, or math- and science-course participation. Morgan, Gelbgeiser, and Weeden (2013) reported that sex differences in occupational plans expressed by high school seniors (in 2002) could not be explained by differences in math coursework in high school (where the gender difference in taking advanced calculus courses is much smaller) or future family-formation plans. In their words, their evidence:

shows that male high school students were more than four times as likely as female students to have listed only STEM occupations in their plans, whether the sample includes all students who later enrolled in post-secondary institutions (17.9% vs. 4.3%) or only 4-year college-bound students (20.7% vs. 4.8%). (Morgan et al., 2013, p. 997)

Note that their definition of “planned STEM occupations” here is similar to our GEEMP classification—that is, it excludes medical, biological, health, and clinical sciences. In fact, girls were more likely than boys to plan a biological/health occupation (27% vs. 11%).

The link between ability and high school interest is not completely clear. Returning to the College Board data on AP tests, in every one of these STEM subjects, boys score on average higher than girls, as Figure 13a illustrates. Furthermore, boys are more likely than girls to appear in the right tail of the test distribution. Figure 13b shows the percentage of girls and boys who achieve the top score of 5 on AP exams. In every field, a higher percentage of boys than girls achieved the top score on the exams. For both average (Fig. 13a) and high (Fig. 13b) scoring, the male advantages are true both in those STEM fields in which the majority of test takers are girls (biology and environmental science) and in which the majority of test takers are boys.

Xie and Shauman (2003) studied the association between academic achievement and the high school expectation of becoming a science and engineering major. They found that math and science achievement (whether on standardized tests or in high school math and science grades) at the top levels explained only a small fraction (<4%) of the gender gap in high school students’ expectation to major in STEM fields, and that mean math achievement explained none of it. (Even students’ basic attitudes toward math added no explanatory power beyond this.) However, the data used in their analysis were for high school students from the 1980s. As we observed in Figure 1, the likelihood of girls’ receiving bachelor’s degrees in STEM fields has risen substantially since then. However, using data through 2004, Riegle-Crumb, King, Grodsky, and Muller (2012) also showed that math achievement at the top levels explained only a very small portion of the gender gap in high schoolers’ expectations of a STEM major.

It might not be the absolute sex differences in math ability that affect choice of college majors, but rather sex differences in math ability *in relation to* verbal ability—that is, the relative advantage of one ability over the other—as Lubinski and Benbow (2006) and Wang, Eccles, and Kenney (2013) have shown.¹¹ Riegle-Crumb et al. (2013) investigated sex differences in the comparative advantage in math/science compared with English

reading abilities and found that it explained a bit more of the gap in expected college major. In their 2004 sample, 85% of the gap remained unexplained. Rosenbloom, Ash, Dupont, and Coder (2008) showed that comparative advantage explained women’s lack of participation in computer-science and information-technology careers.

Lubinski, Benbow, Shea, Eftekhari-Sanjani, and Halvorson (2001, p. 311) provided an analysis of quantitative “tilt” that supports this interpretation. Male and female STEM graduate students’ high school SAT-Mathematics minus SAT-Verbal difference is similar and fairly large (92 vs. 79, respectively), and this difference is similar to that seen in high school SAT scores in a separate sample of males (of the same age cohort) identified at age 13 as being in the top 0.5% in math ability (SAT-Mathematics score – SAT-Verbal score = 87). In contrast, the top-math-ability sample of females (identified at age 13) was more intellectually “balanced” (SAT-Mathematics score – SAT-Verbal score = 31).

In related work on relative abilities based on interviews with high school students in 1992 and then again, when they were adults, in 2007, Wang et al. (2013) found that relative abilities do matter: Individuals with high verbal *and* high mathematical ability (of whom the majority were female) were less likely to later be in a STEM occupation than were those with high mathematical ability and moderate verbal ability (of whom the majority were male). However, Wang et al. did not measure how much of the gender gap relative ability explained. These studies suggest that girls are choosing not to enter STEM majors in part because of their high school comparative advantage in verbal subjects relative to STEM subjects; however, this factor is likely responsible for a minority of the gap.

Does early expression of STEM versus non-STEM occupational preference and intended major translate into majoring in STEM versus non-STEM fields in college? A number of studies have examined whether high schoolers’ occupational preferences for STEM fields or their expectations of college majors (expressed in high school or earlier) are predictive of their actual college majors.

Both Morgan et al. (2013) and Xie and Shauman (2003) found that sex differences in occupational plans expressed by high school seniors are a strong predictor of actual gender differences in college STEM majors, even controlling for differences in high school math coursework (which favors boys taking more advanced calculus courses than girls do —21.7% vs. 16.7%) or future family-formation plans. Morgan et al. (2013) found that of those who intended a STEM or doctoral (doctor of medicine or PhD) medical occupation while in high school (and proceeded to a postsecondary institution), 66.5% of males but only 50.0% of females actually declared a major in a STEM field (including biology) or doctoral-track medicine.¹² These

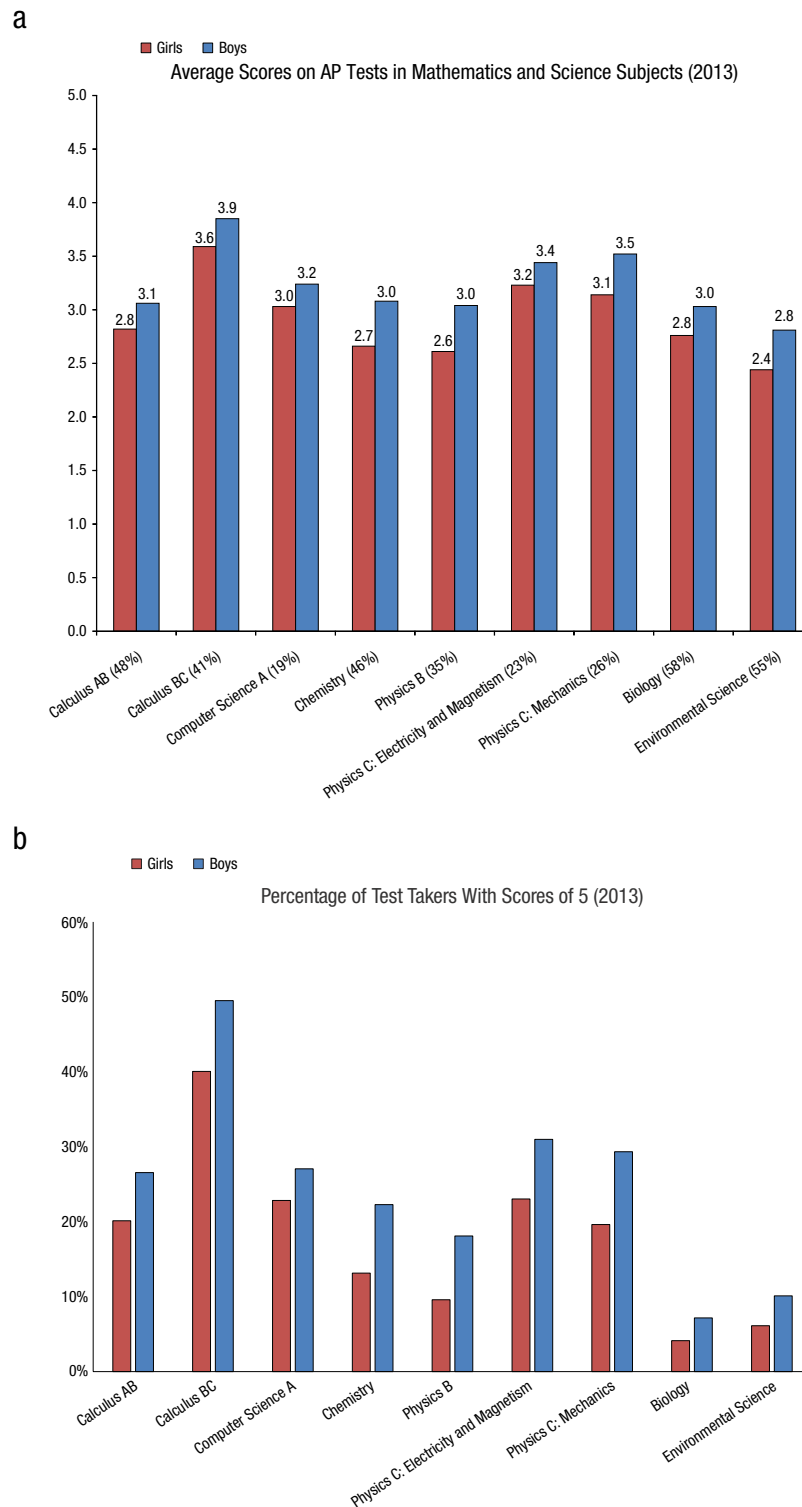


Fig. 13. Average score on Advanced Placement (AP) tests in mathematics and science subjects (a) and percentage of test takers who achieved the top score of 5 (b) in 2013 as a function of sex. Numbers in parentheses are the percentages of female test takers for each test. Data shown here are from the College Board (<http://research.collegeboard.org/programs/ap/data/participation/2013/>).

high school intentions alone explained 23% of the gender gap in actual college majors in these fields.

Interestingly, Xie and Shauman (2003) also found that males who intended in high school to pursue science and engineering majors were more likely to do so than were females who intended in high school to pursue science and engineering majors (28.5% v. 16.0%).¹³ However, Xie and Shauman also reported that female science and engineering majors were more likely to enter these tracks for the first time during college as opposed to beginning college already majoring in science and engineering disciplines; in contrast, for men, the majority of science and engineering majors entered college already expecting to major in these areas. Notwithstanding these perturbations, high school expectations of future college major alone are enough to explain 28.1% of the gender gap in science and engineering baccalaureates!

Finally, Perez-Felkner et al. (2012) examined how subjective orientations to math in high school—as measured with survey questions about perceived math ability, math engagement, valuations of math importance, and beliefs about whether most people can be good at math—affect declarations of GEEMP¹⁴ majors later. They found that gender differences in these subjective orientations—particularly differences in perceived math ability and beliefs about whether most people can be good at math—were important in determining declared GEEMP majors, although their analyses do not allow us to calculate the exact percentage of the gender gap explained by these subjective factors.

In sum, numerous investigators have shown that gender differences in attitudes and expectations about math and science careers and ability become evident by kindergarten and increasingly thereafter. These differences are not stable. To some extent, they are influenced by actual math ability, and they also seem to be heavily influenced by perceived math ability, controlling for actual ability. Stereotypes about the gendered nature of math and the appropriateness of females in this domain are already apparent by high school. Ultimately, if society deems it important to increase the presence of women in the most mathematical fields, it will be necessary to plan pre-high school and high school interventions for increasing math and science identification and advanced coursework. Although women are majoring in mathematics in college in numbers approaching those of men (see Fig. A1a in Appendix), they are not majoring in most of the other GEEMP fields. By the time they become seniors in high school, expectations to major in science are already very different between male and female students. However, this literature has also shown that women who enter college not expecting to major in math can be influenced to do so by their college experiences—female science and engineering majors are more likely to enter the

science and engineering track for the first time during college than to enter college as science and engineering majors, which suggests an important practical implication: that all entering students should be encouraged to take science and math as early as possible.

Potential explanation #3: Sex differences in college majors and in proceeding to a PhD

Recall that in Figure 1, we presented data on college STEM majors. Although women are the majority of bachelor's-degree recipients—reaching a likely plateau of 57% by 2010—they are only 25% of GEEMP majors while being almost 70% of LPS majors. Further, we showed that the percentage of female GEEMP baccalaureate majors has been decreasing since 2002, while continuing to increase in LPS fields. Thus, the single largest bottleneck in the representation of women in math-intensive academic fields is the low number of women majoring in GEEMP disciplines. Above, we discussed the extent to which this reflects high school expectations. Below, we reprise several studies of gender segregation in college majors that have examined international differences in college majors as well as aspects of college education that affect them.

Barone (2011) found across eight European nations (approximately 23,000 college graduates finishing in 1999–2000) sex differences in majors similar to those observed in the United States and Canada. However, he argued that the results indicated that there is not a single gender divide between STEM and humanities fields, but two gender divides, the second one representing a care-versus-technical dimension that cuts across the first. He reported that over 90% of gender segregation in European majors could be accounted for by these two dimensions: the humanities-versus-STEM choices of men and women as well as students' care-versus-technical preferences. This latter distinction echoes the people-thing dimension mentioned earlier, in that it reflects the “cultural opposition between disciplines emphasizing the role of psychological feeling and empathy in understanding and disciplines ruled more by law-governed reasoning” (Barone, 2011, p. 164).

There is some inconsistency regarding sex differences in persistence in college STEM majors. T. R. Stinebrickner and Stinebrickner (2011; R. Stinebrickner & Stinebrickner, 2013) measured students' stated major at the beginning and the end of their college years at a single college, and several times in between. Although far more men (28.1%) than women (16.0%) entered college intending to major in science and engineering—as others have found as well (e.g., Morgan et al., 2013)—more men dropped a science major during college, so that by the end, the ratio of male

to female majors in this particular college had fallen from 1.76:1 to the insignificantly different 1.08:1. Moreover, although their data did not permit them to conduct a detailed analysis of gender differences in science majors, T. R. Stinebrickner and Stinebrickner (2011) found that, compared with female students, male students were overoptimistic and therefore more likely to leave science majors in college because of low actual performance relative to their expected performance. Studying four selective colleges, Strenta, Elliott, Adair, Matier, and Scott (1994) found that science was not homogeneous because, with grades held constant,

gender was not a significant predictor of persistence in engineering and biology; gender added strongly to grades, however, as a factor associated with unusually large losses of women from a category that included the physical sciences and mathematics (p. 513).¹⁵

Grades have been shown to be an important predictor of persistence in a science major by others as well, including some researchers who differentiated the impact by sex and found the impact of grades to be larger for women. In short, research has shown that females attach greater importance to getting high grades than do males and are therefore more likely to drop courses in which their grades may be lower—the so-called “fear of B–.” For instance, Seymour and Hewitt (1997) found that the difficulty with science and engineering coursework and the loss of self-esteem caused by low grades in introductory science and mathematics courses were factors associated with women’s leaving science and engineering majors. Strenta and his colleagues (1994) found that the strongest cognitive predictor of attrition from science majors among those initially interested in science was low grades in science courses during the first 2 years of college, but did not differentiate its impact between women and men. In an investigation at a state university, Jackson, Gardner, and Sullivan (1993) found a similar importance of grades for engineering.

If grades are so crucial in determining majors, whose grades are higher: men’s or women’s? In a representative study, Sonnert and Fox (2012) found that women majoring in biology, the physical sciences, or engineering had cumulative GPAs that were about 0.1 higher than those of men in 2000 (a difference equivalent to approximately 0.3 of a standard deviation, although their measure did not separate out only science courses), and that this female superiority had been increasing over previous years. Other, less representative studies have found contradictory evidence on whose college grades in science are higher. In some, women were found to have lower science grades than men (in the natural sciences: Strenta et al., 1994;¹⁶ in

life science at one public university: Creech & Sweeder, 2012). In other nonrepresentative studies, women were found to have higher grades in a life-science course at one university (Casuso-Holgado et al., 2013); in accounting, math, and statistics courses in one college (C. I. Brooks & Mercincavage, 1991); in introductory physics courses at 16 universities (Tai & Sadler, 2001); and among engineering students at one university (Seymour & Hewitt, 1997). More recent representative studies of college grades in science courses would be useful for understanding this very basic fact regarding comparative grades.

One potential explanation for the dearth of women majors in GEEMP fields may be the lack of role models. Recent data we analyzed from WebCASPAR and the SDR shows that in 2010, STEM fields with more female faculty also produced more female bachelors, with engineering and mathematics/computer science at the low end and psychology at the high end. However, this relationship may not be causal but simply reflect the female participation in GEEMP fields at all life stages.

Using data matching individual students to college instructors, several economists have found that female students are more likely to pursue a major if they have had female faculty (Bettinger & Long, 2005; Canes & Rosen, 1995; Rask & Bailey, 2002) and that females perform better in courses with female faculty (Hoffman & Oreopoulos, 2007). Dee (2005, 2007) found that assignment to a same-gender teacher improves the achievement of both boys and girls, and also improves student engagement. Because these studies were based on observational data, selection (i.e., students’ selection of teachers) may have been responsible for some of these results. However, two recent studies employing random assignment to courses have indicated that having a female faculty member has a causal effect on women majoring in STEM disciplines. Carrell, Page, and West (2010) used the fact that students are randomly assigned to courses at the Air Force Academy and demonstrated that female students who had female professors in introductory STEM courses were more likely to pursue a STEM major than were peers assigned to male professors. (Female instructors had no impact on male STEM majors.) Using a similar identification strategy in a liberal arts college where students were unaware of the gender of the instructor when taking the course, Griffith (in press) found that female students earned higher grades with female faculty members, especially in male-dominated disciplines. However, she found no effect of having a female faculty member on the likelihood that women would pursue majors in male-dominated fields.

Taken together, these recent studies indicate that girls are more likely to pursue STEM majors if they have a female STEM instructor. This underscores the importance of women in academic science careers. However, the

importance of pre-college choices, preferences, and expectations also suggests the importance of female role models in kindergarten through 12th grade as well as in college science courses.

In most cases, the college major is the gatekeeper to pursuing a PhD in that discipline. However, once people graduate from college, are there sex differences in the likelihood of proceeding to a PhD in that field? In Figure 2, we showed that women were less likely than men to transition to the PhD within 7 years of receiving their bachelor's degree in both GEEMP and LPS fields in the 1990s, but that women in GEEMP fields had narrowed the gap significantly in the most recent cohorts. Xie and Killewald (2012) used a much shorter window of time (2–3 years since receipt of bachelor's degree) and found that by 2006, women and men were equally likely to proceed to a higher degree in science or engineering within that time frame (although men had been more likely than women to do so in 2003), but that women were more likely than men to proceed to a professional degree (presumably mostly to medical school). Also, women were less likely to proceed from a master's degree to a PhD in science or engineering, but this was primarily because men were more likely to be in the physical sciences, which had by far the highest probability of proceeding from a master's degree to a PhD (44%). Elsewhere we have noted that women are more likely to have career interruptions in general (W. M. Williams & Ceci, 2012), and in the section of this article dealing with sex differences in productivity, we document that more women PhDs are out of the STEM workforce than men.

To add to these findings, we have analyzed the American Community Survey (2012; <https://www.census.gov/acs/www/>) to investigate educational degree attainment of women and men who majored in science by ages 30 to 35. Considerably more women than men had master's degrees (27.0% vs. 21.8%) and professional degrees (9.1% vs. 7.9%), whereas almost the same percentage had PhDs (5.9% vs. 6.0%), and fewer had halted their education after attaining their bachelor's degree (38.2% vs. 45.5%).¹⁷ Although far more women than men (among the 30- to 35-year-old college STEM majors) were not in the labor force, labor-force participation of women was still high (84.3% vs. 95.9% for men). Of those who *were* employed, men were more likely to be called “scientists” or “engineers” (18.3% vs. 37.4% for women and men, respectively). However, if we include health practitioners and educators as probably being involved in STEM fields in some way, the proportions of men and women involved in science in their careers are much more equal (45.2% vs. 51.2%). Again, we see women choosing people-related occupations (e.g., as health professionals, teachers) rather than thing-related occupations, even within STEM fields.

Studies of sex differences in PhD completion are hampered by a lack of data. One notable exception was a study by Nettles and Millett (2006), who followed over 9,000 students in the 1990s through their graduate careers. They found no sex differences in the completion of and time to degree for the doctorate. However, they found that male PhD students rated their interactions with faculty more highly than did female PhD students. In the next sections, we discuss what happens to those women in GEEMP and LPS fields who obtain PhDs in terms of subsequent career outcomes.

Potential explanation #4: Sex-based biases in interviewing and hiring

Figure 4 showed that in LPS fields (although not GEEMP fields), women are less likely to become assistant professors than the numbers of doctorates awarded to them might lead one to expect. We now consider whether biases in interviewing and hiring explain these gaps.

This section contrasts two forms of contradictory evidence that can be used to argue for or against sex bias in the hiring and promotion of women. On the one hand, numerous small-scale experiments have been reported that strongly suggest that interviewers and evaluators are biased against hypothetical female applicants and their work products, some of which we have briefly reviewed above. On the other hand, actual hiring data across GEEMP fields and the largest experimental test of sex bias in hiring for tenure-track positions in two GEEMP fields (economics and engineering) and two LPS fields (psychology and biology) strongly suggest that the playing field is level as far as interviewing and hiring. We review this evidence below and attempt to reconcile this inconsistency.

To set the stage for its presentation and to justify the length of this section, we begin by citing numerous national blue-ribbon panels, society white papers, and gender-equity reports that continue to allege biased hiring as an important source (if not the most important source) of women's underrepresentation in the academy. Some are at the level of anecdote, and others draw on systematic research. In a *New York Times Magazine* article, Pollack (2013) argued that the underrepresentation of women in math-intensive fields is due—at least in part—to men's underestimation of women's competence and that this is why women are not hired for tenure-track jobs. This essay was replete with anecdotal reports—for example, it quoted a male mathematics professor at Yale's explanation for the shortage of female math professors there: “I guess I just haven't seen that many women whose work I'm excited about” (p. 5).

Consider the following comments that cite research on implicit bias and generalize these findings as a causal

explanation for the underrepresentation of women in academic science:

Women are obtaining doctoral degrees at record rates, but their representation in the ranks of tenured faculty remains below expectations, particularly at research universities. . . . Colleges and universities are not taking advantage of the widest talent pool when they discriminate on the basis of gender in hiring or promoting faculty. . . . At universities that award doctorates, women have filled graduate programs as indicated above, but have not been welcomed into the faculty ranks at comparable rates. (West & Curtiss, 2006, pp. 4–7)

An impressive body of controlled experimental [research] . . . shows that, on the average, people are less likely to hire a woman than a man with identical qualifications [and] are less likely to ascribe credit to a woman than to a man for identical accomplishments. (Institute of Medicine, National Academy of Sciences, and National Academy of Engineering, 2007, p. S2)

Research has pointed to [sex] bias in peer review and hiring . . . The systematic underrating of female applicants could help explain the lower success rate of female scientists in achieving high academic ranks. (Hill et al., 2010, p. 24)

These experimental findings suggest that, contrary to some assertions, gender discrimination in science is not a myth. Specifically, when presented with identical applicants who differed only by their gender, science faculty members evaluated the male student as superior, were more likely to hire him, paid him more money, and offered him more career mentoring. (Moss-Racusin, 2012, para. 8)

In the past, fewer women worked outside the home and as that gradually shifted, there was hiring bias, which means historically women have had fewer science citations than men. That's simple numbers, just like fewer handicapped people and conservatives get citations in modern academia. But is that bias? (Science 2.0; retrieved from http://www.science20.com/news_articles/are_journal_citations_biased_against_women-126192)

One possible explanation for limited progress [in pace of faculty diversification] is that gender and racial or ethnic biases persist throughout academia . . . Evidence suggests that academic scientists express “implicit” biases, which reflect widespread

cultural stereotypes emphasizing white men's scientific competence . . . implicit biases are automatically activated and frequently operate outside of conscious awareness. Although likely unintentional, implicit biases undermine skilled female and minority scientists, prevent full access to talent, and distort the meritocratic nature of academic science . . . Without a scientific approach to diversity interventions, we are likely perpetuating the existing system, which fails to uphold meritocratic values by allowing persistent biases to influence evaluation, advancement, and mentoring of scientists. (Moss-Racusin et al., 2014, p. 616)

Social psychological research repeatedly demonstrates that institutionalized gender bias hinders women's progress in academic science (including medicine). In a recent experiment, for example, men and women science faculty evaluated a job application from a woman less favorably than the identical application from a man. (Connor et al., 2014, p. 1200)

One might imagine that, given the plethora of allegations, there would be compelling evidence that biased interviewing and hiring is a cause of women's underrepresentation in STEM fields and/or that discriminatory remuneration and promotion practices are responsible for the gender gap in pay and rank. However, the evidence in support of biased hiring *as a cause of underrepresentation* is not well supported, and even points in the opposite direction, as we show below. (In the following section, we deal with sex differences in productivity, promotion, and remuneration.) We do not claim that there have not been many excellent demonstrations of implicit bias or stereotyping and explicit bias; rather, our claim is that the literature has failed to demonstrate a *causal link* between such demonstrations and the underrepresentation of female faculty.

First, we review three large-scale analyses of actual tenure-track interviewing and hiring in the United States, which present a consistent picture of gender fairness or even of female preference. That is, female applicants for tenure-track positions are invited to interview and offered jobs at rates higher than their fraction of the applicant pool—the opposite of the bias claim. Following this, we delve into the experimental evidence for gender bias in hiring.

As one of the three examples, a National Research Council (2010) national survey of six math-intensive disciplines examined faculty experiences and institutional policies in place from 1995 through 2003. It included over 1,800 faculty members' experiences, as well as policies in almost 500 departments at 89 R1 universities (note

Table 1. Percentage of Female Applicants for Tenure-Track Positions Invited to Interview and Offered Positions at 89 U.S. Research Universities

Field	Mean percentage of female applicants	Mean percentage of women invited to interview	Mean percentage of women offered position
Physics	12%	19%	20%
Biology	26%	28%	34%
Chemistry	18%	25%	29%
Civil engineering	16%	30%	32%
Electrical engineering	11%	19%	32%
Mathematics	20%	28%	32%

Note: Data shown here were drawn from Sections 3-10 and 3-13 of "Gender Differences at Critical Transitions in the Careers of Science, Engineering and Mathematics Faculty" (National Research Council, 2010).

that these places not only have fewer women faculty than do teaching-intensive institutions but are among the best-paying, most prestigious institutions, so they are a good place at which to examine biased hiring). Although a smaller proportion of female than male PhDs applied for 545 assistant-professor tenure-track positions at these 89 universities, those who did apply were invited to interview and offered positions more often than would be predicted by their fraction of the applicant pool. For example, in the field of mathematics, out of the 96 hires at the assistant-professor tenure-track level from 1995 through 2003, only 20% of applicants for these positions were female, but 28% of those invited to interview were female, as were 32% of those offered tenure-track positions. As another example, out of the 124 hires made in physics at the assistant-professor level, only 13% of the applicants were females (which is much less than the percentage of females among PhDs during that period), but 19% of those invited to interview were women, as were 20% of those actually hired.

As seen in Table 1, similar findings were found in all six STEM disciplines shown (five of which are GEEMP fields)—that is, female applicants were invited to interview and offered positions at higher rates than men. In the words of the National Research Council panel, "in every instance, the mean percentage of female interviews exceeds the mean percentage of applications from women . . . results are similar if we compare median percentages (rather than mean percentages)" (p. 46). Not shown in this table is the finding that a comparable overrepresentation of women were also offered posts for 96 more senior (tenured) positions.

The other two large-scale analyses (Glass & Minotte, 2010; Wolfinger, Mason, & Goulden, 2008) accord with these findings and we do not describe them in detail, other than to note that Wolfinger et al. analyzed over 30,000 respondents interviewed between 1981 and 1995 and found that although women were less likely to obtain tenure-track positions, controlling for such variables as differences in family formation and the

presence of young children revealed that women during this epoch were also hired at rates comparable to or better than men's. For example, in the most common demographic group—unmarried without children—females were 16% more likely to get tenure-track jobs than were males. And Glass and Minotte's analysis of hiring at one large state university over a 6-year period found that of 3,245 applicants for 63 tenure-track positions in 19 STEM fields, 2.03% of male applicants were hired compared with 4.28% of females. It is noteworthy that no counterevidence exists in actual hiring studies—simply put, no real-world hiring data show a bias against hiring women.

Are female applicants superior to male applicants? Some have attempted to explain the preference for interviewing and hiring female applicants for both tenure-track and tenured positions summarized above by arguing that the female applicants are, on average, superior to their male competitors. Thus, according to this argument, the data on actual, real-world female hiring advantages do not rule out biases against female applicants because as a group those who apply for tenure-track jobs are superior to their male counterparts. Consider:

This finding suggests that once tenure-track females apply to a position, departments are on average inviting more females to interview than would be expected if gender were not a factor, or females who apply to tenure-track or tenured positions in research-intensive (R1) institutions are, on average, well qualified. It is important to note that these higher rates of success do not imply favoritism, but may be explained by the possibility that only the strongest women candidates applied for R1 positions. This self-selection by female candidates would be consistent with the lower rates of application by women to these positions. (National Research Council, 2010, p. 49)

In the blogosphere, it is frequently suggested that female applicants are of higher quality than males by virtue of having survived a biased-pipeline process that weeded out many more women.¹⁸ Thus, the argument is that if the pool of female applicants is of higher quality than the male pool, then the high proportion of female PhDs hired may mask bias that prevented an even higher proportion from being hired.

Later, we review evidence and an argument that run counter to this claim, showing that, when taken together, objective measures of productivity (publications, grant dollars, citations per article) do not indicate that women in the applicant pool are stronger than men—publication measures favor men, as do total citations to their work; grant success is similar for both sexes; and citations per article tilt in favor of women—but on the whole, there is no evidence for the superiority of either gender applying for tenure-track jobs.

First, however, we review the key experimental findings on hiring decisions, culminating in a description of the largest and best-sampled study of this genre. Those who make the claim that hiring is not “gender-blind” but instead is biased against women, who—being of superior quality—would have otherwise been hired bolster this claim by relying on experimental evidence regarding hypothetical female applicants’ curricula vitae (CVs), as opposed to actual real-world hiring data, some of which we summarized in Table 1. Several of these experiments have supported the claim of sex biases in evaluations of hypothetical women and their work products (e.g., lectures, papers, hiring) in academic settings, sometimes with moderate to large effect sizes (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012; Reuben, Sapienza, & Zingales, 2014).

In these experiments, raters—either faculty or undergraduates—have been asked to judge the quality of hypothetical applicants’ CVs or work products. Several experiments have revealed that both female and male raters downgrade hypothetical job applicants who are female (Foschi, Lai, & Sigerson, 1994; Moss-Racusin et al., 2012; Reuben et al., 2014; Steinpreis, Anders, & Ritzke, 1999). For example, when 127 science faculty at six U.S. universities evaluated hypothetical applicants with bachelor’s degrees for a staff-lab-manager post, they rated males higher and recommended higher starting salaries and more mentoring for them than for female applicants, even though there was no difference between their applications (Moss-Racusin et al., 2012). Yet with a single exception, these experiments have dealt with biases against female undergraduates—for instance, applicants for lab-manager posts, computational tasks, or summer jobs (Foschi et al., 1994; Moss-Racusin et al., 2012; Reuben et al., 2014; Swim, Borgida, Maruyama, & Myers, 1989)—have involved undergraduates rating the work

products or teaching effectiveness of lecturers (e.g., Bug, 2010), or have involved bias against female applicants for non-professorial jobs (Heilman, Martell, & Simon, 1988). Although these findings are extremely important, it is unclear whether they generalize to the hiring of tenure-track professors in STEM fields (particularly in life science—the field in which the transition from PhD to assistant professor is most difficult).

One observational analysis and one experiment revealed a similar bias in the case of academic hiring of postdocs and tenure-track professors. The experiment examined hypothetical applicants for faculty positions in psychology and found that they were rated as being more hireable if they were male, even though the CVs of males and females were identical (Steinpreis et al., 1999). Again, both female and male faculty raters exhibited this bias. Numerous other experiments have shown bias against females’ teaching skills and work products (e.g., Bug, 2010; Foschi et al., 1994; cf. Swim et al., 1989), but the Steinpreis et al. study is the only study to show bias against women in tenure-track hiring. (Also buried amid these findings are examples of females being overvalued in gender-discordant fields—e.g., Heilman et al., 1988.) In an observational analysis, Sheltzer and Smith (2014) reported that high-achieving male faculty members train fewer women (postdocs and graduate students) in their laboratories than are trained by either elite female investigators or less elite male investigators. Because this study was not an experiment, it is impossible to know the basis of the observed sex differences.

A recent large-scale national tenure-track-hiring experiment was specifically designed to address the question of whether the dearth of women in math-intensive fields is the result of sex bias in the hiring of assistant professors in these fields. This study sampled faculty from 347 universities and colleges to examine bias in the hiring of tenure-track assistant professors in various STEM fields (W. M. Williams & Ceci, 2014¹⁹).

This finding is consistent with the other evidence on productivity presented below, which also fails to show female superiority in hiring outcomes as being due to objectively higher female quality. These experimental findings are compatible with the hiring data showing gender neutrality or even a female preference in actual hiring. There are a variety of methodological and sampling factors that may explain the seeming divergence between earlier experiments and the Williams and Ceci experiment. Notably, in this experiment, candidates for tenure-track positions were depicted as excellent, as short-listed candidates almost always are in real-life academic hiring.²⁰ In contrast, many of the most prominent experimental studies have depicted candidates as “ambiguous” with respect to academic credentials. For instance, Moss-Racusin et al. (2012) described candidates for a

lab-manager position, which are a level of applicants very different from those who are finalists for a tenure-track position, as having ambiguous academic records (i.e., in addition to having a publication with their advisor, they had unremarkable GPAs and had withdrawn from a core course).

Bias may exist in ambiguous cases because of what economists call “statistical discrimination,” which occurs when evaluators assign a group’s average characteristics to individual members of the group. For example, women publish fewer papers than men. Thus, when evaluating a potential female hire, evaluators may assume that as a woman, the candidate will be less productive, based on the group averages. However, this is no guarantee that bias exists in cases in which candidates are clearly competent, such as in the competition among short-listed candidates for tenure-track posts.

Moreover, for implicit bias to be driving the underrepresentation of women in LPS tenure-track hiring, we should observe similar patterns of women being underrepresented in *all* fields and in *all* high-skilled careers. Yet this is not what we observe in the data. Within GEEMP fields, we do not see this hiring bias (see Table 5). Moreover, women are increasing their numbers and success in LPS careers (e.g., as doctors of medicine and veterinarians) more than they are in tenure-track academic science in LPS fields. Goldin and Katz (2012) recently argued that “pharmacist” is the most egalitarian and family-friendly occupation. Thus, the only way bias would fully explain the hiring differential would be if academics in LPS fields were relatively more biased than both academics in GEEMP and nonacademics in life science.

Furthermore, academic hiring decisions are typically made by committees or entire departments, not by individuals. Thus, the hiring process may mitigate the effects of implicit bias, given what is known about the reluctance to express prejudice in public: People are aware of the social norms against expressing prejudice in public situations, so they suppress their private bias, or express it in more subtle ways (e.g., Crosby, Bromley, & Saxe, 1980; Plant & Devine, 1998). Thus, if anything, the results of this experiment are likely to be a conservative estimate of the female preference that would have been observed in a group context.

In light of this recent experiment finding no implicit bias, we would need far more field-specific evidence of biased tenure-track hiring to believe that this is true.

Of course, academic careers may begin with being hired onto the tenure track, but they do not end there. As was clear in our introductory presentation of data, there are many gender differences in sciences at later stages of academic life as well. The next section discusses whether productivity differences exist that explain some of these differences.

Potential explanation #5: Sex differences in productivity

Early in this review, we showed evidence regarding the low—although increasing—percentage of females among tenured and tenure-track faculty. We have begun to track where these differences arise through the point of being hired as a tenure-track assistant professor, and in later sections, we will follow women’s careers as they do or do not get promoted to higher ranks. Economists tend to point to differences in productivity as the primary underlying explanation for sex differences in employment outcomes. In this section, we examine whether there are gender differences in the number of publications, the productivity factor that is most likely to affect academics’ hiring, salary, and promotion. Indeed, Long (1992) found that number of publications increases the academic promotion (to full professorship) of women considerably *more* than it does that of men. (Another important dimension of productivity is the impact or quality of publications, as measured by citations. Later, we will address citations and citations per publication.)

Table 2 summarizes the publication differences identified by some of the many studies on gender differences in publications in STEM fields. In this table, we review the evidence from the articles that are the most comprehensive, covering more recent periods, as well as some of the more highly cited studies covering earlier years. We exclude studies based on a sample of *publications* rather than of people, because these do not allow us to calculate productivity per person.²¹ We also exclude studies limited to specific fields (e.g., Helmreich et al., 1980, on psychologists; Symonds, Gemmell, Braisher, Gorringer, & Elgar, 2006, on life scientists; and Keith, Layne, Babchuk, & Johnson, 2002, on sociologists) or other countries (e.g., Borrego et al., 2010, on Spain; Prpić, 2002, on Croatia; and Symonds et al., 2006, on the United Kingdom and Australia). As we will see in Figure 14, interfield differences are substantial enough to make it impossible to compare across fields to identify time trends. The same is likely to be true across countries. Finally, we exclude studies that are not limited to STEM fields,²² with one exception included because it measured time trends.

Also included in Table 2 are gender differences in publications from analyses that we have done for this report based on the most recently available publication data in the NSF’s 2008 SDR.²³ Specifically, the SDR measured articles accepted for a refereed professional journal in the previous 5 years. We start with this metric in order to describe the current state of the “gender publication differences.”

The overall average difference in the 5-year publication count for male and female academics in the 2008

Table 2. Gender Differences in Publications

Article	Data years	Field	Specific population	Female disadvantage (female publications as percentage of male publications)	Number of observations	Measure of publications	Controls and other notes
Present article	2003–2008	STEM	Assistant professors Associate professors Full professors	27%, 2.1 articles 11%, 0.86 articles	6,694	Mean	Nationally representative sample; 5-year panel
	1990–1995	STEM	Assistant professors Associate professors Full professors	22%, 2.8 articles 32%, 2.1 articles 21%, 1.54 articles	6,694	Mean	
Xie and Shauman (2003)	1969 1973 1988 1993	STEM	Academics	19%, 1.8 articles 42.2%, 1.4 articles 57.9%, 1.3 articles 30.6%, 1.5 articles 18.3%, 1.0 articles	12,167 6,998 847 1,845	Mean	2-year period; nationally representative sample; no controls
	1969 1973 1988 1993	STEM	Academics	4.80% 6.40% 22.50% 6.90%	12,167 6,998 847 1,845	Mean	2-year period; nationally representative sample; controls for field, academic rank, type of institution, teaching, and residential funding
Fox (2005)	1990–1994	Chemistry, comp science, microbiology, physics, electrical engineering	Full-time tenured or tenure-track faculty at PhD-granting universities	22%	769	Mean	3-year period
Levin and Stephan (1998)	1973–1981	Physics, earth science, biochemistry, physiology	Full-time tenured or tenure-track faculty at PhD-granting universities	M = 17.1% (range = –5%–51%)	2,209	Mean	Did not use self-reported publications
Sax, Hagedorn, Arredondo, and Dicrisi (2002)	1972–1973 1989–1990 1998–1999	All fields (not just STEM)	Academics	20.6% more females with zero publications 15.2% more females with zero publications 10.5% more females with zero publications	8,544	Zero publications	2-year period; nationally representative sample; no controls
Cole and Zuckerman (1984)	1972–1973 1989–1990 1998–1999 1968–1979	STEM cohort	All PhDs (not just academics)	14.6% (three or more publications) 13.5% (three or more publications) 11.0% (three or more publications) 27%, 2.1 articles (5 years post-PhD) 58%, 4.4 articles (5 years post-PhD) 43%, 4.8 articles (11 years post-PhD) 49%, 2.7 articles (11 years post-PhD)	526	Three or more publications Mean Median Mean Median	Cumulative publications; single cohort
Long (1992)	Through 1980	1956–1963 biochemistry cohorts	All PhDs (not just academics)	21% (1 year post-PhD) 40% (4 years post-PhD) 48% (9 years post-PhD) 37% (17 years post-PhD)	786	Mean Median Mean Median	Calculated by years since PhD; smoothed annual rate; did not use self-reported publications

Note: STEM = science, technology, engineering, and math.

SDR is 2.1 articles, which represents 19.6% of the average number of male publications. In the 1995 SDR, the average 5-year publication gender gap was 22.5%, just 2.9 percentage points higher than the gap in 2008. This suggests only a modest improvement over these 13 years. However, these male advantages could be highly misleading, because on average male academics are more senior to female ones, and annual publications tend to change over the course of one's career.²⁴ We therefore differentiate by academic rank in Figures 14a, 14b, and 14c, which show gender differences in academic productivity by STEM field from the NSF's 1995 and 2008 SDR at the ranks of assistant professor, associate professor, and full professor, respectively.

Assistant professors represent the new generation of women and men entering the field. The average difference in publications for assistant professors is 2.1 articles, representing 27% of male assistant-professor publications. Figure 14a shows these differences by field. In each field in both time periods, point estimates indicate that men published more on average than women. The largest statistically significant productivity gaps for assistant professors in 1995 were in the fields of engineering, life science, mathematics/computer science, and the physical sciences. In the fields of engineering and mathematics/computer science, these gaps have fallen and were no longer statistically significant in 2008. In life science, the gap has narrowed, but it was still significant in 2008, and in the physical sciences, the gap has grown. In almost every field, assistant professors of both sexes are publishing more articles in 2008 than they were in 1995. The exceptions are notable—men in life science are publishing slightly fewer articles in 2008 compared with men in 1995 (although this result is not statistically significant), whereas women are publishing more, but not enough more to close the significant gender publication gap. In economics, men were publishing more and women were publishing less in 2008 than in 1995, leading to a newly significant 2008 publication gap. Although both women and men were publishing more in psychology in 2008, men's productivity increased relatively more, giving rise to a newly significant 2008 publication gap as well.

For associate professors, the average 2008 publication gap was much smaller: 10.5%, or 0.9 articles. In Figure 14b, we see that in 1995, although males published more than females in all fields, these differences were small and were significant only in mathematics/computer science and social science. By 2008, women, on average, were publishing more in six of the eight fields (economics, engineering, geoscience, mathematics/computer science, psychology, and social science), although none of these gender differences were statistically significant. The only field-specific publication advantage that was statistically significant in 2008 was the male advantage in the physical sciences.

The average 2008 gender publication gap was much larger for full professors: 21.6%, or 2.8 articles. Figure 14c indicates that several fields had substantial gaps in 2008, including the physical sciences, psychology, life science, geoscience, and economics. However, given the small samples of women in some of these fields, the only gender gaps for full professors that are significant are for the physical sciences and psychology. For five of the eight fields, these gaps were larger in 2008 than in 1995, when women actually had a publication advantage in three fields.

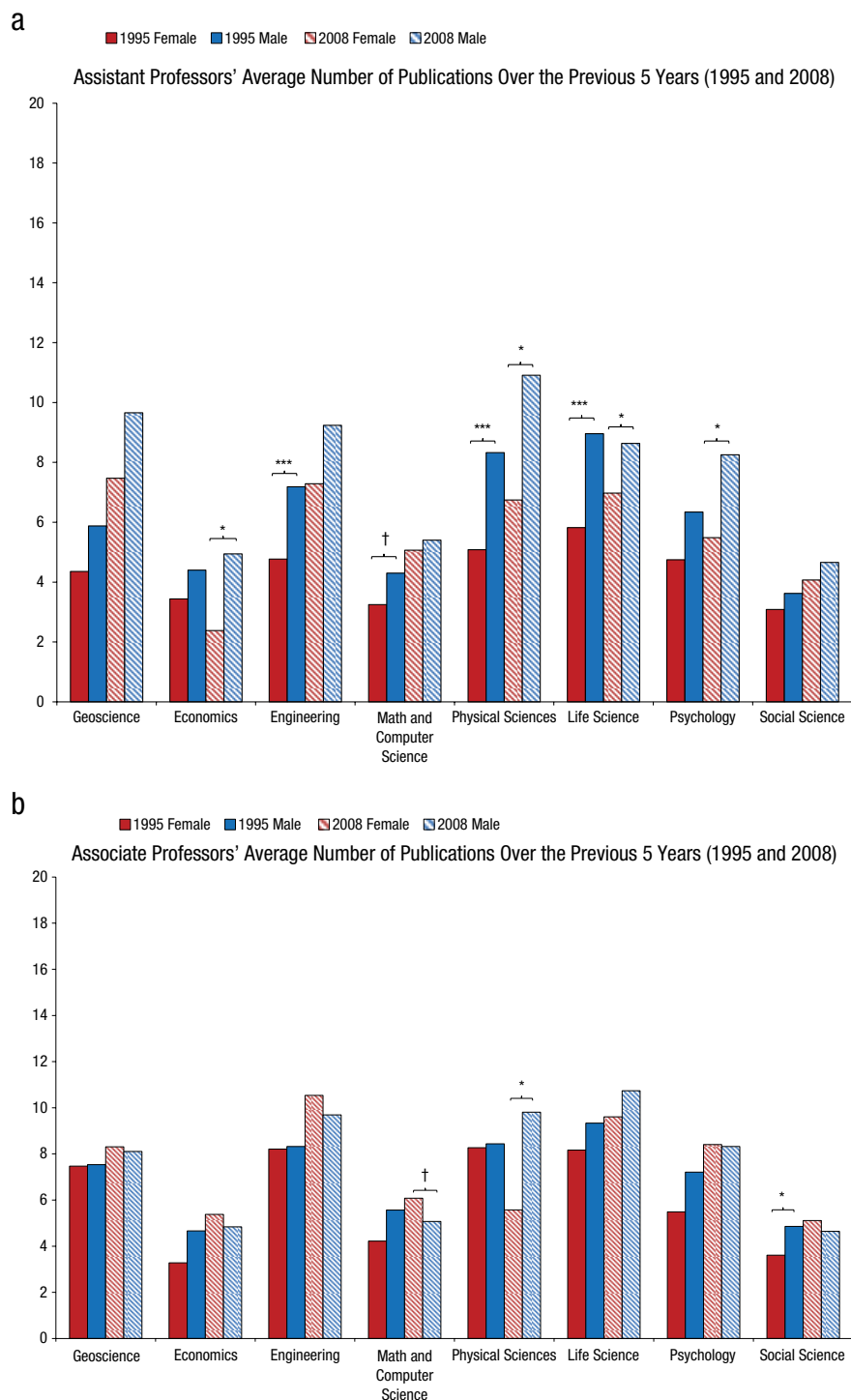
In sum, the male publication advantage narrowed by only a relatively small amount during the 13 years from 1995 to 2008. Disaggregating this average by rank (Table 2), there was a sizable narrowing in the gender gap (by 10 percentage points) at the level of associate professor, with the gap reversing in 2008 in most fields. There was only a smaller narrowing of the gap for assistant professors (by 5 percentage points), the newer generation of academics. And there was a *widening* of the gender publication gap (by 3 percentage points) for full professors. Although this widening at the full-professorship level might be attributable to a greater historic selectivity of female scientists, we have no evidence that shows this to be the cause. We conclude that overall, any equalization in publication rates in the years from 1995 to 2008 was small. And in one field—the physical sciences—the publication advantage of men increased at all ranks.

Other studies have indicated that the productivity gap did narrow *before* the mid-'90s. Xie and Shauman (1998, 2003) reported on three comparable cross-sections of STEM faculty between 1969 and 1993. The 2-year publication averages in the top row showing Xie and Shauman's results in Table 2 do not control for rank or anything else. We see that from 1973 through 1993, the gender gap in publications narrowed from 58% to 18%. This 1993 gender gap is quite similar to the average of 22% in the 1995 SDR. It is also similar to the 22% (for five STEM fields serving as proxies for all of the natural sciences) identified by Fox (2005) for the 1990-through-1994 period. On the other hand, the large 1973 and 1988 gender differences of Xie and Shauman were much higher than the 17% average gap identified by Levin and Stephan (1998) for the 1973-through-1981 period for four STEM fields serving as proxies for all of the natural sciences. The Levin and Stephan study was different from these other studies in that it did not rely on self-reported publication counts—which may be over- or underreported to different extents by women and men.

We could find no time-series comparisons of gender differences in publications that isolate STEM fields, other than Xie and Shauman's. However, Sax, Hagedorn, Arredondo, and Dicrisi (2002) studied publication gender gaps at three points of time between 1972 and 1999 for

all faculty, STEM and non-STEM combined. The Sax et al. study must be regarded as suggestive only, given that gender differences in STEM fields are different from those in the humanities in many other ways, and that during these decades, there were large compositional changes in the percentage of female academics engaged in STEM fields relative to the humanities.

Sax et al. (2002) gives the distribution, rather than the mean, of 2-year publication counts. As Table 2 indicates, the gender gap in the percentage of academics with no publications was halved between 1972 to 1973 and 1998 to 1999, whereas the gender gap in the proportion with three or more publications (over 2 years) improved only slightly. Several other studies have addressed the



(continued)

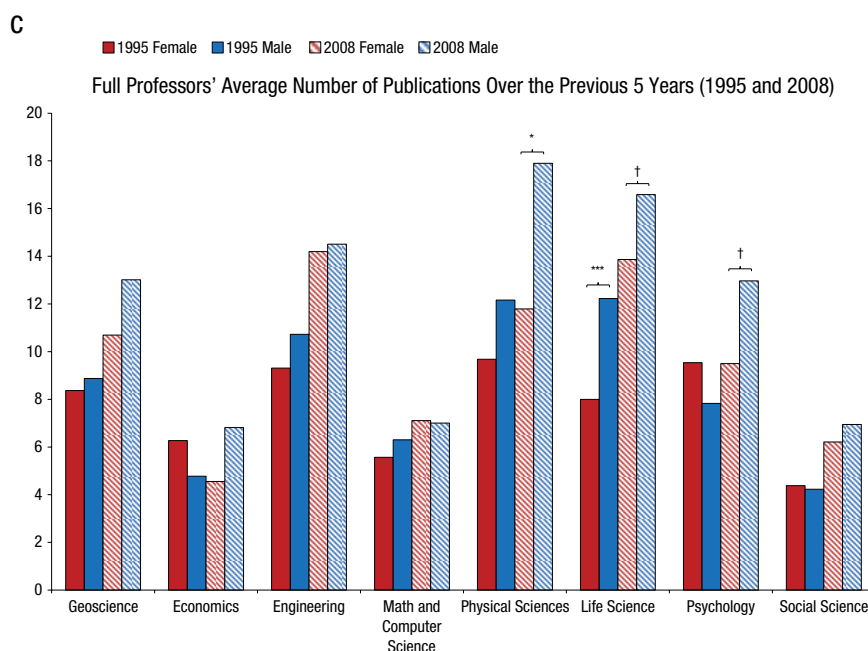


Fig. 14. Average publications over the previous 5 years for assistant professors (a), associate professors (b), and full professors (c) in 1995 and 2008 as a function of subject and gender. Values shown are weighted averages. Asterisks indicate significant differences between males and females ($^{\dagger}p < .10$, $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$). Data shown here were drawn from the National Science Foundation's Survey of Doctorate Recipients (<http://www.nsf.gov/statistics/srvydoctoratework/>).

distribution rather than the mean or median. They, too, have found women considerably more likely to have zero publications (typically publications over a 1- to 5-year period rather than cumulative publications). At the other extreme, men are considerably more likely to be in the top tail (e.g., to have more than 10 publications per period, perhaps echoing the literature on scientific ability).

Two early, highly cited studies are informative but not comparable to those already discussed. Both Cole and Zuckerman (1984) and Long (1992) studied specific cohorts of PhDs as their careers progressed. This is very different from the studies discussed above, which were limited to (full-time) academics or even more narrowly to tenured or tenure-track academics. Using only current tenured or tenure-track academics drops many PhD recipients who did not enter academic jobs, and many more who were unsuccessful in academia and consequently left it. Because women drop out of academia and science more than men (as we discuss below), gender publication gaps will be much larger for complete PhD cohorts. Thus, Long (1992) found a 48% gender gap at 9 years post-PhD for biochemists (1956–1963 cohorts, through 1980), and Cole and Zuckerman (1984) found a 43% gender gap at 11 years post-PhD for all STEM fields.

To compare to these earlier figures, we use the SDR data on publications between 2003 and 2008 for all PhDs at different time spans since PhD, not limiting ourselves

to those in academia. We found an average 20% gap at 6 to 10 years post-PhD and an average 25% gap at 11 to 20 years post-PhD for all STEM fields, noticeably smaller than the 43% gap that Cole and Zuckerman found 25 years earlier, yet still substantial. In life science alone—which includes biochemistry and is thus most comparable to Long's cohort showing a 48% gap—the gaps are 28% to 29% at both ranges post-PhD. In three fields—of which two are GEEMP fields (math, engineering, and social science excluding economics)—women's average publications actually exceed men's average at 6 to 10 years post-PhD, and are very similar at 11 to 20 years post-PhD (analogous to our earlier results for associate professors). Again, this suggests sizable improvement over the decades, yet with considerable gaps remaining, larger than those found among academics alone.

Although patenting is not a widespread academic activity, Table A1 in the Appendix shows that the gender gaps in patenting by academics are large and significant (favoring males) in fields in which patenting is prevalent—engineering, life science, and the physical sciences. This is consistent with findings from Ding, Murray, and Stuart (2006).

Various studies have examined several explanations for gender differences in academic productivity. First, women, especially women with children, may work fewer hours than men. Figure 15—again based on the 2010 SDR—shows surprisingly small differences in the

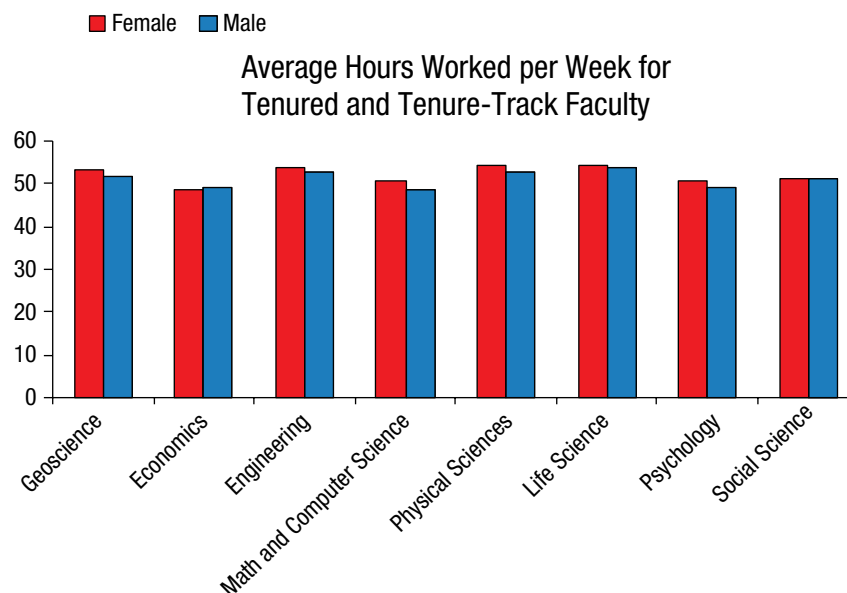


Fig. 15. Average number of hours worked per week as a function of field and gender. Values shown are weighted averages. Data shown here were drawn from the National Science Foundation's 2010 Survey of Doctorate Recipients (<http://www.nsf.gov/statistics/srvydoctoratework/>).

weekly hours worked outside the home by sex and field of tenured and tenure-track STEM faculty. In fact, women work more on average than men in five fields, but the only statistically significant difference is that women work more than men in mathematics/computer science ($p < .05$). Ecklund and Lincoln (2011), in their survey of 1,175 faculty in biology, astronomy, and physics, similarly found no significant difference in the number of hours women and men without children worked.

Lubinski and his colleagues (Ferriman, Lubinski, & Benbow, 2009; Lubinski & Benbow, 2006) have measured actual and preferred weekly hours of women and men in their mid-30s from two groups: those who were identified at early adolescence as being of high mathematical ability, and those who were math and science graduate students in the natural sciences and engineering 10 years previously. In both groups, women preferred to and actually did work fewer hours per week. In particular, there were more women who worked (and preferred to work) less than 40 hours per week—although always less than 15%—and fewer women who worked 60 hours or more per week. Comparing this with the evidence on faculty hours suggests that women who excel in STEM fields but who prefer to limit their work hours may shun academic (tenure-track or tenured) positions. To test this, we again looked at PhD scientists in the 2010 SDR and found that, controlling for age, whereas women in (tenure-track) academia worked the same hours as men, women working outside academia worked 4.4 fewer hours ($p < .001$); in addition, women were 5.4 percentage points more

likely to be out of the labor force. Further, there was a difference between GEEMP and LPS women, with women in GEEMP fields less likely than women in LPS fields to work part-time and on average working more hours than women in LPS fields (about 2.3 hours more); not surprisingly, GEEMP women were more likely to be in tenure-track positions. We conclude that STEM faculty women's lower number of publications cannot be explained by their work hours, but that work hours may hold the key to understanding why women with PhDs in LPS fields are less likely to enter academia, a point that we return to in the Conclusion.

The detailed study by Xie and Shauman (2003) indicates that the higher demands on women faculty's time for teaching (and/or service, mentioned by others) can explain much of women's lower research productivity. Indeed, they found that faculty teaching 11 or more hours per week had much lower research productivity.²⁵ Thus, the second row showing Xie and Shauman's results in Table 2 lists much smaller gender differences, with controls added for teaching and research funding, as well as field, academic rank, and institution type. This finding has been echoed by Misra, Lundquist, Holmes, and Agiomavritis (2011), who found significant differences in time use by STEM faculty at a single research university: Men spent almost twice as much time on research than women and significantly less time on mentoring and service. Using a sample of 150 economists, Manchester and Barbezat (2013) examined how time allocation and time concentration (i.e., how research time was spread across

the academic year and summer months) were related to academic productivity. They found that male economists spend more time on research and concentrate more of their research during the academic year, whereas women concentrate their research in summer months as a result of child-care responsibilities. Time concentration was associated with women submitting fewer articles for publication.

However, to use the economists' terminology, hours spent teaching is highly endogenous. In other words, those academics who cannot or prefer not to do research will teach more (and do more service), including by choosing teaching institutions over R1 institutions, and similarly will get less research funding. In fact, Winslow (2010) used the 1999 National Study of Postsecondary Faculty (NSOPF) to examine gender differences in time allocation of all faculty, and found that women prefer to teach more and do teach more than men, whereas men prefer to do research and spend more time on research. As a result, the much smaller gender productivity gaps seen when controlling for teaching time and research funding in our judgment underestimate the true gender gap in publications.

A second possible explanation for the gender productivity gap is that having children reduces research productivity. Many articles have addressed the relationship between family variables and the gender productivity gap by measuring separate impacts of family-related variables for women and men. Xie and Shauman (2003), Stack (2004), and Sax et al. (2002) found that family-related variables had no effect or relatively small effects on productivity; however, these studies again controlled for highly endogenous variables that can soak up much of the effect of children, such as teaching hours, federal support, research orientation, and salary. Without these controls, some studies have revealed large negative effects of children on productivity. For instance, Fox (1995) found that elementary-school-age children (only) and divorce significantly decreased the productivity of female STEM academics, whereas both marriage and children (of any age) significantly *increased* the productivity of male STEM academics. However, in an older study of chemists, Hargens, McCann, and Reskin (1978) found that children (significantly) slowed the productivity of men and women equally. Likewise, Hunter and Leahey (2010) showed that children decrease productivity growth in the social-science fields of sociology and linguistics for both women and men, but more so for women. Ginther and Kahn's results (2009, in press), controlling for exogenous variables such as PhD quality and rank but not for arguably endogenous variables such as research funding and teaching hours, have shown that women assistant professors with children in social

science and geoscience—but not other fields—publish fewer papers than women without children.

To bring the analysis of publications and children up to date, we again use data from the 2008 SDR. In Figure 16, we examine whether the presence of children is associated with the productivity gender gap by looking at publications of assistant professors without children. As shown in Figure 16a, publications by single, childless females are not significantly different from those by single childless males in any field except economics and the physical sciences. Publications in life science and psychology, which had significant gender differences on average for assistant professors, are not different when limited to childless singles.

However, a visual comparison of Figures 16a and 16b is enlightening. In each field except mathematics/computer science, the physical sciences, and psychology, men with children published *more* than men without children. This pattern is likely to be due to selection bias; positive correlations of men's being married and/or having children and their productivity and/or wages are seen across the labor market. The differences between these two groups of men are the major differences in the two graphs, rather than any differences between women with and without children. Women without children publish noticeably more than women with children only in geoscience and psychology. Thus, except for these two fields, the presence of children cannot explain the overall gender productivity gaps.

We have also investigated whether having children reduces academics' work hours, and whether this effect is larger for women (who typically have the major responsibility for child rearing) than for men. We repeated the hours analysis of Figure 15 for men and women with children (see Section D in the Appendix) and found, not surprisingly, that women *and* men with children both work less (outside the home) than do those without children. However, we found that the only field in which women with children work *significantly* less than men with children was physical science ($M = 2.9$ hours less per week; $p < .10$). There were no other statistically significant gender differences in hours of work as a function of field and presence of children. Ecklund and Lincoln (2011), in their survey of 1,175 faculty in biology, astronomy, and physics, also found that while children lowered work hours, the impact was similar for women and men.

Before we leave this discussion of children and productivity, we note that all of the literature in this area is only descriptive. Links observed between children and research productivity cannot be disentangled to evaluate whether the associations found are due to selection—that is, whether either more or less able researchers end up having children—rather than being the effect of children

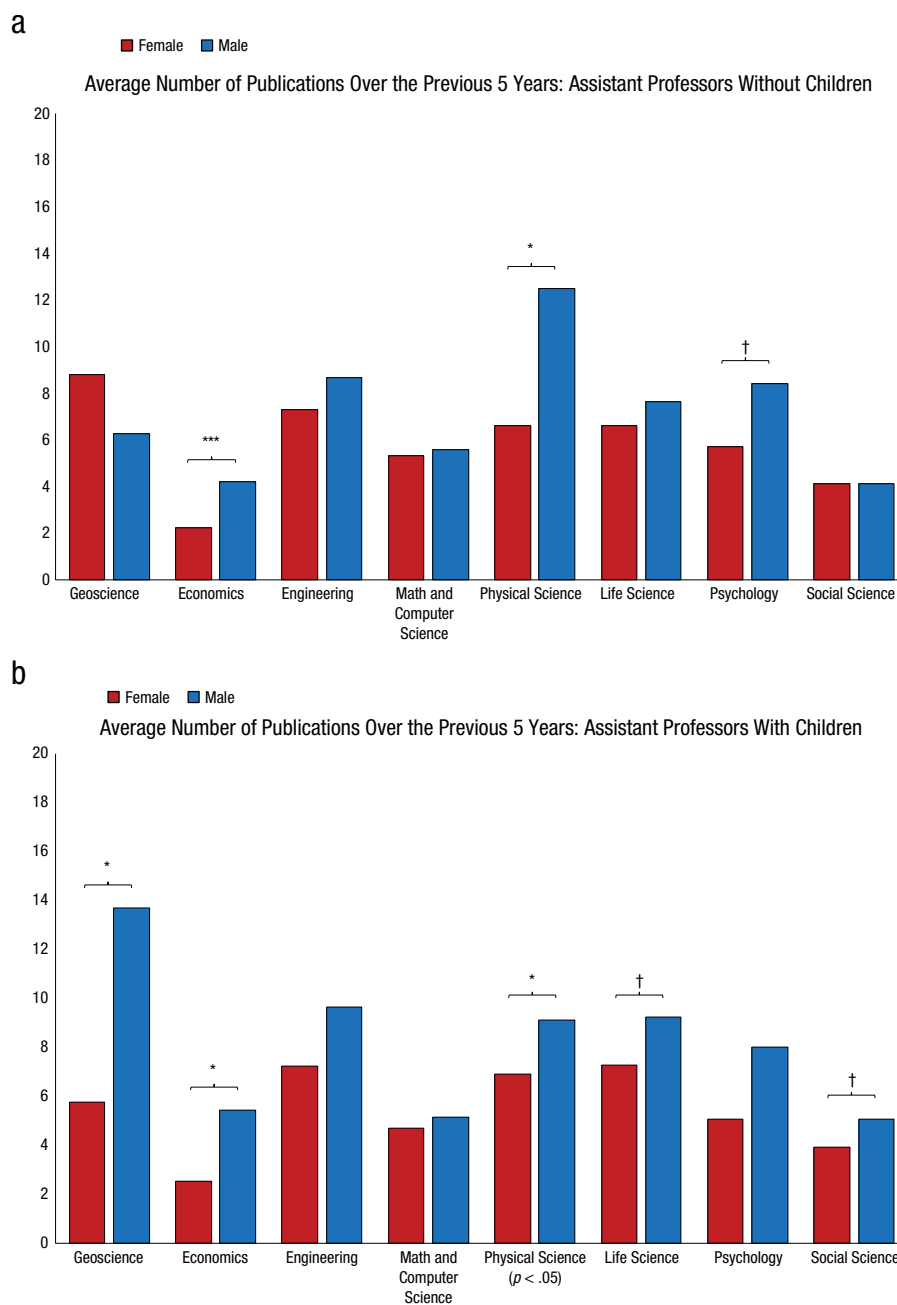


Fig. 16. Average number of publications over the previous 5 years for assistant professors without children (a) and assistant professors with children (b) in 2008 as a function of field and gender. Values shown are weighted averages. Asterisks indicate significant differences between males and females ($p < .10$, $*p < .05$, $**p < .01$, $***p < .001$). Data shown here were drawn from the National Science Foundation's Survey of Doctorate Recipients (<http://www.nsf.gov/statistics/srvydoctatework/>).

on productivity. This is a problem for both those studies with control variables and those without controls. In order to tease out selection explanations from the impact of children on productivity, a better study would look at the effect of childbearing on number of publications before and after having children. Cole and Zuckerman (1987) did this with a small sample and found that

marriage did somewhat lower women's productivity but the birth of children did not. More recent and larger analyses of this sort are needed.

A third explanation given for sex differences in productivity is the existence of smaller professional networks and fewer coauthors, ultimately resulting in fewer publications. Sex differences in the frequency of coauthorship

may result from faculty preferring to collaborate with researchers of the same sex, such that fewer women in a field may result in fewer papers in that field coauthored by women (Bukvova, 2010; McDowell, Singell, & Stater, 2006; McNeely & Schintler, 2010). Unsurprisingly, research has shown that coauthorship can strengthen a scientist's publication record (Bukvova, 2010; Fox & Mohapatra, 2007; McDowell et al., 2006; McNeely & Schintler, 2010).

A recent article (Duch et al., 2012) offers evidence of an additional reason that women in STEM may publish less than men. These authors documented an association between STEM fields that have large resource requirements for research (e.g., molecular biology) and a larger gender publication gap at selected top research institutions in the United States. They related this gender gap to the fact that historically, women have had less access to institutional resources and support (see, e.g., Massachusetts Institute of Technology, 1999). Industrial engineering, the field in their study with the fewest resource requirements, has the smallest gender publication gap. A final possible reason was identified by Leahey (2006), who found that women in sociology and linguistics are less likely to specialize in their research topics, with less specialization resulting in fewer publications.

In summary, the data show that women in STEM fields on average are significantly less productive than men at the assistant-professor rank. Economists, including those coauthoring this article, believe that this is *prima facie* evidence that shifts at least part of the responsibility for women's limited academic success away from employing departments—unless one wants to argue, as some do, that more should be done to help mothers with young children. Unfortunately, there are relatively few studies that provide compelling explanations for these productivity differences. The Duch et al. (2012) argument based on resources is a promising new direction, arguing that the problem lies in institutional resource decisions. This and other possible explanations warrant additional investigation to illuminate causes of the gender productivity gap, research that could benefit from new data sources that link individual researchers to their research output. As seen above, scientific fields often differ significantly in the productivity of women with children. It has been suggested that the fastest-paced GEEMP fields—those that experience the most rapid knowledge decay and require regular technical updating—will have the largest penalties for family leaves, an argument advanced by Lubinski and Benbow (2007, p. 93). Taking time out from fast-paced careers for even a limited amount of time is difficult, given the rapid accumulation of knowledge and technical advances in such careers. In some GEEMP fields, much more than 40 hours a week may be needed to stay up-to-date to have a high-impact career (Lubinski & Benbow, 2007). Future research might be directed at this issue.

Potential explanation #6: Biased work-product evaluation

It could be that women are less productive than men in terms of publications because the review process is biased against them. A number of recent analyses of journal reviewing have been reported, and we summarize them here. (The interested reader should consult Ceci & Williams, 2011, for a more detailed report.) To adumbrate the conclusion of this article, there have been no systematic sex differences in work-product evaluation during the past two decades: There are similar journal-acceptance rates as well as grant-funding rates for male and female authors, a finding that has been in evidence for at least two decades (e.g., J. R. Gilbert, Williams, & Lundberg, 1994; Grant, Burden, & Breen, 1997; Hammerschmidt, Reinhardt, & Rolff, 2008).

Sex differences in journal acceptance rates. Some have claimed that editors and reviewers are biased against accepting women's manuscripts. Budden et al. (2008) reported that women's acceptances rose 33% for the journal *Behavioral Ecology* after it implemented blind review so that reviewers did not know the gender of the authors:

Research on anonymous refereeing shows fairly clearly that biases play a role in evaluating work . . . one such journal, *Behavioral Ecology*, recently decided to [implement blind review]. They found that it led to a 33% increase in representation of female authors. (quoted in Saul, 2013, p. 41)

According to some, this reviewing bias contributes to women's underrepresentation because it results in fewer publications, which has both direct and indirect effects on women's career advancement. Relatedly, a number of experiments have shown that the same manuscript is rated higher when it has a male name on it, although raters in these studies are usually students, not reviewers (i.e., Foschi et al., 1994; Swim et al., 1989). When it comes to actual manuscripts submitted to actual journals, the evidence for gender fairness is unequivocal: There are no sex differences in acceptance rates. As seen below, the reason women have fewer publications is not that their manuscripts are rejected at higher rates, but rather that they submit fewer manuscripts.

One way to test the hypothesis that journal reviewers are biased against women's manuscripts is to send identical versions of a manuscript, with female and male names, to the same reviewers. When this has been done, women and men have been treated similarly (e.g., Borsuk et al., 2009). Another type of evidence is overall review outcomes for male and female authors. Nearly all such studies reveal gender neutrality. For example, nearly

3,400 publication recommendations submitted to the *Journal of the American Medical Association* revealed no association between author gender and acceptance (J. R. Gilbert et al., 1994), a finding repeatedly found for other journals (e.g., Blank, 1991; J. Brooks & Della Sala, 2009; Grant et al., 1997; Hammerschmidt et al., 2008; Nature Neuroscience, 2006; Tregenza, 2002).

In a recent study on economists, the authors tested directly for the results of bias in the review process. Abrevaya and Hamermesh (2012) examined gender dynamics in refereeing economics papers. Using over 5,000 papers from an economics journal in which the sex of the authors was matched to the sex of the referees, they found no evidence of bias (either same-sex favoritism or opposite-sex discrimination) in the referee process.

Sex differences in citation rates. Another measure of the evaluation of research is whether a paper is cited by other researchers. Because women publish fewer papers, they have reduced opportunities to receive citations compared with men. However, whether gender differences in citation rates per article exist is a different question.²⁶

Citations per article can only be compared within fields, or analyzed while controlling for fields, because fields differ considerably in their citation protocols. Controlling for field, citations are typically viewed as a measure of the quality of the article. On the other hand, citations may be a measure of the authors' networks (as in the old-boy network) or the authors' networking skills. Gender bias, if it exists, could affect citations per article in two ways. First, an article's citations depend considerably on the prestige of the journal it is in, so any editorial or reviewer bias could also be reflected in citations per article. Second, people's evaluation of any article's quality and cite-worthiness may be influenced by the authors' gender. Third, citations could be a function of the "narrowness" of the topic or field, though we know of no data showing that women work on narrower topics in narrower fields.

The literature on gender differences in citation rates per article is summarized in Table 3. Most of this literature is limited to a single field, a single country, or a single field within a single country, and therefore may not be representative of all fields and countries. Also, many studies have involved limited numbers of observations. (This table excludes studies with fewer than 75 subjects.) Finally, Table 3 includes only citation studies that controlled in some way for the period of time that the citations had to accumulate (since the publication date). Note that Table 3 indicates a much more limited and less comprehensive set of articles than were available on publication rates. A major reason for this is that although the availability of citation data is rapidly expanding,

matching these data to individual characteristics (including gender) remains a slow process.

Perusal of this table leads to the conclusion that in general, there is no gender difference in citations per article. Women do seem to have fewer average citations per publication than men in archaeology (Hutson, 2002), international relations (Maliniak et al., 2013), and in Norway (Aksnes, Rorstad, Prio, & Sivertsen, 2011). Women seem to have more average citations per publication than men in ecology (Duch et al., 2012) and in political science in Canada (Montpetit, Blais, & Foucault, 2008). Also, before 1990, women had more citations per publication in biochemistry. Those authors who find a female advantage typically emphasize the quality dimension, arguing that women produce higher-quality, if less, work.

Two articles limited the analysis both by country of citing as well as cited author. Aksnes et al. (2011) found lower citations to Norwegian women in Norwegian publications. Kahn and MacGarvie (2014) found that when they considered only *citers* from outside the United States (and more heavily in developing countries), women were cited less. But for the same group of authors and publications, Kahn and MacGarvie found that women were not cited less by citers in the United States. Together, these two articles suggest that authors in some countries other than the United States may cite women less than men.

In sum, an overwhelming amount of evidence reveals no gender differences or higher citation rates for women that, if found, might have suggested that women publish higher quality articles.

Sex biases in grant funding rates. Numerous commentators have claimed that sex bias in grant review is responsible for fewer women getting funded (or getting funded at lower levels) and that this failure to gain grants is responsible for women's lower rate of persistence and lower rate of promotion. For example, Lortie et al. (2007) wrote that "it is now recognized that (sex) biases function at many levels within science, including funding allocation, employment, publication, and general research directions" (p. 1247).

Notwithstanding such claims and myriad others (e.g., Wenneras & Wold, 1997), there are no systematic sex differences in grant-funding rates, although men's grant awards tend to be for higher dollar amounts, likely as a result of men's greater likelihood of being principal investigators on large center grants and program projects. Overall, however, men and women have very similar funding rates for their grant proposals (e.g., Jayasinghe, Marsh, & Bond, 2003; Ley & Hamilton, 2008; Marsh, Bornmann, Mutz, Daniel, & O'Mara, 2009; Mutz, Bornmann, & Daniel, 2012; Pohlhaus, Jiang, Wagner,

Table 3. Gender Differences in Citations per Publication

Article	Data years of citations	Field	Country of citing authors, cited authors, or both	Whose gender?	Gender difference	Inclusion of self-cites	Statistic	Other quality controls
Aksnes, Rorstad, Prio, and Sivertsen (2011)	2005–2009	Science and engineering	Norway (both citing and cited authors)	Any author	Women lower	Yes	Mean	Yes, including journal
Borrego, Barrios, Villaroya, & Olle (2010)	1992–2006	Science and engineering	Spain (cited PhD-holding authors)	Any author	Women higher	Yes	Percentage with one or more publications	None
					None	No	Percentage with one or more publications	
					Women higher	Both ways	Median	
Duch et al. (2012)	Through 2010	Ecology	United States (cited authors)	Any author	Women higher	Unknown	<i>b</i> -index adjusted for number of publications	Only top departments; only highly cited articles 10 or more years after publication
		Chemistry, psychology, molecular biology			Women higher			
Hopkins et al. unpublished results (same data as Hopkins, Jawitz, McCarty, Alex Goldman, & Basu, 2013)	Through 2005	Biochemistry, water resources, anthropology, economics	International (majority from United States; cited authors)	Any author	No significant difference at assistant- or associate/full-professor levels	Yes	Mean	Rank
Hutson (2002)	1998–2000	Archaeology	Two U.S. journals (cited authors)	Only male versus only female authors	Women lower	No	Mean	Journal
Kahn and MacGarvie (2014)	1994–2008	STEM	United States (citing authors)	Random author	None	No	Top-coded	Yes, including limiting count to early career
			Various countries, mostly developing (citing authors)		Women lower			
Lewison (2001)	1980–2000	Clinical medicine	Iceland (cited authors)	Percentage of female authors	No significant difference (women higher)	Yes	Mean	No
Long (1992)	1956–1991	Biochemistry	United States (cited authors)	Any author	Women higher	Yes	Mean	No
Maliniak, Powers, & Walter (2013)	1980–2013	International relations	Twelve international journals (cited authors)	Only female versus mixed or only male authors	Women lower	Yes	Mean	Yes, including journal and institutional quality
Montpetit, Blais, and Foucault (2008)	1985–2005	Political science	Canada (cited authors)	Majority of authors	Women higher	Yes	Mean	Yes, including journal impact
Slyder et al. (2011)	Through 2010	Geography/forestry	United States (cited authors)	First author	No significant difference (women higher)	Yes	Mean of maximum citations per author	Yes, including journal
Smart and Waldfogel (1996)	1980–1990	Economics/finance	United States (cited authors)	Any author	None	Yes	Median and mean	Yes, including journal
Symonds, Gemmell, Braisher, Gorringe, and Elgar (2006)	1990–2005	Ecology, evolutionary biology	United Kingdom and Australia (cited authors)	Any author	No significant difference (women higher)	Yes	Median	None
					Women higher			Total publications

Schaffer, & Pinn, 2011; RAND, 2005). Below, we summarize this literature.

In the aftermath of Wenneras and Wold's (1997) finding of biased grant reviews of 114 Swedish postdoctoral fellowships, many have claimed that women's success in tenure-track positions is stymied by biases in grant awards, arguing that for a woman to be funded, she had to have on average 2.5 more major publications (i.e., in top journals) than comparable male competitors to get the same score. However, a comprehensive analysis of the data does not accord with this claim, and the full corpus of evidence does not reveal an anti-female bias in grant reviews (Ceci & Williams, 2011). Here, we add to the evidence.

The European Molecular Biology Organization (EMBO) funds scientists through various award mechanisms and has been tracking the gender fairness of its grants for 15 years. It has repeatedly found that the success rate of male applicants is approximately 20% higher than that of females. Ledin, Bornmann, Gannon, and Wallon (2007) analyzed the EMBO data over two rounds of reviews in 2006. Their first analysis eliminated all references to gender contained in the grant applications, letters of recommendation, and reports sent to reviewers for scoring:

Nevertheless, the difference in success rate persisted. The finding that the committee reached the same conclusions when [proposals and letters] were gender-blinded challenges some of the usual explanations given for the differences in success between male and female scientists. We therefore looked for bias introduced from an external source. A recent publication suggested that letters of recommendation are written differently for men and women (Trix & Psenka, 2003), and we wondered whether this was the case . . . We independently read the 283 reports from the Spring 2006 deadline and tried to deduce the gender of the applicants from the language used, as described by Trix & Psenka (2003) . . . We concluded that it was not possible to accurately determine the gender . . . so this could not be an alternative explanation for the lower average success rate of women. (Ledin et al., 2007, p. 982)

Men's proposals continued to be funded by the EMBO at a rate 20% higher than women's, even when reviewers were unable to distinguish the gender of the principal investigator. These data provide no evidence of anti-female bias, consistent with the following review of large-scale analyses of U.S. grants.

Sex differences in federal grant funding in the United States are much less apparent. Hosek et al. (2005)

reviewed sex differences in grants awarded by the NSF, the U.S. Department of Agriculture (USDA), and the National Institutes of Health (NIH) between 2001 and 2003. The study found no significant sex differences in NSF and USDA awards. However, it found that women received smaller amounts of funding than men at the NIH. This resulted from women's being less likely to receive very large awards. These results must be qualified because the study could not distinguish whether the size of women's awards was smaller because they asked for less money. In addition, women were less likely to reapply within 2 years of submitting their initial proposal to both the NSF and the NIH.

Other studies have provided a more detailed look at sex differences in NIH funding. An analysis of roughly 100,000 grant applications to the NIH over a 5-year period (2003–2007) by Ley and Hamilton (2008) is one of many large-scale studies that has supported the claim of gender-neutral grant reviewing. These studies have provided compelling evidence against gender bias. As will be seen, the overall grant pattern is one of gender neutrality, not male superiority, notwithstanding isolated grant categories in which one sex or the other excels (see Figs. 1–6 in the Supplementary Materials for Ley and Hamilton, 2008), a finding that is also true for virtually all other grant agencies:

Despite the oft-held perception that women do not fare as well in the NIH grantee pool as men, the data show that funding success rates for nearly all grant (categories) were essentially equal for men and women, regardless of degree (Ph.D., M.D., Ph.D./M.D.) . . . When the data are pooled for all investigators and all grants studied from 2003 to 2007, the success rates for men and women are virtually equivalent (31% success for women, and 32% for men). (Ley & Hamilton, 2008, p. 1473)

More recently, Pohlhaus et al. (2011) examined sex differences in NIH applications and funding rates. They found few sex differences in NIH awards. However, like Ley and Hamilton, they found that women were less likely to apply for Research Project Grant (R01) awards conditional on applying for Clinical Investigator (K08), Small Grant (R03), and Exploratory/Developmental Research Grant (R21) awards. Although women were equally likely to be successful in applying for new R01 awards, they were less likely to be funded for R01 renewals than men. Pohlhaus et al. did not investigate whether differences in research productivity from the first R01 grant explain women's diminished award rates for R01 renewals. The limited sex differences in NIH funding stand in marked contrast to the large race/ethnicity differences in NIH funding uncovered by Ginther et al. (2011), who found

that black researchers were one third less likely to receive NIH funding compared to white researchers.

These analyses of U.S. grants is consistent with a half dozen other large-scale analyses based on hundreds of thousands of grant applications to agencies throughout the world, which together lead to an unequivocally bias-free conclusion. The latest of these was conducted on nearly 24,000 reviews of 8,500 grants (Mutz et al., 2012), again confirming that the decision to award grants was not associated with applicant's gender or the interaction between it and the reviewer's gender. (See Ceci & Williams, 2011, for four other large-scale analyses that accord with this conclusion.)

This massive evidence of gender fairness in grant reviewing, based on hundreds of thousands of reviews, is seldom cited by studies claiming sex discrimination, which instead emphasize Wennerås and Wold's (1997) study of 114 Swedish postdoctoral applicants. Although the 1997 study implied that sex discrimination had not been entirely eliminated by that point in time, now the overwhelming picture is one of gender neutrality in the grant-review process. Approval rates of women and men are "virtually equivalent" (Ley & Hamilton, 2008), with occasional exceptions that benefit each sex—such as EMBO funding, which favors men in gender-blind competitions, NIH center grants, which also favor men, and NIH K01, K08, and Loan Repayment Program (LRP) grants, all of which favor women.

Does this mean that every analysis of manuscript reviewing has shown gender-neutral outcomes? Of course not. However, the departures from gender-neutral outcomes have been rare and as likely as not to result in greater female acceptance rates as greater male acceptance rates. As Ceci and Williams (2011) concluded:

Although there are occasional instances of sex effects, they are rare, of small magnitude, and are as often in favor of women as against them; the largest aberrations were not close to Wennerås and Wold's finding that women had to be 2.5 times more productive than men to obtain similar scores . . . Sandstrom and Hallsten analyzed more recent data from the Swedish [Medical Research Council] (same data set used by Wennerås and Wold) and found that the gender bias reported by Wennerås and Wold (29) had reversed itself several years later, with a small but significant effect in favor of funding women's grants compared to men's with the same score. (p. 3157)

Potential explanation #7: Sex differences in academic promotion

As shown above, women make up small percentages of STEM graduate students, tenure-track faculty, and tenured

faculty, and this is especially true in physical-science and engineering disciplines (Ginther, 2006a, 2006b; National Research Council, 2001). Many studies have tracked the numbers of women in science at various stages of their academic careers (National Science Foundation, 2012)—for example, showing that women continue to be less likely than their male colleagues to be full professors and more likely to be assistant professors. A related literature has examined sex differences in faculty tenure and promotion (Nettles, Perna, & Bradburn, 2000; Perna, 2001a, 2005), but these analyses have tended to combine all academic fields, and Ginther and Kahn (2004, 2009, in press) have cautioned that one cannot generalize the findings from one academic discipline (e.g., engineering) to others (e.g., life science). Thus, we focus on research that has examined sex differences in promotion in disaggregated academic STEM disciplines. This literature is thin, and the findings are somewhat different than expected (Ginther, 2006a, 2006b; Ginther & Kahn, 2009; Long, 2001; National Research Council, 2010). The results suggest few barriers to women's advancement from tenure-track jobs to tenured ones in math-intensive fields, once researchers control for observable characteristics including academic productivity. There is some evidence of barriers in life science and in economics.

The NSF did identify barriers in STEM fields (as a whole) in its comprehensive study of the factors contributing to promotion in academic careers of scientists and social scientists combined (NSF, 2004). This work showed that, controlling for human capital, personal characteristics, and institutional factors, there remains a significant female disadvantage in the likelihood of being in a tenure-track job, of receiving tenure, and of being promoted to full professor. However, in most of the NSF researchers' specifications, they find that these gender differences become statistically insignificant when family characteristics are allowed to affect women and men differently, as they likely do in the real world. Furthermore, they combined all STEM fields in their analysis, which masks important differences across fields.

The National Academies surveyed the departments of biology, chemistry, civil engineering, electrical engineering, mathematics, and physics at R1 universities to evaluate sex differences in promotion to tenure, promotion to full professor, and time in rank (National Research Council, 2001). The report found few sex differences in academic career progression. Although women were less likely to be considered for tenure than men, once considered they were more likely than men to receive tenure. The report also found no significant sex differences in being considered for and promoted to full professor. Finally, it found that women spend more time in rank as assistant professors than do men, but there were no sex differences in time in associate rank.

Kaminski and Geisler (2012) tracked the retention and promotion of almost 3,000 science tenure-track or tenured faculty at 14 universities in the fields of electrical, mechanical, civil and chemical engineering, physics, mathematics, computer science, and biology. They found no significant sex differences in promotion or retention of faculty. The one exception was in mathematics, in which retention times of faculty are short, and significantly shorter for women compared with men.

Using the SDR, Ginther and Kahn (2009) echoed many but not all of these results, despite the fact that Ginther and Kahn controlled for a wide variety of background characteristics and for publications as well. They found no significant gender differences in promotion to tenure in physical-science and engineering fields—fields in which women tend to be underrepresented. They did find differences in life science, however. In their subsequent analysis of social science, Ginther and Kahn (in press) found significant and sometimes large differences in promotion to tenure in economics and psychology. Some of these differences are explained by family characteristics, which will be discussed later in this review. With the exception of economics, the fields with promotion issues are ones in which women have a critical mass of female students and faculty.

Taken together, the research indicates no significant sex differences in promotion to tenure and full professor in the GEEMP fields. However, women are significantly less likely to be promoted in some of the fields in which they are most prevalent: life science and psychology.

Economics is an outlier, with a persistent sex gap in promotion that cannot be readily explained by productivity differences.

Potential explanation #8: Sex differences in academic salaries

Although evidence of gender differences in promotion in science and engineering fields is scant, there are sizable gender differences in salaries in these fields. These differences may in turn lead to women leaving academia or leaving science entirely.

One recent survey reported that women in life science earn significantly less than men across almost all job categories (Dunning, 2012). Each year, the American Association of University Professors produces its *Annual Report on the Economic Status of the Profession*, which often includes salary comparisons by sex; almost all of these data show that men earn more than women. Myriad papers have examined sex differences in salaries across all academic disciplines (e.g., Perna, 2001b; Toutkoushian, Bellas, & Moore, 2007; Toutkoushian & Conley, 2005), finding that men earn more than women. However, as with promotion, comparing sex differences in salaries

across fields and academic ranks is problematic. Different fields pay different salaries (e.g., engineers earn more than life scientists), women and men select different academic fields, and men are more prevalent in the senior ranks—which also pay higher salaries. Thus, the goal of any salary comparison should be to make apples-to-apples comparisons: Individuals in the same fields, in the same academic ranks (and years in rank, when possible), and in similar institutions should be compared in order to reveal whether there are significant sex differences in salaries.

In Table 4, we present sex differences in average annual salaries by field and academic rank from the 2010 SDR. There are significant salary differences across fields—engineers earn more on average than those in other fields. Within fields, there are significant differences across rank—assistant professors earn less than associate and full professors. And within institutions, faculty employed at R1 universities earn more than colleagues at other institutions that tend to place a greater emphasis on teaching.

However, when we look within field and rank, we see only a few significant sex differences in salaries. In 2010, only 6 of the 24 field-rank cells have salaries of males significantly greater than those of females: assistant and full professors in economics, assistant professors only in life science, associate and full professors in engineering and the physical sciences, and full professors only in geoscience. When we isolate salary differences at R1 institutions alone (see Table 4), they tend to be smaller in the majority of the rank/field cells (15 of 24 cells); conversely, salary differences at non-R1 universities tend to be larger (13 of 24 cells).²⁷

Has this always been the case? Figure 17 presents the ratio of average female salaries to male salaries in 1995 and 2010 by rank and field. Values less than 1 indicate that men earn more. These graphs allow us to determine whether the gender salary gap has narrowed over this time period.

For assistant professors (Fig. 17a), women earned significantly less than men in 1995 in four fields (engineering, life science, the physical sciences, and psychology). In 2010, gender differences in salaries in these fields were smaller and no longer significant in all but one field, life science. There was not uniform improvement. In fact, the gender gap narrowed in four fields but widened in the other four, including life science. In economics, the gap widened sufficiently that it became significant.

Figure 17b shows salary differences at the associate-professor rank. Women earned significantly less than men in 1995 in only two fields (geoscience and life science). In 2010, gender differences in salaries in both of these fields were smaller and no longer significant, and no newly significant gaps emerged. In four of the eight fields, women earned relatively more, and in fact in three of these fields

Table 4. Annual Salaries for Tenured and Tenure-Track Academics by Sex, Rank, and Field

Field and gender	Assistant professor	Associate professor	Full professor
All institutions			
Economics			
Male	95,897*	102,749	150,006***
Female	83,310	106,728	110,872
Engineering			
Male	93,122	100,798 [†]	137,004*
Female	93,436	93,302	122,283
Geoscience			
Male	65,438	75,587	113,293*
Female	66,020	76,584	99,993
Life science			
Male	85,118**	91,005	135,624
Female	73,376	86,657	130,803
Mathematics/computer science			
Male	73,923	80,498	112,758
Female	70,907	79,824	115,401
Physical sciences			
Male	68,751	77,511	127,826*
Female	64,535	75,343	106,115
Psychology			
Male	65,701	79,736	124,186
Female	63,037	74,789	112,806
Social science			
Male	62,354	77,748	117,999
Female	60,811	78,567	111,368
Research I institutions			
Economics			
Male	112,774	119,813	188,087
Female	104,996	177,455	181,124
Engineering			
Male	113,072	110,392	145,722
Female	114,465	99,468	149,175
Geoscience			
Male	71,003	87,243	131,831 [†]
Female	79,797	82,853	110,635
Life science			
Male	101,826	105,062	157,078
Female	91,301	100,306	154,351
Mathematics/computer science			
Male	85,326	93,780	131,010
Female	89,849	104,168	152,468
Physical sciences			
Male	84,665	87,127 [†]	155,234
Female	78,275	109,470	143,269
Psychology			
Male	76,630	97,418	160,122
Female	75,598	87,835	139,573
Social science			
Male	72,505	84,188	144,606
Female	67,648	83,231	132,418

Note: Data shown here were drawn from the National Science Foundation's 2010 Survey of Doctorate Recipients (<http://ncesdata.nsf.gov/doctoratework/2010/>).

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

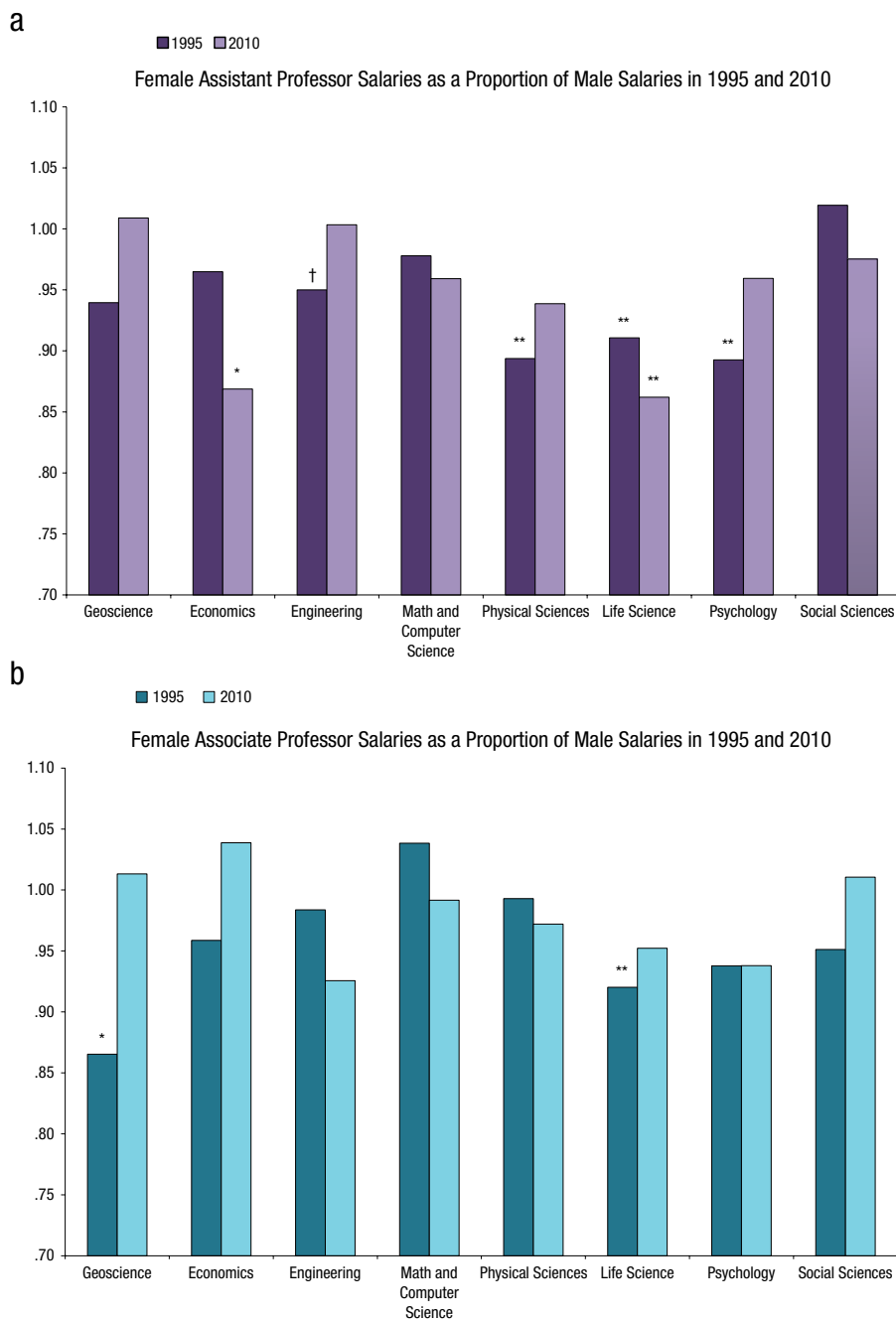
(economics, geoscience, and social science), women on average earned more than men in 2010.

Sex differences are larger at the full professor rank (Fig. 17c). The gap in 1995 was significant in five of the eight fields; in each of these fields, the gender gap in salary narrowed by 2010 and remained significant only in engineering, geoscience, and the physical sciences. However, the gap widened in the other three fields. As was the case with assistant professors, economics is the outlier—female full professors in economics went from

earning 95% of what male full professors earned in 1995 to less than 75% of what male full professors earned in 2010, a large and statistically significant difference.

On balance, progress toward salary equalization has been made in some fields and ranks between 1995 and 2010. However, notable salary gaps remain, and in 9 of the 24 field ranks, the salary gap widened.

Economists have many competing explanations for sex differences in salaries, including preferences, productivity, job matching, and negotiation. Women's



(continued)

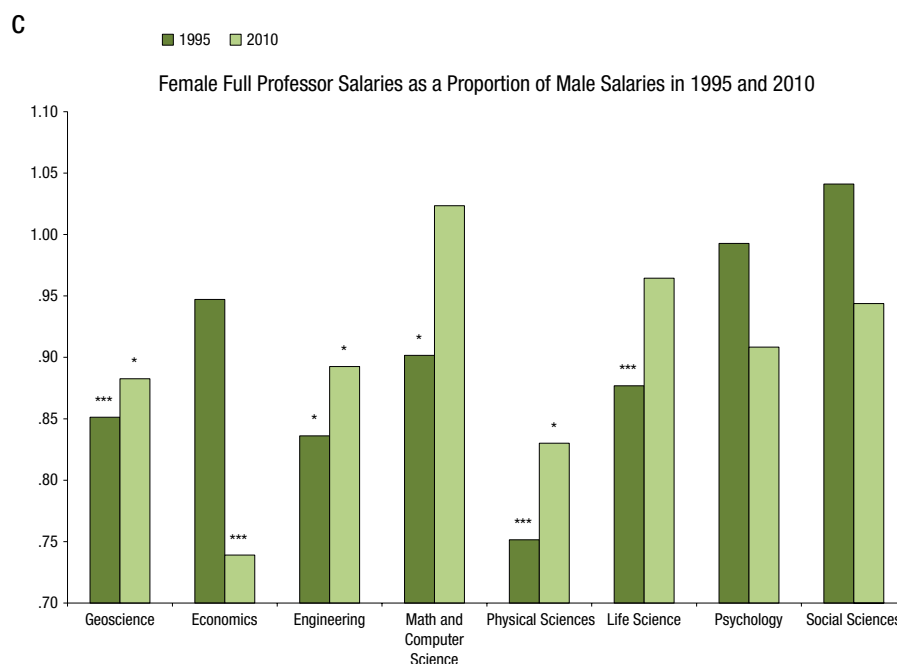


Fig. 17. Female salaries as a proportion of male salaries in 1995 and 2010 among assistant professors (a), associate professors (b), and full professors (c). Values shown are ratios of weighted averages. Asterisks indicate significant differences between males and females ($p < .10$, $*p < .05$, $**p < .01$, $***p < .001$). Data shown here were drawn from the National Science Foundation's Survey of Doctorate Recipients (<http://www.nsf.gov/statistics/srvydoctoratework/>).

“preferences” for spending time in child care²⁸ may be directly related to their salaries. Hundreds of studies have identified a “child salary penalty” for women in the labor market as a whole, while at the same time identifying a marriage and child premium for men (e.g., Budig & England, 2001; Waldfogel, 1997). Much of this child premium has been attributed to weekly hours, gaps in careers, and occupational choices that accommodate flexibility with hours worked (Goldin, 2014). It is a different question whether this child penalty extends to high-skilled occupations like those of STEM academics, who have signaled their commitment to their field.

Using SDR data through 2001, Ginther (2004) examined the economic explanations for gender differences in academic salaries in science and engineering fields and found that children and marriage did not explain substantial gender salary gap. She did not find that including measures of productivity appreciably reduced the salary gap, although these results must be qualified because the SDR measure of publications is imperfect and, at its most accurate, is based on selected surveys from the previous 5 years. Furthermore, publication counts may not be as important a determinant of salary as citations to those publications.

In contrast, Kelly and Grant (2012), using the 2004 NSOPF, did find that marriage-based salary penalties in the fields of science, engineering, and math²⁹ are explained

by productivity differences caused by married women's publishing fewer papers (which leaves open the question of why this does not lead to penalties for hiring, tenure, and promotion as well). Controlling for productivity, field, and rank, married mothers in science, engineering, and math made salaries similar to those of married fathers and substantially higher than those of single women (with or without children). Using 2008 SDR data, Kahn (2013) found that PhD women engineers working in tenure-track academic positions earned an insignificant 2% less than men on average, but single women (with no employment gaps) earned 4% more than men, indicating that there had been a small marriage/motherhood penalty. Children also make a difference in academic biomedicine, cutting salary differences by approximately half (Kahn & Ginther, 2012). This latter work on biomedicine also showed substantial salary differentials for single childless academics, as did Kelly and Grant (2012).

A related preference-based argument suggests that women may devote less time to work, and earn less as a result. However, Figure 15 shows that this is clearly not the case in most science fields. Taken as a whole, this literature suggests that when children lower publications or cause gaps in careers, they do seem to create a negative marriage/child effect on women's salaries.

Babcock and Laschever (2003) argued that sex differences in negotiation may generate observed differences in

compensation. They showed evidence that “women don’t ask” for or expect salaries that are as high as men’s. In academia, where salaries are negotiated, women’s not asking for or receiving comparable starting salaries can lead to large salary differences over careers. Furthermore, salaries may grow at different rates for women and men because of outside offers. For example, the Massachusetts Institute of Technology (MIT) Report of the School of Humanities, Arts, and Social Sciences (2002, page 18) found that at MIT, an increase in salary “partly depends on obtaining outside offers from other universities, and such outside offers have become increasingly important drivers at MIT over the last decade.” Thus, if women are less likely to seek or receive outside offers, their salaries may not grow as fast as those of their male colleagues. Blackaby, Booth, and Frank (2005) showed that even if women in economics do ask for salary adjustments, their increases may not be as high as those received by men. They found that after controlling for productivity, British women economists were less likely to get outside offers and got lower returns to outside offers than men in U.K. academia.

Institutional policies may lead to sex differences in salaries. One recent study showed that stopping the tenure clock had no significant effect on the probability of being promoted at a large public university, but it did significantly reduce salaries (Manchester, Leslie, & Kramer, 2013). Although there was no evidence of bias in the review process for journals, the same cannot be said of annual peer evaluations of salaries. Carlin, Kidd, Rooney, and Denton (2013) examined the productivity record and peer review of faculty at a public university. They found that men’s productivity, as measured by publications, earned higher salary increments than women’s productivity.

Taken together, these results present a mixed picture of the barriers to women’s progress in academic careers. Gender differences in promotion and salaries can largely be explained by observable characteristics, including productivity and field. However, in some cases, even single childless women continue to earn less than men in the same field and rank. Moreover, Ginther (2004), studying STEM fields as a whole through 2001, and Kahn and Ginther (2012), studying biomedicine, found that the gender gap in academic STEM salaries increases as careers unfold.³⁰ This leaves open the possibility that bias may be playing some role in the remaining gender salary gap found in some fields.

The Role of Women’s Choices to Opt Out

Two popular narratives have emerged related to professional women’s commitment to their careers: “opting out” (Belkin, 2003) and, more recently, “leaning in” (Sandberg,

2013). The “opting out” narrative discussed in Belkin (2003) holds that women cannot have both a family and a high-powered career. Professional women often partner with professional men and can afford to stay at home and take care of their families. Sandberg’s (2013) “lean in” narrative argues that women’s own choices restrain them from leadership roles and exhorts women to be more assertive about having both a satisfying career and a family life. We now consider evidence for opting out and leaning in in academic science.

We have presented evidence that women’s underrepresentation in math-intensive (GEEMP) fields starts very early, so that by high school, we see far fewer girls taking AP GEEMP courses while more girls than boys take AP LPS courses. In college, the preponderance of women in LPS rather than GEEMP fields persists. Yet post-bachelor’s, gender differences in attrition are much smaller in GEEMP than in LPS fields. Thus, the gender difference in proceeding from a GEEMP bachelor’s degree to a PhD has been shrinking and is now small, and women and men proceed to get tenure-track jobs and then to achieve tenure at an equal pace. In contrast, the gender difference in proceeding to a PhD is large and growing for women in LPS fields, and far fewer women move from PhD to tenure-track jobs, despite the large numbers of women faculty in LPS fields. The cause of this is not that women applicants are not being hired, but rather that they are choosing to opt out of academic science.

In life science particularly, women are able to opt out of academic science because there are multiple opportunities to do science in non-academic settings (Monosson, 2008). As of 2008, 55% of jobs for biomedical PhDs were outside of academia (NIH, 2012). Furthermore, a greater proportion of female scientists are married to male scientists than the reverse, and the difficulty in finding two jobs may cause women to opt for non-academic positions (Mavriplis et al., 2010).

So far in this review, we have skirted around the issue of work–life balance as the source of STEM pipeline leakage. However, this is the explicit or implicit reason that many believe keeps women from devoting themselves to a science career. In this section, we consider how family-related choices affect attrition from PhD to tenure-track academia and, more generally, the likelihood to leave science or the labor force.

Children and pipeline leakage from PhD to tenure track

A major source of attrition among STEM women occurs between the receipt of the PhD and the attainment of tenure-track positions in LPS fields and, to a lesser extent, in GEEMP fields. While outright biases in hiring and promotion in the past may have been a significant reason

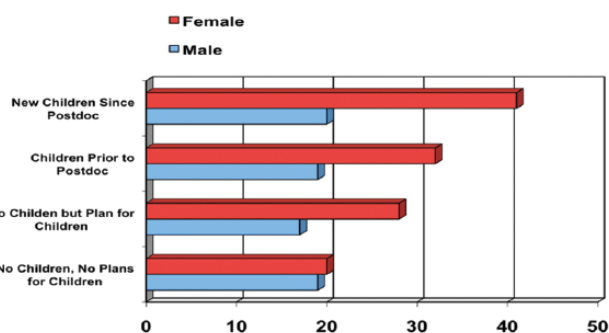


Fig. 18. Percentage of University of California postdocs who switched away from an emphasis on a career as a research professor as a function of presence of children and gender. Data shown here were drawn from Goulden, Frasch, and Mason (2009).

that many female PhDs did not apply for tenure-track positions, or applied but were not hired or promoted, the current most important barrier at this transition point, at least in statistical terms, is the perception among female PhD recipients and postdocs that these positions are not compatible with family formation.

Ginther and Kahn (2009) estimated the transition from PhD to tenure-track job separately by broad field and found that within the life sciences, married women and women with children were significantly less likely to transition to tenure-track jobs compared with single, childless women. Mason and her colleagues found that women PhDs with no children and no plans to have children fared as well as men in applying for and getting STEM tenure-track jobs, whereas those with plans to have children opted out of the R1 tenure-track pipeline in favor of careers they believed were more compatible with their plans, such as positions at teaching-intensive colleges or adjunct posts (Goulden, Frasch, & Mason, 2009; Wolfinger et al., 2008, 2009). To develop an idea of the magnitude of this factor, female postdocs in Mason et al.'s survey experienced over 50% more attrition if they planned to have families compared with men who planned to do so (28% vs. 16%), or if they already had children prior to the postdoctoral position (31% vs. 19%; see Fig. 17). Martinez and her colleagues (2007) found similar child-related attrition in a survey of 1,300 NIH postdocs. And Ecklund and Lincoln's (2011) survey of 3,455 biologists, astronomers, and physicists in top-20 departments found that four times as many female as male graduate students and 50% more female than male postdocs were worried that a science career would keep them from having a family. As the authors noted, "It is not surprising then that by the time they reach the postdoctoral level, women are much less likely than men to report considering a tenure-track job at a research university."

Why do children have more impact on obtaining a tenure-track job in life science than in GEEMP fields? The

answer is likely to lie in the postdoctoral position itself. As Kahn and Ginther (2012), Mason, Wolfinger, and Goulden (2013) and Monosson (2008) have pointed out, postdocs postpone getting started in biomedical careers. Moreover, postdocs in life science require long hours of work with little discretion over when those hours are, which, as Goldin (2014) pointed out, keeps women from vying for the most prestigious jobs across the spectrum of jobs in the U.S. labor market.

For those women who "lean in" to their academic careers, work-life balance poses significant challenges despite the widespread adoption of family-friendly policies in academia, including parental leave and the option to stop the tenure clock. Fox, Fonseca, and Bao (2011) surveyed STEM faculty at nine research universities between 2002 and 2004 to examine work/family conflict (whereby work interferes with family commitments) and family/work conflict (whereby family commitments interfere with work). Both women and men reported that work interfered with family more than family interfered with work, but that conflict was higher for women in both the work/family and family/work domains. Women's family/work conflict also increases with seniority.

Drago et al. (2006) surveyed faculty in English and chemistry and found that workplace norms in academia did not support family commitments. As a result, faculty women were more likely to stay single, to have fewer children, to have children after tenure, and to miss children's events in order to avoid perceived bias against caregiving. Ecklund and Lincoln (2011) found that among biologists, astronomers, and physicists in top-20 departments, roughly twice as many women as men claimed that career demands caused them to have fewer children than desired, and this was the only factor that was significantly associated with plans to seek a career outside science. Moreover, all of these studies may have underreported work/family conflict because individuals with the highest amount of conflict may have already opted out of academia.

Despite the significant work/family conflict, female faculty can and do become mothers. Ward and Wolf-Wendel (2012) interviewed 87 female faculty across a wide variety of disciplines and institution types in order to determine how academic mothers manage work and family demands. Among the STEM faculty interviewed, several common themes emerged. In particular, STEM academic mothers talked about being the only women in their department and being called upon to meet with students and do extra service. Ward and Wolf-Wendel (2012) noted that faculty members "were very aware of the extra work that comes with being the only woman, the only scientist, the only mother, and the only one for people to turn to for myriad activities" (p. 93).

Ward and Wolf-Wendel found that STEM faculty working in labs were more productive during family-related leaves of absence because the communal nature of lab work kept the research going. Compared with other academic mothers who took parental leave, women in laboratory science did not experience a slowdown in research productivity compared with women in other disciplines (Ward & Wolf-Wendel, 2012).

Family-friendly policies, including stopping the tenure clock and paid parental leave, have been adopted by universities in recognition of work/family conflict. Both the NIH and the NSF have instituted policies to promote career flexibility.³¹ However, these policies only assist with work-life balance if they are utilized by faculty. Drago et al. (2006) found that women faculty (in chemistry and English) engaged in "bias avoidance" to hide family commitments, such as by not taking parental leave or stopping the tenure clock. Lundquist, Misra, and O'Meara (2012) found that the overwhelming majority of faculty taking advantage of parental leave (72%) were women, but that only 26% of STEM faculty took leave. Rhoades and Rhoades (2012) surveyed 181 married tenure-track faculty working at institutions with paid parental-leave policies who had children under the age of 2. They found that 69% of women took parental leave but only 12% of men. Among those who took parental leave, men participated less in child care than women. Conversely, female faculty's husbands worked significantly more hours outside of the home than male faculty's wives. Thus, academic mothers work a "second shift" (Ward & Wolf-Wendel, 2012) even with parental-leave policies.

Also, the majority of family-friendly policies focus on the birth of the child and do not recognize faculty's need for work-life balance related to caring for children beyond infancy. In fact, Mason et al. (2013) argued for expanding family-friendly policies to include part-time tenured or tenure-track positions to meet these needs.

That said, research also suggests that family-friendly policies may have unintended consequences. Women in STEM disciplines may be less likely to stop the clock or take parental leave because the time off could lead to professional isolation or have a negative impact on research (Mavriplis et al., 2010). Manchester, Leslie, and Kramer (2013) found that faculty who took leave at a Midwestern research university were paid less.

One way to estimate the eventual impact of the relatively new family-friendly policies adopted by many U.S. universities on women's representation in tenure-track faculty is to look at Sweden and Finland, where family-friendly policies such as paid parental leave have been available to women for decades. However, Mayer and Tikka (2008) found that women in these countries have no higher representation in academia than in the United

States. In contrast, Mason et al. (2013) found that the adoption of family-friendly policies in the University of California system increased the number of tenure-track faculty having children. Thus, although family-friendly policies may help individual female faculty members achieve work-life balance, these studies indicate that they are not a panacea.

Job satisfaction

One reason that women may decide to opt out of academic science careers is their overall level of job satisfaction. The 1997 and 2010 SDR asked respondents whether they were very satisfied, somewhat satisfied, somewhat dissatisfied, or very dissatisfied with their jobs. We combined the satisfied categories for tenured and tenure-track academics and graphed the percentage satisfied in Figure 19. In 1997, women were less satisfied than men in all fields, although significantly less so in economics, engineering, and life science. In each of the natural sciences, these gaps narrowed to near zero by 2010. However, women in economics and social science were significantly less satisfied than their male colleagues. In the case of economics, the gap grew over time: Men were more satisfied and women less satisfied in 2010 than in 1997.

Others have found more widespread statistically significant gender differences in job satisfaction. Ecklund and Lincoln's (2011) survey found women faculty (in the fields of biology, astronomy, and physics) about 30% less satisfied with their careers than men in 2008 and 2009. In earlier work, Callister (2006) conducted a small study of 308 faculty members in science and engineering fields and found that women were significantly less satisfied with their work than men. She attributed these sex differences to negative departmental climates. In a larger survey, Trower and Bleak (2004) surveyed faculty at six research universities during 2002. They found that female faculty were less satisfied than their male colleagues with the climate of their departments. Of 28 areas probed, females were less satisfied in 19 of them, and in 9 there were no sex differences. In no area did males profess to be less satisfied than females. Specifically, females were significantly less satisfied than males with the commitment of their department chair and senior faculty to their success, and they were dissatisfied with their professional interactions with senior colleagues, leading to dissatisfaction with how well they fit in in their department.

Institutions that receive funding from the NSF's ADVANCE program are required to conduct climate surveys for STEM faculty. Part of the criteria for receiving ADVANCE funding is to demonstrate that women in science experience institutional barriers to career

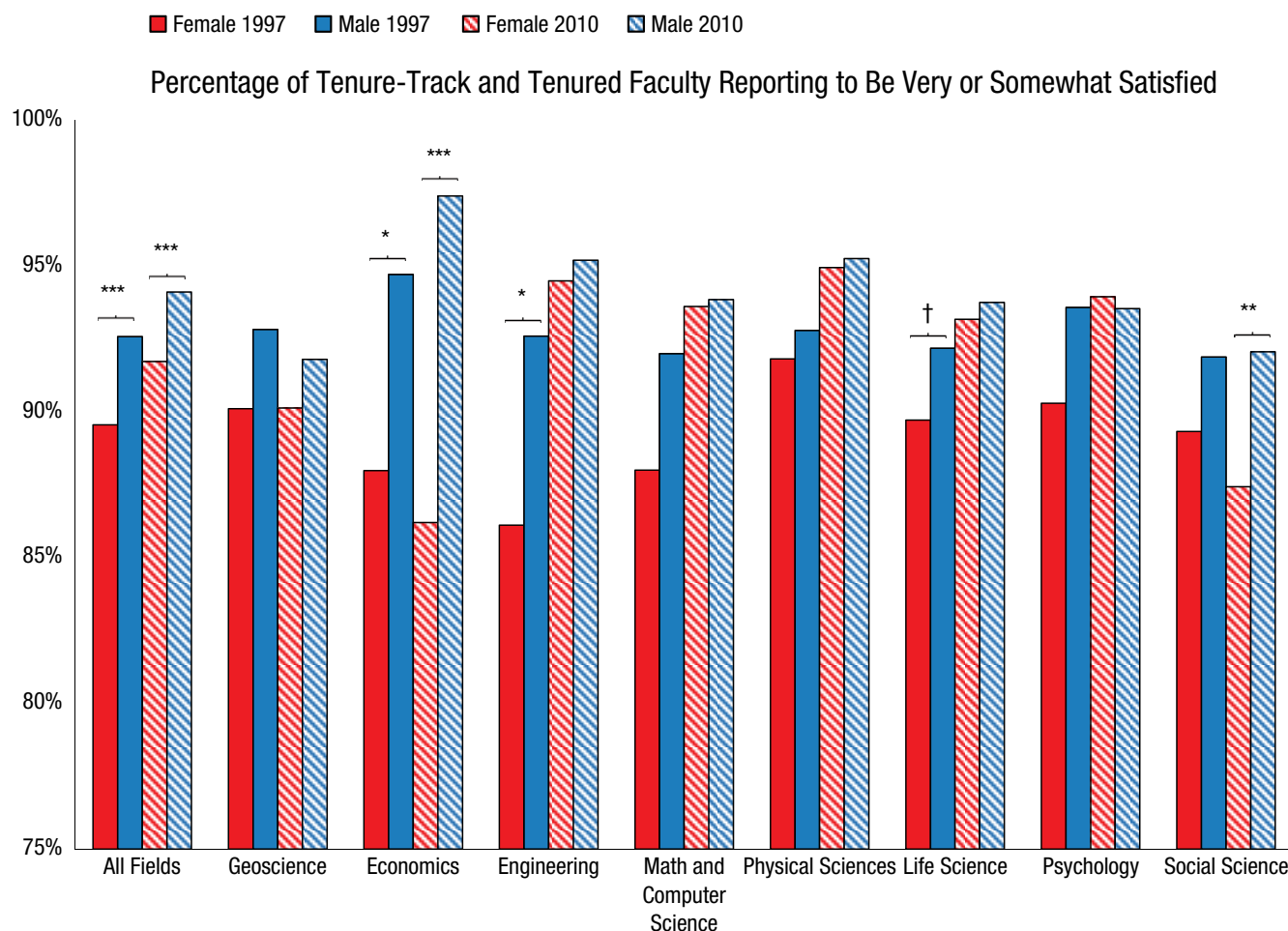


Fig. 19. Percentage of tenure-track and tenured faculty reporting to be "very satisfied" or "somewhat satisfied." Values shown are weighted percentages. Asterisks indicate significant differences between males and females ($^{\dagger}p < .10$, $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$). Data shown here were drawn from the National Science Foundation's Survey of Doctorate Recipients (<http://www.nsf.gov/statistics/srvydoctoratework/>).

advancement, so the results of these climate surveys may not generalize to all female STEM faculty. Nevertheless, the results suggest that institutional climate is associated with job satisfaction. Bilimoria, Joy, and Liang (2008) reviewed the results of climate surveys at six NSF ADVANCE grant institutions. They found that women STEM faculty were isolated and had fewer role models and lower job satisfaction. Settles, Cortina, Malley, and Stewart (2006) surveyed 208 STEM faculty at a Midwestern university in 2001. Perceptions of sexist behavior were associated with lower job satisfaction, and positive climate (measured by collaboration, cooperation, and collegiality) was associated with higher job satisfaction. Women in physical- and life-science fields reported more sexist behavior than did women in social science.

In sum, women faculty in STEM overall have been less satisfied with their jobs, with climate-related complaints being the primary cause. Figure 18, however, suggests

that this is no longer true in the natural sciences but remains—and has grown—in economics and social sciences. Low job satisfaction and climate affect women's choices to leave science altogether, as we show in the following section.

Leaving academic science, all science, and the labor market

Women with children face many obstacles in academic science, and women in some fields are unhappy in their academic jobs. Does this actually lead them to leave tenure-track jobs more than men? In the most comprehensive, and relatively recent, study, Xu (2008a, 2008b) used the 1999 NSOPF and found that women and men were equally likely to leave STEM academic careers. However, women were more likely to change academic jobs because of dissatisfaction with research support and advancement. Consistent with this, Callister's (2006) small

study of 308 found that women's lower satisfaction with their work made them more likely to quit faculty jobs than men.

The most dramatic form of opting out is leaving science altogether. Preston (1994, 2004) examined why scientists in general and women scientists in particular leave science (including leaving before obtaining their first job and leaving the labor market completely). These studies were based on data from the 1980s and 1990s and did not separate academics from other scientists. Although Preston found large differences in the 1980s, in the more recent analysis of the early 1990s she found that women and men PhDs were equally likely to leave science. Some of these differences were explained by exiting the labor force, but surprisingly, family characteristics did not explain female PhDs' greater likelihood of leaving science. Men were more likely to leave science because of unmet career and salary expectations, whereas women left because of work-life balance and lack of mentoring.

In a more recent study, Hunt (2010) used the 1993 and 2003 National Survey of College Graduates to examine why women with degrees in science leave science careers. Unlike previous studies in which women have been compared with men in science, Hunt compared women scientists to women in nonscience fields. She found that women are more likely to leave engineering but not science careers. Women engineers leave their careers because of pay and promotion concerns. She also found that the higher the concentration of men in the field, the more likely women are to exit it. Hunt concluded that a lack of mentoring, a lack of networks, and possible discrimination may play a role in women's exit from engineering careers. However, it is important to note that her study included scientists and engineers at all levels of education and was not limited to academic science.

We have used the SDR to determine whether there are significant sex differences in working in a field that is unrelated to the PhD degree, our measure of leaving science. In 2010 more men than women with STEM PhDs were working in jobs unrelated to their degrees: 8.4% compared with 7.5% ($p < .05$). When we compared the reasons for working outside of their field, job-related reasons dominated for both sexes ($p < .001$): 94.9% of men and 85.2% of women attributed their decision to work outside their field to issues such as pay, promotion, and location. Another way to interpret these same numbers, however, is that women were almost three times more likely to leave their science field because of family concerns (14.8% vs. 5.1%).

On the other hand, women are more likely than men to leave the labor force entirely. Analysis of the 2010 SDR indicates that women were more than twice as likely as

men to have left the labor force—either from a job or directly after obtaining their PhD (7.8% vs. 3.8%; $p < .0001$). The most frequently reported reason for leaving the labor force for both sexes was retirement (41.1% of women leavers and 75.8% of men leavers). Only 0.9% of men, but 4.6% of women, left the labor force for reasons other than retirement. The majority of these women (61.5%) reported leaving because of family considerations, compared with only 29.6% of the few non-retiring men. Thus, women scientists are more likely than men to leave the labor force and much more likely to leave because of family considerations.

Combining those who leave science to take another job or to leave the labor force (excluding those who retire), women are indeed more likely than men to leave science (12.5% vs. 9.6%).³² However, PhD women who continue working are less likely than men to be in tenure-track academia but are also less likely than men to opt out of science altogether.

Conclusions: Refocusing Today's Debate on Women in Science

In this final section, we present our conclusions in light of the best currently available empirical data to provide an empirically informed understanding of the causes for women's underrepresentation in academic science and how best to address these current causes. Our hope is that this research synthesizes, coupled with the numerous new analyses we have provided in this article, will help to redirect the debate toward critical issues that are most important in limiting the careers of women scientists *today*, and hopefully move closer to solving them.

Claims that biases against female scientists have been remedied, if untrue, can have serious negative societal consequences. Such false claims can lull policymakers into an unwarranted sense of complacency; they can be invoked by those desiring to terminate important programs and interventions. On the other hand, claims that bias against female scientists is a primary explanation for their underrepresentation, if untrue, can also be detrimental to women scientists—by diverting resources from needed interventions and directing these resources to addressing former barriers that no longer explain women's underrepresentation.

We began this article by noting how rapidly women have increased their representation in all STEM disciplines. In the 1970s, women received less than half of all STEM degrees. Since that time, progress has been uneven. In the case of LPS fields, women have attained a critical mass, sometimes comprising over 50% of assistant professors and a quarter to a third of full professors. In contrast, women are in shorter supply in GEEMP fields. Women comprised about 10% of bachelor's-degree

recipients in these fields (in engineering, only 1%) in the 1970s and are now receiving between 20% and 40% of bachelor's degrees. The most recent figures indicate that in these fields, women comprise only 25% to 44% of tenure-track assistant professors and only 7% to 16% of full professors. The goal of this article was to explore and explain the basis of this difference.

Myriad causes have been alleged to explain women's underrepresentation in GEEMP disciplines—chilly climate, biased interviewing and hiring, lack of female role models, lack of mentors, biased tenure and promotion, unfair salary, sex differences in quantitative and spatial abilities, lower productivity and impact, stereotype threat, and sex differences in career preferences. So what explanation (or explanations) do the data support? In Table 5, we summarize the vast body of literature that was part of our synthesis.

We begin with a seeming paradox. The fields in which women are in shortest supply are (with the exception of economics) the very fields in which they appear to have the most success at being hired, promoted, and remunerated as professors. Recall that it is in GEEMP fields (often with the exception of economics) that women and men proceed from college major to PhD in equal proportions, in which women applicants are invited to interview and are hired at higher rates than their male colleagues, and in which women are tenured, promoted, and remunerated comparably to their male colleagues. And it is in GEEMP fields that women persist in their careers as long as their male peers—with the occasional exception. But we want to underscore that the exceptions *are* occasional, and that the overall picture is one of gender neutrality in GEEMP fields, notwithstanding frequent claims to the contrary. Thus, the paradox: Why are women in shortest supply in the very fields in which they appear to fare best?

Our analyses and research synthesis led us to dismiss as important causes of women's underrepresentation pay or citations per published article, both of which were equivalent (with some exceptions) for the two sexes. Our analysis also ruled out discriminatory grant and journal reviewing and biased hiring and promotion decisions, none of which have been consistently demonstrated to occur. These factors are not related to women's underrepresentation in academic science careers, which has led us to the conclusion that the overall state of the academy is largely one of gender neutrality. There are some important ways in which women and men differ, and some of these may be related to differential outcomes by sex. For example, our research shows that women on average publish fewer papers than men (although their citations per published paper are the same). Given the central importance of publications to the progression of academic careers—in fact, of all variables, it is

publications that are often argued to be the single most important measure of academic success—women's lower productivity in publishing may be seen as a key variable in some differential outcomes. Furthermore, there is no evidence that women's relatively fewer publications are higher in quality and impact. Better data on individual publications and citations, and more research in general, is required to determine whether differences in productivity (quantity and quality) account for some of the observed differences in salaries and promotion.

Given that the factors just discussed do not explain the gender gap in math-intensive GEEMP fields, what does? This long list of exclusions still leaves occupational preferences (women are more likely to prefer organic fields that involve living things, whereas men tend to prefer fields that emphasize symbol manipulation), and the roots of these differences can be seen in the type of AP coursework that high school students take as well as in their choice of college majors.

Also, the list of potential causes still leaves open the key issue of the impact of children. However, why would the presence of children reduce a woman's likelihood of entering GEEMP fields and not, say, life science, in which the temporal inflexibility of lab work would seem to be at least as great? Ginther and Kahn (2009) found that family variables did not dissuade women from pursuing academic careers in the physical sciences and engineering. But that study was based on data through 2001, and given the dramatic changes we have documented here, it is entirely possible that as more women have entered GEEMP fields, family variables may now have begun to influence career decisions. As we have suggested elsewhere (W. M. Williams & Ceci, 2012), childrearing may have adverse effects on women's careers in all fields, but these effects might be more apparent in GEEMP fields because women's numbers are relatively low to begin with. That is, if the plan to have children dissuades some female PhD students or postdocs from entering the competition for tenure-track jobs in psychology or biology (the two fields in which the transition showed the largest loss of females), this effect will be less apparent because women currently constitute 40% to 67% of assistant professors in these fields. In contrast, the reduction of women resulting from a similar decision not to enter tenure-track job competition in engineering will exacerbate an already-existing dearth of women. Thus, the negative impact of children on the work lives of women is foregrounded in fields in which they are relatively sparse to begin with.

This still leaves as potential causes of underrepresentation sex differences in occupational preferences that are evident by middle school and that result in different high school course choices (i.e., accelerated and AP coursework in physics, Calculus BC, and chemistry).

Table 5. Narrative Summary of the Large Body of Literature and Our Own Analyses That Were Synthesized in This Review

Variable	LPS	GEEMP
Math/spatial aptitude	No evidence that a math/spatial gap affects women's later representation.	Strong evidence that males are overrepresented at the right tail, although the extent is mutable and differences are probably narrowing; not clear if and how much this affects women's later entry into GEEMP fields.
High school STEM interest	No evidence of sex differences; females prefer LPS careers.	Surveys indicate wide gender differences in plans/expectations, especially from middle school. Females take fewer Advanced Placement science/Calculus-BC coursework and show less interest in GEEMP-based careers.
College major	For roughly 20 years, females have been well represented in LPS majors, often in the majority.	Strong evidence that females earn 40% or less of baccalaureates in GEEMP fields, despite growth in recent decades.
PhD programs	Fewer females than males transition from majors to PhDs.	The percentage of bachelors who transition to PhDs is equivalent for both sexes.
Biased interviewing/hiring	A lower percentage of female PhDs apply for tenure-track posts. Track posts less often than men, but good evidence that they are offered these posts more often, both in real-world hiring and in experimental studies.	There is no evidence that women applicants are hired for tenure-track posts more often, both in real-world hiring and in experimental studies.
Publications	Men are more productive in journal publishing.	This gap narrowed through mid-'90s but then plateaued.
Citations	Although women publish fewer papers, there are no sex differences in citations per article.	
Grants	Women apply for fewer grants, particularly follow-up ones. Women are less likely to receive follow-up grants at the National Institutes of Health. Men on average get larger grants because they are principal investigators on larger projects.	
Biased work evaluation	Editors and grant reviewers rate both sexes equivalently.	
Tenure and promotion	Some evidence of women having harder time getting tenure in biology and psychology.	No evidence of women having harder time getting tenure in any GEEMP field but economics.
Salary	Mixed evidence—women in life science are paid less than men at the assistant-professor rank. The presence of children explains half of this gap. Others are paid equivalently.	Among assistant and full professors in economics and full professors in geoscience, engineering, and the physical sciences, women are paid less than men. Others are paid equivalently.
Opting out	Fewer female PhDs apply for tenure-track positions. In life science, more women leave between PhD and obtaining a tenure-track position.	There are no sex differences in the likelihood of entering a tenure-track job or in leaving academic science. However, women are more likely to leave non-academic engineering careers.
Job satisfaction	Women are less satisfied in social science. Recently, women and men are similar in other LPS fields.	Women are less satisfied in economics. Recently, women and men are similar in other GEEMP fields.
Children and opting out	In life science and social science, women are more likely to leave between PhD and tenure-track positions if they are married and/or have children.	No evidence that marriage or children deter women from entering tenure-track jobs in the physical sciences or engineering.

Females profess to be more interested in medicine, biology, law, psychology, and veterinary medicine from a relatively young age, whereas males are more likely to prefer engineering and computer science. Perhaps it should not be surprising to discover that the sexes trod divergent paths after all.

Our review of the evidence leaves open one possibility raised recently by Goldin (2014) in her presidential address to the American Economic Association. Her analysis revealed the premium that women disproportionately place on flexible work conditions, which in turn results in lower wages and promotion, but only in some fields. Her analyses, model, and arguments suggest that women's status in science is the result of personal choices and time-flexibility preferences, as opposed to sex differences in human capital and sex-based salary discrimination (except insofar as one can argue that companies are implicitly biased for not making flexibility available at a lower cost to those desiring it). Her analyses explain why sex differences in salaries usually begin very small, grow over time, and manifest nonlinearly (with hours worked) in some fields but not in others. Goldin has shown that if women place a premium on flexibility, then purchasing that amenity can be quite costly in some fields because of the nature of the work and the size of typical workplaces (e.g., for MBAs in finance or Juris Doctors), but not in others (e.g., in pharmacy or large practices in which workers are interchangeable). Thus, she has brought personal choice, a factor that has been derided by some gender-equity advocates, back into the picture.

In a related vein, Ferriman et al. (2009) showed that men and women from top graduate programs in STEM fields expressed quite similar preferences for flexibility in their future work schedules and for working less than 50 (or 60) hours per week while they were still in graduate school, but by their mid-30s, the women with children were much less like men (with or without children) and women without children with respect to preferring to avoid long hours and wanting flexible schedules. Further, Lubinski and Benbow (2006) found in both this same sample of ex-graduate students as well as a sample of people who had been exceptionally able in mathematics as adolescents (also in their mid-30s) that women not only were more likely to prefer to work fewer hours but actually did so. Relatedly, in graduate school, both sexes actually reported devoting 20 hours a week to studying and 30 hours a week to research: no sex differences (see Lubinski et al., 2001, Table 3, p. 315). Time flexibility is thus a priority for at least a portion of women who excel in science. And, as we argued earlier, it seems that these women often eschew faculty jobs in order to accommodate these preferences. Looking at current statistics (from

the SDR), we found that women with LPS PhDs were the ones least likely to pursue tenure-track positions and most likely to have shorter hours in their work outside tenure-track academia. This suggests that Mason et al.'s (2013) policy prescription of part-time tenure may serve to increase the number of female faculty.

Another way in which our review of the evidence is limited is by what is not available in the scientific literature and by the adequacy of available evidence. At a number of junctures we presented survey data and anecdotal claims because they were the only evidence available. In discussing sex differences in job satisfaction and climate, for example, we presented claims by women from surveys or interviews stating that they were on average less satisfied with their work climate or jobs than men. However, the meaning of such self-reports is unclear. Is women's self-reported lower level of satisfaction due to salary and promotion factors, or is women's relatively lower reported level of satisfaction due to men's socialization inhibiting them from readily disclosing (or even acknowledging) unhappiness with their jobs, potentially because such acknowledgement signals weakness (men have been shown to underreport pain, for example—Ellermeier & Westphal, 1995; Fillingim, King, Ribeiro-Dasilva, Rahim-Williams, & Riley, 2009)? That is, is women's lower satisfaction due to their greater willingness to label themselves as dissatisfied? And what are the measurable outcomes of this gap in satisfaction between men and women (when it exists, given that it is not consistent)? We do not know.

Currently, we lack outcome measures to validate such self-reports (e.g., does lower satisfaction predict leaving the scientific workforce, or is it tied to productivity differences?). It is particularly important to investigate the meaning of self-reports when they diverge from the large literatures showing that, with few exceptions, men and women fare similarly in grant and journal reviews, hiring, persistence, and promotion rates. Furthermore, it is important not to ante up "studies" on one side or the other of an issue, equating the validity and meaning of different types of evidence. For example, an objective analysis of sex differences in grant awards based on 100,000 grant applications should not be offset by a self-reported survey on job satisfaction at a few universities.

In sum, depending on the life-course transition point, the cause of early lack of interest in GEEMP subjects and later attrition from GEEMP fields is the result of one or more of a confluence of variables. Attempts to reduce these causes to a single "culprit" (e.g., bias by search committees against female applicants; women's preference for other fields or lack of math aptitude; publication rates; salary differentials) are not supported by the

full corpus of data and research findings. Granted, one can cherry-pick aberrant examples that seem to suggest bias or aptitude gaps or differences between the sexes in productivity or impact, but the entire scientific corpus reveals that no single cause can account for the dearth of women in GEEMP careers. The most significant implication of our analysis is that failure to acknowledge the nature, complexity, and timing of causes limits progress in increasing women's representation in math-intensive careers, by directing resources to areas that are not currently major reasons for the dearth of women in math-intensive fields. It is our hope that we have helped move the debate from slogans and rallying cries to a judicious consideration of the full corpus of scientific data. This may make it harder to make sweeping indictments, but that is a price worth paying for scientific accuracy.

Finally, we resort to the hackneyed convention that "future research is needed," because it is true. There were many junctures in this analysis at which it was not possible to narrow a list of potential causes down to a single cause. For example, the question of whether differences in perceived math ability and beliefs about whether most people can be good at math influenced high school students' decisions to declare and pursue GEEMP majors remains open. Much more research will be needed in the coming years to narrow the many uncertainties unveiled in this report. It is our belief that we have advanced the debate by ruling out many dead ends and shining a spotlight on areas in need of future attention, rather than continuing to pursue issues that may have been historically relevant but no longer predict women's underrepresentation.

Appendix

A. Data sources

We used several data sources to provide a statistical portrait of women in academic science. We describe these sources below.

Survey of Doctorate Recipients. The Survey of Doctorate Recipients (SDR; www.nsf.gov/statistics/srvydocto) is a biennial longitudinal survey of doctorate recipients from U.S. institutions conducted by the National Science Foundation (NSF). The SDR's respondents are drawn from the Survey of Earned Doctorates, the NSF's annual census of doctorates awarded in the United States. The SDR collects detailed information on doctorate recipients, including demographic characteristics, educational background, time use, employer characteristics, and salary. We have used repeated cross-sections of data to show sex differences in reported outcomes. When possible, we have used the latest data available (2010) and

compared it with data from 1995. The SDR collects information on publications only periodically, so we have used the latest survey with those questions (2008) and compared it with that from 1995. Data on job satisfaction was collected in 2010 and in 1997 (the most proximate year to 1995).

WebCASPAR. The WebCASPAR database (<https://ncesdata.nsf.gov/webcaspar/>) provides access to statistical-data resources in science and engineering at U.S. academic institutions. We used WebCASPAR to tabulate data from the National Center for Education Statistics's Integrated Postsecondary Education Data System on bachelor's-degree, master's-degree, and PhD recipients at U.S. colleges and universities by sex and academic field.

Current Population Survey. The Current Population Survey (CPS; <http://www.bls.gov/cps/#data>) is collected monthly by the U.S. Census Bureau on behalf of the U.S. Bureau of Labor Statistics, and it is the primary source of labor-force statistics for the population of the United States. We used the Current Population Survey to count the number and percentage female high school graduates aged 20 to 25 from 1992 to 2011.

American Community Survey. The American Community Survey (ACS; <https://www.census.gov/acs/www/>) is collected annually by the U.S. Census Bureau. Recent waves of the ACS have collected information on the undergraduate major of respondents. We used 2012 ACS data to examine the probability that individuals who majored in science obtained higher degrees within 10 years of receiving their bachelor's degree and to examine whether these degree holders were working in science occupations.

B. Sex differences in mathematics and computer-science degree attainment

In Figures 3a and 3b in the main text, the field of mathematics/computer science shows a large decrease in the percentage of bachelor's and PhD degrees awarded to women. In this appendix, we investigate these trends in greater detail by disaggregating the fields into mathematics and statistics on the one hand and computer science on the other. Figures A1a and A1b show the percentage of bachelor's and PhD degrees awarded to females in the fields of computer science, economics, engineering, and mathematics and statistics. The drop in the percentage of bachelor's degrees awarded to females in mathematics/computer science is entirely explained by the drop in degrees awarded to females in computer science. Women

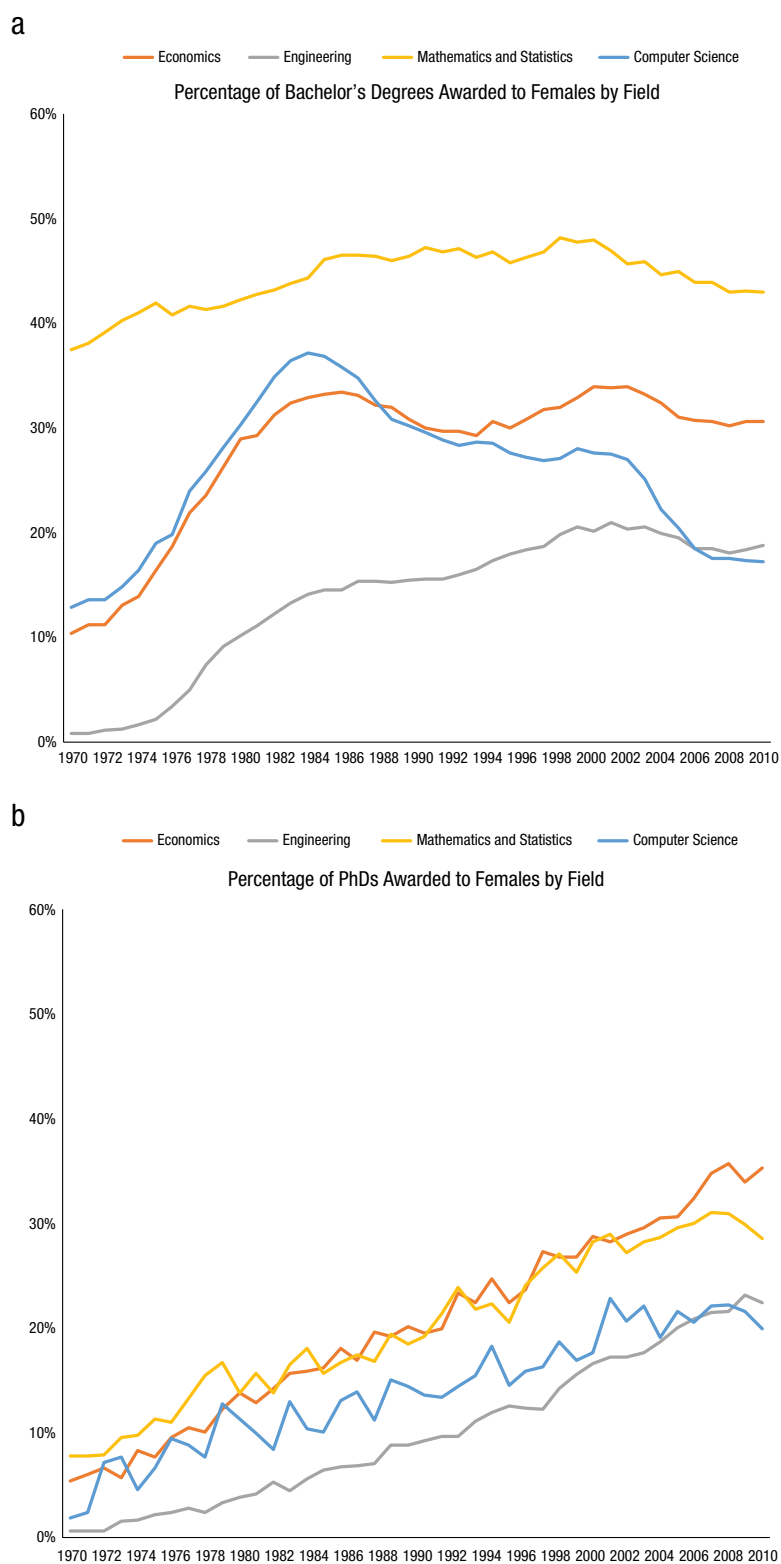


Fig. A1. Percentage of bachelor's degrees (a) and PhDs (b) awarded to females from 1970 to 2010 as a function of field. Data shown here were drawn from the National Science Foundation's WebCASPAR database (<https://ncesdata.nsf.gov/webcaspar/>).

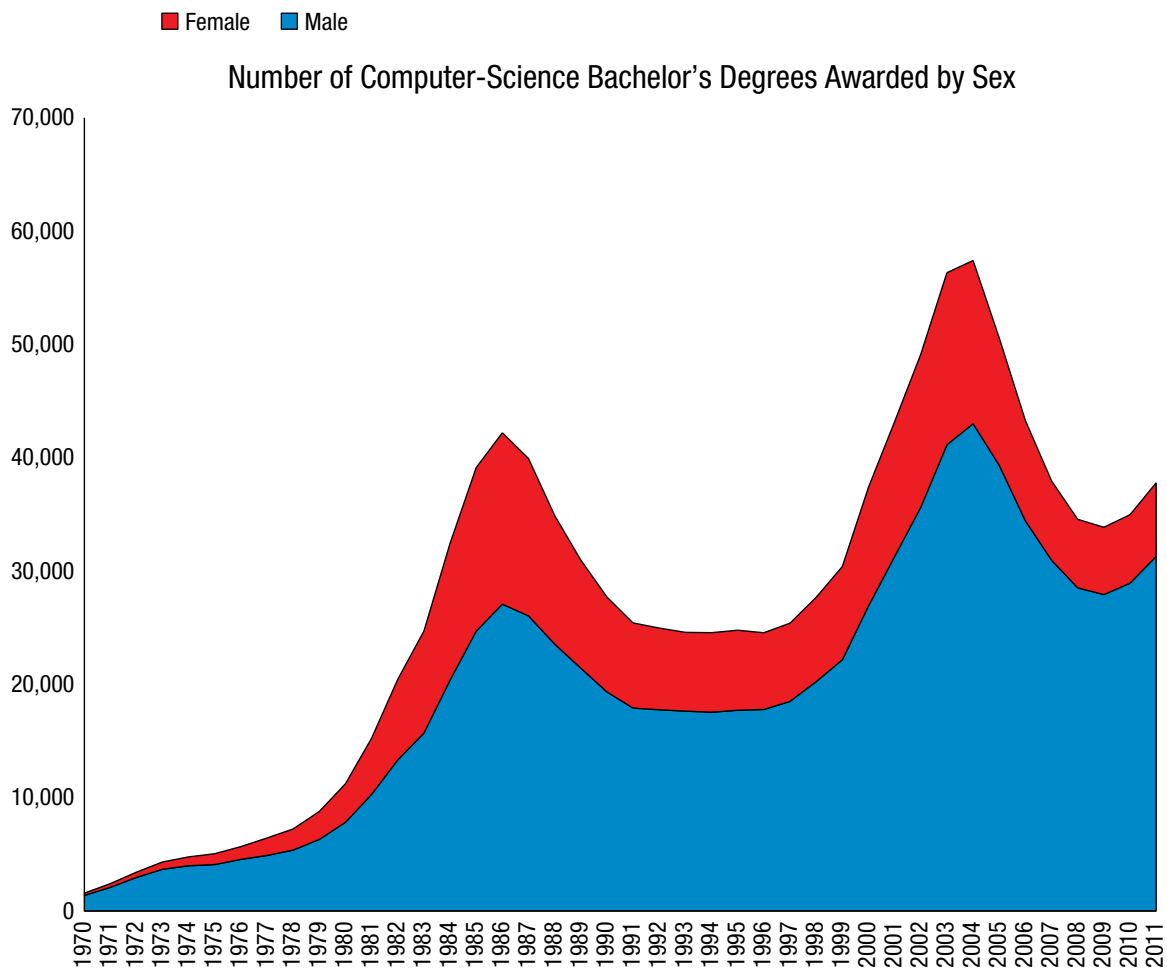


Fig. A2. Number of computer-science bachelor's degrees awarded from 1970 to 2011 as a function of sex.

as a percentage of computer-science majors peaked in the mid-1980s at 37% and decreased to less than 20% by 2011, the lowest percentage in any field considered. In contrast, mathematics and statistics held steady, with percentages of degrees awarded to females ranging between 43% and 48% since the mid-1980s. Figure A1b shows that the percentage of doctorates awarded to females grew in these four fields between the 1970s and 2011, but the growth rate was slowest in computer science.

We also investigated whether changes in the numbers of men and women obtaining bachelor's degrees in computer science could explain the observed differences (see Fig. A2). Two trends are immediately apparent. First, computer-science degrees exhibit a good deal of cyclicity, peaking in the mid-1980s and again in the early 2000s. Second, men have increased their relative share as computer-science majors, such that the number of male majors grew by 59% between 1986 and 2004 (the 2 peak years) whereas the number of female majors fell by 5% in the same periods. When we compare the

trough years of 1994 to 2011, male computer-science majors increased by 78%, whereas female computer-science majors decreased by 5%. Thus, much of the difference in percentage of bachelor's degrees awarded to females in computer science can be explained by the increase in male majors and a smaller decline in female majors.

C. Sex differences in publications, conference papers, and patents

In addition to publications, we used the 2008 SDR to examine sex differences in academic productivity in Table A1. Besides the publication results in the main text, we observe some notable sex differences in productivity in selected fields. In engineering and the physical sciences, males present more conference papers than females. In fields in which patenting is prevalent (engineering, life science, and the physical sciences), male faculty apply for and receive more patents than female

Table A1. Gender Differences in Academic Productivity From 2003 Through 2008 by Field

Institutions and field	Gender	Publications	Conference papers	Patent applications	Patents granted
All universities					
Economics	Male	6.07*	8.97	0.00	0.00
	Female	4.15	7.33	0.10	0.02
Engineering	Male	12.38 [†]	19.94**	1.34**	0.70**
	Female	9.81	15.70	0.50	0.30
Geoscience	Male	11.19	18.82	0.04	0.01
	Female	8.78	17.36	0.08	0.04
Life science	Male	13.33***	14.73	0.62**	0.34**
	Female	10.38	13.16	0.27	0.16
Mathematics/ computer science	Male	6.09	9.09	0.16	0.11
	Female	5.99	9.29	0.19	0.09
Physical sciences	Male	14.75***	15.72***	0.85*	0.57*
	Female	7.93	10.02	0.33	0.40
Psychology	Male	11.06**	14.79	0.01 [†]	0.01 [†]
	Female	7.95	14.12	0.00	0.00
Social science	Male	5.85	9.97	0.01	0.01
	Female	5.12	9.65	0.00	0.00
Research I universities					
Economics	Male	8.56*	12.07	0.00	0.00
	Female	5.71	9.93	0.28	0.05
Engineering	Male	18.19	26.27*	1.93**	1.00**
	Female	14.21	20.04	0.59	0.33
Geoscience	Male	17.91 [†]	26.36	0.08	0.02
	Female	12.86	20.75	0.30	0.08
Life science	Male	17.29	17.12	1.05*	0.57*
	Female	15.14	18.26	0.48	0.35
Mathematics/ computer science	Male	8.44	12.55	0.36	0.30
	Female	9.54	13.61	0.47	0.20
Physical sciences	Male	25.17 [†]	25.38	1.69 [†]	1.06 [†]
	Female	19.55	20.35	0.60	1.43
Psychology	Male	19.33*	21.89	0.02	0.02
	Female	13.11	19.47	0.00	0.00
Social science	Male	9.71 [†]	14.88 [†]	0.00	0.00
	Female	7.36	12.61	0.00	0.00

Note: Data shown here were drawn from the National Science Foundation's Survey of Doctorate Recipients (<http://www.nsf.gov/statistics/srvydoctoratework/>).

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

faculty. As would be expected, sex differences are somewhat less prevalent among faculty at research-intensive institutions.

D. Hours of work by presence of children

In the main text, we report results showing that male and female faculty with children work fewer hours (see Fig. A3). However we did not observe any sex differences in work hours for faculty with children. The one exception

is that men with children work on average 3 hours more than women in the physical sciences ($p < .10$).

Acknowledgments

The use of National Science Foundation (NSF) data does not imply NSF endorsement of the research, research methods, or conclusions contained in this report. We thank Dasha Milakhina, Pat Oslund, and Carlos Zambrana for their assistance with the data and figures for this paper. We also thank Lisa Wolf-Wendel for insightful comments. Data for figures are archived and available at the second author's University of Kansas web page: people.ku.edu/~dginther. Raw data for Williams and Ceci study of

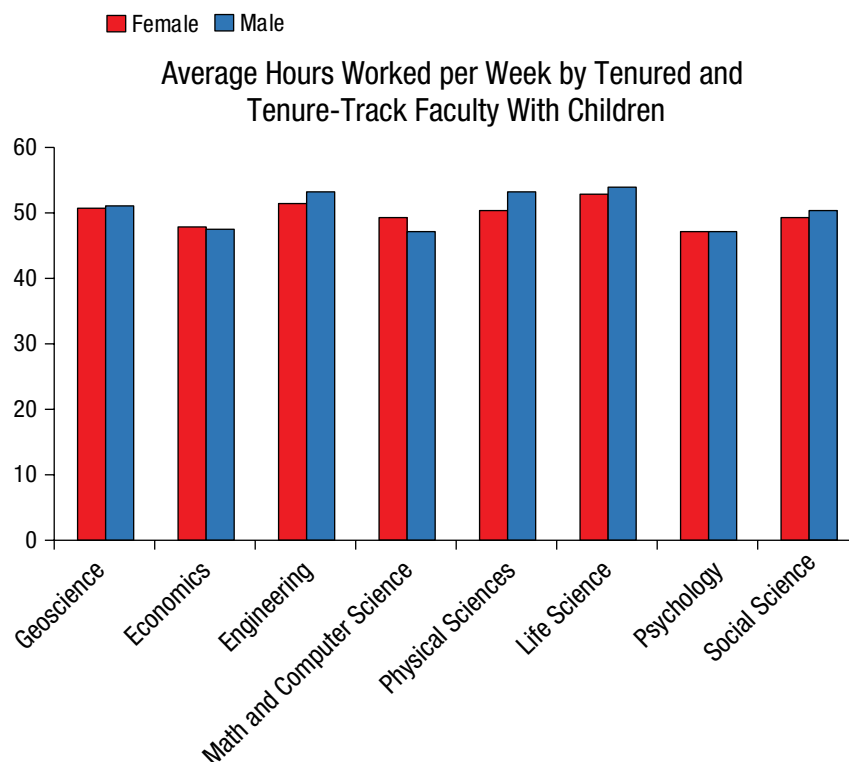


Fig. A3. Average number of hours worked per week by academic faculty as a function of field and sex.

hiring (W. M. Williams & Ceci, 2014) is archived at the Cornell Institute for Women in Science Web page www.ciws.cornell.edu.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

Portions of this work were supported by National Institutes of Health Grant 1 R01-NS069792-01 to W. M. Williams and S. J. Ceci. D. K. Ginther and S. Kahn acknowledge financial support from National Institutes of Health Grant R01-AG036820.

Notes

1. We focus on math-intensive fields because of the oft-heard argument that women do not enter science because of lower math ability. We categorized math-intensive fields on the basis of the mean GRE quantitative scores of graduates students in each field.
2. In this article, we use the terms “gender” and “sex” interchangeably, not reserving the latter for exclusively biological contexts.
3. We will not address issues related to women of color or immigrant women in science because race/ethnicity and

nativity present additional academic career challenges for women (National Research Council, 2013).

4. In the appendix, we disaggregate the fields of mathematics and computer science and discuss these different trends for bachelors and PhDs.
5. Recently, Valian (2014) has argued that despite the fact that these two dimensions produce empirical clusters, they are not conceptually coherent, with the three “thing” components (Realistic, Conventional, and Investigative) representing conceptually heterogeneous attributes and, similarly, the three “people” components (Social, Enterprising, Artistic) being conceptually heterogeneous. Thus, she argues that the people-thing distinction may not distinguish exclusively between people versus thing orientations. Her argument combines the Social-Enterprising-Artistic themes of the Holland codes and contrasts them with the Realistic-Conventional-Investigative themes. Her point about the measurement of occupational performance and the malleability of occupational choices is important to bear in mind with regard to the implications of the people-thing discussion. However, the people-thing dimension that runs from the so-called Realistic component to the Social one (the other components are not involved) leads to the conclusion that this dimension reflects a psychologically significant parameter of human individuality. Barring empirical evidence for a lack of sex differences in these components that is predictive of occupational choices, or evidence that the two components (Realistic-Social) are

more relevant for males but orthogonal for females, the people-thing dimension seems pertinent to maintain.

6. It is surprising, in view of the extensive documentation and effect sizes involved, that researchers are sometimes dismissive of this construct (e.g., Morgan et al., 2013).
7. This synopsis is necessarily highly abbreviated and does not discuss the myriad complexities that have prodded hormone researchers to repeatedly qualify earlier positions. Initially, animal studies appeared to suggest a simpler, straightforward picture of male hormones resulting in subsequent male-typed behavior. For example, male rats perform better than female rats on some spatial tasks, but female rats exposed to higher levels of androgen or its metabolites during early development show improved performance, and conversely, males exposed to lower levels of androgens display reduced performance (C. Williams & Meck, 1991). Subsequently, researchers found a U-shaped function wherein male hormones enhanced spatial cognition to a point, but beyond this point, they actually decreased performance in males.
8. See Miller and Halpern (2014) for examples of performance on spatial tasks that varies depending on whether the tasks are speeded versus non-speeded or involve transformations of rigid surface versus nonrigid surfaces.
9. The effect size, d , is the mean for males minus the mean for females, divided by the pooled within-gender standard deviation.
10. Xie and Shauman defined science and engineering as the fields of life science, physical science (including geoscience), mathematics/computer science, and engineering.
11. The economist Lorne Carmichael (2005) made this point by analogy:

The overall conditions for producing softwood in the Southeastern United States are much better than they are in Northern Ontario. So why does Canada export lumber to the United States when the Americans could produce it faster and better? The answer [is that] countries specialize in producing goods for which they have a “comparative” advantage. By this logic, the reason that land in Northern Ontario is devoted to the growing of trees is that there is precious little else that can be done with it. In Georgia the land can be used for many other things; golf courses, peach orchards, and stately homes made from cheap imported Canadian lumber. In discussions of the low number of women choosing to study Science and Engineering, much is often made of the fact that women do just as well in their high school science and mathematics courses as do men. But the data are just as clear on another fact—women on average do much better in English, and indeed in every class other than science and math. And high school students must choose what to do with their time just as countries must choose what to do with their land. . . . When we consider overall performance in high school, the question about enrollments in Engineering is not: “Why are there so few women?” The real question is rhetorical: “What else are boys going to do?” (p. 6)

12. Note that this is a different measure than reported above on STEM only. The male-to-female ratio of high school

students with occupational plans in STEM fields requiring a doctor of medicine is 2:1, compared with 4:1 for STEM fields alone.

13. Note that Xie and Shauman’s (2003) estimates of the percentage of students who persisted in STEM fields were much lower than those of Morgan et al. (2013), perhaps because Xie and Shauman used completed STEM majors (whereas Morgan et al. used *declared* STEM majors), because they studied a cohort 20 years earlier, or because they defined “science” differently.
14. “PEMC,” in their terminology, for physical sciences, engineering, mathematics, and computer science (Perez-Felkner et al., 2012).
15. Note that this difference in female persistence was not divided cleanly between GEEMP and LPS fields.
16. As a result, female students were *less* likely to persist in a science major than males (48% vs. 66%).
17. These percentages do not sum to 100% because they exclude those presently enrolled in a graduate or professional degree (these numbers were similar across sexes but not differentiated by type of degree).
18. Consider: a blogger at *Science* magazine asserted:

Given qualified women drop out of math-intensive fields at higher rates than their male peers . . . the women who remain are probably, on average, better than their male colleagues and should be having better (hiring) outcomes on average. If their salaries, resources, publication rates, etc. are similar, it then indicates gender discrimination still exists, not that this problem has been solved. (<http://blogs.sciencemag.org/sciencecareers/2011/02/the-real-cause.html>; retrieved on June 22, 2014)

A commentator on another blog post, speaking of her “experience at a national lab,” wrote:

Female scientists were either not retained or not hired so that only a couple of super-brilliant female scientists were working in staff-scientist positions. On the other hand, several mediocre male scientists were hired and retained, many rising to staff-scientist positions or higher. If you compare these super-brilliant female scientists with their mediocre male counterparts, of course you will not see the difference in their treatment. (Kali, 2011)

In a critique posted on February 15, 2012, Cathy Kessel argued that claims of gender equivalence were misleading because they overlooked the possibility that women’s scholarship is better than men’s. In her words:

The studies [claiming gender neutrality] examined odds ratios rather than details of the proposals submitted. This does not rule out the possibility of gender bias. As Marie Vitulli and I said in 2011 [Kessel & Vitulli, 2011], “selection bias can also explain why, in the presence of gender discrimination, female scientists might still fare as well as their male colleagues in some respects if their work was better on average than that of their male peers.” (Kessel, 2012)

19. The embargo policy of the journal to which this report has been submitted prohibits our discussion of these findings before they are published.

20. That is, they have successfully completed doctoral programs, garnered publications and glowing letters of reference, and been rated by the hypothetical faculty as "excellent" to "exceptional."
21. This includes results from the Eigenfactor Project (<http://chronicle.com/article/Woman-as-Academic-Authors/135192>; Symonds, Gemmell, Braisher, Gorringer, & Elgar, 2006).
22. For instance, Bellas and Toutkoushian (1999) and Blackburn and Lawrence (1995) included all academics, not just scientists.
23. The 2010 SDR did not ask questions about publication.
24. Among others, Long (1992) showed that annual STEM-faculty publications change as a function of years since the receipt of PhD, rising for at least 12 years, with patterns over career years that differ by gender. In the SDR analysis, we found that 5-year publication rates are greatest for full professors and lowest for assistant professors, although our results may be the result of the least successful dropping out of academe.
25. Bellas and Toutkoushian (1999), who studied all fields, not just STEM, also found a large effect of teaching on research productivity.
26. Many scientists now emphasize the *b*-index, which is an author's maximum number (*b*) of articles that have at least *b* citations. (See, e.g., Hopkins, Jawitz, McCarty, Alex Goldman, & Basu, 2013). Although this is an excellent measure for identifying people's career accomplishments, it is highly dependent on their total publications and career length, does not isolate others' evaluation of research from production of the research itself, and, therefore, cannot be compared to citations per article controlling for years since publication. Given the increasing numbers of female STEM PhDs, there are of course more males than females at senior levels and thus with high *b*-indices. We do include in our summary table one article based on an *b*-index (Duch et al., 2012) because it considers a specific career length (10 years post-PhD) and adjusts for the numbers of publications.
27. Some of the differences at all institutions combined were due to there being more men at R1 universities, which as a group pay higher salaries in all 24 field/rank cells and so did not have simultaneous larger non-R1 and smaller R1 salary gaps.
28. Non-economists might deem these "societal expectations."
29. Kelly and Grant excluded all social scientists from these fields and included life scientists in their analysis.
30. For instance, although childless biomedical tenured and tenure-track women academics earn only 1.5% less than men when hired, by 10 years post-PhD, they earn 9.4% less.
31. In 2004, the NIH announced Research Supplements to Promote Reentry into Biomedical and Behavioral Research Careers (<http://grants.nih.gov/grants/guide/pa-files/PA-08-191.html>). In 2011, the NSF announced a 10-year plan to promote work-related flexibility for research scientists (<http://www.nsf.gov/pubs/2012/nsf12065/nsf12065.pdf>).
32. Calculated as a percentage of those who did not retire.

References

- Abrevaya, J., & Hamermesh, D. S. (2012). Charity and favoritism in the field: Are female economists nicer (to each other)? *The Review of Economics and Statistics*, 94, 202–207.
- Aksnes, D. W., Rorstad, K., Prio, F., & Sivertsen, G. (2011). Are female researchers less cited? A large-scale study of Norwegian scientists. *Journal of the American Society for Information Science and Technology*, 62, 628–636.
- Andreescu, T., Gallian, J. A., Kane, J., & Mertz, J. E. (2008). Cross-cultural analysis of students with exceptional talent in mathematical problem solving. *Notices of the American Mathematical Society*, 55, 1248–1260.
- Auyeung, B., Lombardo, M. V., & Baron-Cohen, S. (2013). Prenatal and postnatal hormone effects on the human brain and cognition. *Pflügers Archiv-European Journal of Physiology*, 465, 557–571. doi:10.1007/s00424-013-1268-2
- Babcock, L., & Laschever, S. (2003). *Women don't ask: Negotiation and the gender divide*. Princeton, NJ: Princeton University Press.
- Barone, C. (2011). Some things never change: Gender segregation in higher education across eight nations and three decades. *Sociology of Education*, 84, 157–176.
- 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.
- Belkin, L. (2003, October 26). The opt-out revolution. *New York Times Magazine*, pp. 42–47; 58; 85–86.
- Bellas, M. L., & Toutkoushian, R. K. (1999). Faculty time allocations and research productivity: Gender, race and family effects. *The Review of Higher Education*, 22, 367–390.
- Bettinger, E. P., & Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review*, 95(2), 152–157.
- Bielock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, 136, 256–276.
- Bilimoria, D., Joy, S., & Liang, X. (2008). Breaking barriers and creating inclusiveness: Lessons of organizational transformation to advance women faculty in academic science and engineering. *Human Resource Management*, 47, 423–441.
- Blackaby, D., Booth, A. L., & Frank, J. (2005). Outside offers and the gender pay gap: Empirical evidence from the UK academic labour market. *The Economic Journal*, 115, F81–F107.
- Blackburn, R. T., & Lawrence, J. H. (1995). *Faculty at work: Motivation, expectation, satisfaction*. Baltimore, MD: The Johns Hopkins University Press.
- Blair, C., Gamson, D., Thorne, S., & Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex. *Intelligence*, 33, 93–106.
- Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from The American Economic Review. *American Economic Review*, 81, 1041–1067.

- Borrego, A., Barrios, M., Villarroya, A., & Olle, C. (2010). Scientific output and impact of postdoctoral scientists: A gender perspective. *Scientometrics*, 83, 93–101.
- Borsuk, R. M., Aarssen, L. W., Budden, A. E., Koricheva, J., Leimu, R., Tregenza, T., & Lortie, C. J. (2009). To name or not to name: The effect of changing author gender on peer review. *BioScience*, 59, 985–989.
- Brooks, C. I., & Mercincavage, J. E. (1991). Superiority of women in statistics achievement. *Teaching of Psychology*, 14, 45.
- Brooks, J., & Della Sala, S. (2009). Re-addressing gender bias in *Cortex* publications. *Cortex*, 45, 1126–1137.
- Budden, A., Lortie, C., Tregenza, T., Aarssen, L., Koricheva, J., & Leimu, R. (2008). Response to Webb et al.: Blind review: Accept with minor revisions. *Trends in Ecology and Evolution*, 23, 353–355.
- Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American Sociological Review*, 66, 204–225.
- Bug, A. (2010, August). Swimming against the unseen tide. *PhysicsWorld*, 16–17. Retrieved from <http://physicsworld.archive.iop.org/index.cfm?action=summary&doc=23%2F08%2Fphwv23i08a27%40pwa-xml&qt=>
- Bukvova, H. (2010). Studying research collaboration: A literature review. *Sprouts: Working Papers on Information Systems*, 10(3). Retrieved from <http://sprouts.aisnet.org/10-3>
- Callister, R. R. (2006). The impact of gender and department climate on job satisfaction and intentions to quit for faculty in science and engineering fields. *Journal of Technology Transfer*, 31, 367–375.
- Canes, B. J., & Rosen, H. S. (1995). Following in her footsteps? Faculty gender composition and women's choices of college majors. *Industrial and Labor Relations Review*, 48, 486–504.
- Carlin, P. S., Kidd, M. P., Rooney, P. M., & Denton, B. (2013). Academic wage structure by gender: The roles of peer review, performance, and market forces. *Southern Economic Journal*, 80, 127–146.
- Carmichael, L. (2005, April 4). Lumber exports and women engineers. *Queen's Gazette*, p. 6. Retrieved from <http://www.queensu.ca/news/sites/default/files/assets/gazette/gazette20050404.pdf>
- Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125, 1101–1144.
- Casuso-Holgado, M. J., Cuesta-Vargas, A. I., Moreno-Morales, N., Labajos-Manzanares, M. T., Barón-López, F. J., & Vega-Cuesta, M. (2013). The association between academic engagement and achievement in health sciences students. *BMC Medical Education*, 13, Article 33. Retrieved from <http://www.biomedcentral.com/1472-6920/13/33>
- Ceci, S. J., & Williams, W. M. (2007). *Why aren't more women in science? Top researchers debate the evidence*. Washington, DC: APA Books.
- Ceci, S. J., & Williams, W. M. (2010a). *The mathematics of sex: How biology and society conspire to limit talented women and girls*. New York, NY: Oxford University Press.
- Ceci, S. J., & Williams, W. M. (2010b). Sex differences in math-intensive fields. *Current Directions in Psychological Science*, 19, 275–279.
- Ceci, S. J., & Williams, W. M. (2011). Current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences, USA*, 108, 3157–3162.
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, 135, 218–261.
- Cole, J. R., & Zuckerman, H. (1984). The productivity puzzle: Persistence and change in patterns of publication of men and women scientists. In M. W. Steinkamp & M. L. Maehr (Eds.), *Advances in motivation and achievement* (pp. 217–258). Greenwich, CN: JAI Press.
- Cole, J. R., & Zuckerman, H. (1987, February). Marriage, motherhood, and research performance in science. *Scientific American*, 256(2), 119–125.
- Conner, A. L., Cook, K. S., Correll, S. J., Markus, H. R., Moss-Racusin, C. A., Muller, C. B., Raymond, J. L., & Simard, C. Obscuring gender bias with “choice.” *Science*, 343, 1200. doi:10.1126/science.343.6176.1200-a
- Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, 106, 1691–1730.
- Correll, S. J. (2004). Constraints into preferences: Gender, status, and emerging career aspirations. *American Sociological Review*, 69, 93–113.
- Cotton, C., McIntyre, F., & Price, J. (2013). Gender differences in repeated competition: Evidence from school math contests. *Journal of Economic Behavior & Organization*, 86, 52–66.
- Creech, L. R., & Sweeder, R. D. (2012). Analysis of student performance in large-enrollment life science courses. *Life Sciences Education*. Retrieved from <http://www.lifescied.org/content/11/4/386>
- Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin*, 87, 546–563.
- Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology*, 38, 1639–1655.
- Daverman, R. J. (2011). Statistics on women mathematicians compiled by the American Mathematical Society. *Notices of the American Mathematical Society*, 58, 1310.
- Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, 42, 528–554.
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity or gender matter? *American Economic Review*, 95, 158–165.
- Ding, W. W., Murray, F., & Stuart, T. E. (2006). Gender differences in patenting in the academic life sciences. *Science*, 313, 665–667.
- Drago, R., Colbeck, C. L., Stauffer, K. D., Pirretti, A., Burkum, K., Fazioli, J., . . . Habasevich, J. (2006). The avoidance of bias against caregiving: The case of academic faculty. *American Behavioral Scientist*, 49, 1222–1247.
- Duch, J., Zeng, X. H. T., Sales-Pardo, M., Radicchi, F., Otis, S., Woodruff, T. K., & Amaral, L. A. N. (2012). The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PLoS ONE*, 7(12), e51332. Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0051332>
- Dunning, H. (2012, November 1). *Life Sciences Salary Survey 2012* [Web log post]. Retrieved from <http://www>

- .the-scientist.com/?articles.view/articleNo/32918/title/Life-Sciences-Salary-Survey-2012/
- Ecklund, E., & Lincoln, A. E. (2011). Scientists want more children. *PLoS ONE*, 6(8), e22590. Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0022590>
- Ellermeier, W., & Westphal, W. (1995). Gender differences in pain ratings and pupil reactions to painful pressure stimuli. *Pain*, 61, 435–439.
- Ellison, G., & Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: Evidence from the American Mathematics Competitions. *Journal of Economic Perspectives*, 24, 109–128.
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136, 103–127.
- Falter, C. M., Arroyo, M., & Davis, G. J. (2006). Testosterone: Activation or organization of spatial cognition? *Biological Psychology*, 73, 132–140.
- Ferriman, K., Lubinski, D., & Benbow, C. P. (2009). Work preferences, life values, and personal views of top math/science graduate students and the profoundly gifted: Developmental changes and gender differences during emerging adulthood and parenthood. *Journal of Personality and Social Psychology*, 97, 517–532.
- Fillingim, R. B., King, C., Ribeiro-Dasilva, M., Rahim-Williams, B., & Riley, J. (2009). Sex, gender, and pain: A review of recent clinical and experimental findings. *Pain*, 10, 447–485. doi:10.1016/j.jpain.2008.12.001
- Finegan, J.-A. K., Niccols, G. A., & Sitarenios, G. (1992). Relations between prenatal testosterone levels and cognitive abilities at 4 years. *Developmental Psychology*, 28, 1075–1089.
- Foschi, M., Lai, L., & Sigerson, K. (1994). Gender and double standards in the assessments of job candidates. *Social Psychology Quarterly*, 57, 326–339.
- Fox, M. F. (1995). Women in scientific careers. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.), *Handbook of science and technology studies* (pp. 205–223). Thousand Oaks, CA: Sage.
- Fox, M. F. (2005). Gender, family characteristics, and publication productivity among scientists. *Social Studies of Science*, 35, 131–150.
- Fox, M. F., Fonseca, C., & Bao, J. (2011). Work and family conflict in academic science: Patterns and predictions among women and men in research universities. *Social Studies of Science*, 41, 715–735.
- Fox, M. F., & Mohapatra, S. (2007). Social-organizational characteristics of work and publication productivity among academic scientists in doctoral-granting departments. *Journal of Higher Education*, 78, 543–571.
- 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.
- Gilbert, J. R., Williams, E. S., & Lundberg, G. D. (1994). Is there gender bias in JAMA's peer review? *Journal of the American Medical Association*, 272, 139–142.
- Ginther, D. K. (2003). Is MIT the exception? Gender pay differentials in academic science 1973–1997. *Bulletin of Science, Technology, & Society*, 23, 21–26.
- Ginther, D. K. (2004, Winter). Why women earn less: Economic explanations for the gender salary gap in science. *AWIS Magazine*, 33(1), 6–10.
- Ginther, D. K. (2006a). Economics of gendered distribution of resources in academia. In *Biological, social, and organizational components of success for women in science and engineering: Workshop report* (pp. 56–60). Washington, DC: Committee on Maximizing the Potential of Women in Academic Science and Engineering, The National Academies Press.
- Ginther, D. K. (2006b). The economics of gender differences in employment outcomes in academia. In *Biological, social, and organizational components of success for women in science and engineering: Workshop report* (pp. 99–112). Washington, DC: Committee on Maximizing the Potential of Women in Academic Science and Engineering, The National Academies Press.
- Ginther, D. K., & Kahn, S. (2004). Women in economics: Moving up or falling off the academic career ladder? *Journal of Economic Perspectives*, 18(3), 193–214.
- Ginther, D. K., & Kahn, S. (2009). Does science promote women? Evidence from Academia 1973–2001. In R. B. Freeman & D. F. Goroff (Eds.), *Science and engineering careers in the United States* (National Bureau of Economic Research Conference Report; pp. 163–194). Chicago, IL: University of Chicago Press.
- Ginther, D. K., & Kahn, S. (in press). Women's careers in academic social science: Progress, pitfalls, and plateaus. In A. Lanteri & J. Vromen (Eds.), *The economics of economists*. Cambridge, England: Cambridge University Press.
- Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., & Kington, R. (2011). Race, ethnicity, and NIH research awards. *Science*, 333, 1015–1019.
- Glass, C., & Minnotte, K. (2010). Recruiting and hiring women in STEM fields. *Journal of Diversity in Higher Education*, 3, 218–229.
- Goldin, C. (2014, January 4). *A grand gender convergence: Its last chapter*. American Economic Association Presidential Address, Philadelphia, PA.
- Goldin, C., & Katz, L. (2012, September). *The most egalitarian of all professions: Pharmacy and the evolution of a family-friendly occupation* (NBER Working Paper No. 18410). Washington, DC: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w18410>
- Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology*, 29, 17–28.
- Goulden, M., Frasch, K., & Mason, M. A. (2009). *Staying competitive: Patching America's leaky pipeline in the sciences*. Center for American Progress. Retrieved from http://cdn.americanprogress.org/wp-content/uploads/issues/2009/11/pdf/women_and_sciences.pdf
- Griffith, A. L. (in press). Faculty gender in the college classroom: Does it matter for achievement and major choice? *Southern Economic Journal*.
- Grissmer, D., Mashburn, A., Cottone, E., Chen, W.-B., Brock, L., Murrah, W., . . . Cameron, C. (in press). Play-based after-school intervention simultaneously improves visuo-spatial,

- executive function, and math skills for disadvantaged children. *Developmental Psychology*.
- Halpern, D. F. (2012). *Sex differences in cognitive abilities* (4th ed.). New York, NY: Psychology Press.
- 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.
- Hammerschmidt, K., Reinhardt, K., & Rolff, J. (2008). Does double-blind review favor female authors? *Frontiers in Ecology and the Environment*, 6, 354.
- Hargens, L. L., McCann, J. C., & Reskin, B. F. (1978). Productivity and reproductivity: Fertility and professional achievement among research scientists. *Social Forces*, 57, 154–163.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41–45.
- Heilman, M., Martell, R., & Simon, M. (1988). The vagaries of sex bias. *Organizational Behavior and Human Decision Processes*, 41, 98–110.
- Helmreich, R. L., Spence, J. T., Beane, W. E., Lucker, G. W., & Matthews, K. A. (1980). Making it in academic psychology: Demographic and personality correlates of attainment. *Journal of Personality and Social Psychology*, 39, 869–908.
- Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. Washington, DC: American Association of University Women.
- Hines, M., Fane, B. A., Pasterski, V. L., Mathews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality: Targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology*, 28, 1010–1026.
- Hoffman, F., & Oreopoulos, P. (2007, June). *A professor like me: The influence of instructor gender on college achievement* (NBER Working Paper No. 13182). Washington, DC: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w13182>
- Hopkins, A., Jawitz, J. W., McCarty, C., Alex Goldman, A., & Basu, N. (2013). Disparities in publication patterns by gender, race and ethnicity based on a survey of a random sample of authors. *Scientometrics*, 96, 515–534. doi:10.1007/s11192-012-0893-4
- Hosek, S., Cox, A., Ghosh-Dastidar, B., Kofner, A., Ramphal, N., Scott, J. S., & Berry, S. (2005). *Gender differences in major federal external grant programs*. Retrieved from http://www.rand.org/pubs/technical_reports/TR307.html
- Huguet, P., & Regner, I. (2009). Counter-stereotypic beliefs in math do not protect school girls from stereotype threat. *Journal of Experimental Social Psychology*, 45, 1024–1027.
- Hunt, J. (2010, March). *Why do women leave science and engineering?* (NBER Working Paper No. 15853). Washington, DC: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w15853>
- Hunter, L. A., & Leahey, E. (2010). Parenting and research productivity: New evidence and methods. *Social Studies of Science*, 40, 433–451.
- Hutson, S. R. (2002). Gendered citation practices in *American Antiquity* and other archaeology journals. *American Antiquity*, 67, 331–342.
- Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139–155.
- Hyde, J. S., Lindberg, S., 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.
- Hyde, J. S., & Mertz, J. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, USA*, 106, 8801–8809.
- Ingallhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., . . . & Verma, R. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences, USA*, 111, 823–828. Retrieved from www.pnas.org/cgi/doi/10.1073/pnas.1316909110
- Institute of Medicine, National Academy of Sciences, and National Academy of Engineering. (2007). *Beyond bias and barriers: Fulfilling the potential of women in academic science and engineering*. Washington, DC: The National Academies Press.
- Jackson, L. A., Gardner, P. D., & Sullivan, L. A. (1993). Engineering persistence: Past, present and future factors and gender differences. *Higher Education*, 26, 227–246.
- Jayasinghe, U., Marsh, H., & Bond, N. (2003). A multilevel cross-classified modeling approach to peer-review of grant proposals. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, 166A, 279–300.
- Kahn, S. (2013, March). *Careers of engineering PhD women in the US 1980-2010: Equal treatment? Equal choices?* Lecture conducted from Tel Aviv University, Tel Aviv, Israel.
- Kahn, S., & Ginther, D. (2012, November). *Gender differences in employment and salaries in biomedical careers. Lecture conducted at the National Institutes of Health Causal Factors and Interventions Workshop, Washington, DC.*
- Kahn, S., & MacGarvie, M. (2014). *Do return requirements increase international knowledge diffusion?* Unpublished manuscript, Boston University, MA.
- Kali. (2011, April 13). Re: Why aren't more women becoming physics professors? [Web log comment]. Retrieved from <http://www.thenewagenda.net/2011/04/13/why-arent-more-women-becoming-physics-professors/>
- Kaminski, D., & Geisler, B. (2012). Gender survival analysis of faculty retention in science and engineering. *Science*, 335, 864–866. doi:10.1126/science.1214844
- Keith, B., Layne, J. S., Babchuk, N., & Johnson, K. (2002). The context of scientific achievement: Sex status, organizational environments, and the timing of publication of on scholarly outcomes. *Social Forces*, 80, 1253–1282.
- Kelly, K., & Grant, L. (2012). Penalties and premiums: The impact of gender, marriage, and parenthood, on faculty salaries in STEM and non-STEM Fields. *Social Studies of Science*, 42, 872–899.

- Kessel, C. (2012, February 15). *The pipeline and the trough* [Web log post]. Retrieved from <http://mathedck.wordpress.com/2012/02/15/the-pipeline-and-the-trough/>
- Kessel, C., & Vitulli, M. A. (2011). Critique of "Understanding Current Causes of Women's Underrepresentation in Science" [Web log post]. Retrieved from <https://sites.google.com/site/awmmath/awm-resources/policy-and-advocacy/critique-of-understanding-current-causes-of-women-s-underrepresentation-in-science>
- Kimball, M. M. (1989). A new perspective on women's mathematics achievement. *Psychological Bulletin*, 105, 198–214.
- Krendl, A., Richeson, J., Kelley, W., & Hamilton, T. (2008). The negative consequences of threat: An fMRI investigation of the neural mechanisms underlying women's underperformance in math. *Psychological Science*, 19, 168–175.
- Leahey, E. (2006). Gender differences in productivity: Research specialization as the missing link. *Gender & Society*, 20, 754–780.
- Ledin, A., Bornmann, L., Gannon, F., & Wallon, G. (2007). A persistent problem. *EMBO Reports*, 8, 982–996.
- Legewie, J., & DiPrete, T. A. (2012a). *High school environments, STEM orientations, and the gender gap in science and engineering degrees*. Retrieved from <http://ssrn.com/abstract=2008733>
- Levin, S. G., & Stephan, P. E. (1998). Gender differences in the rewards to publishing in Academe: Science in the 1970s. *Sex Roles*, 38, 1049–1064.
- Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology*, 48, 530–542. doi:10.1037/a0025913
- Lewison, G. (2001). The quantity and quality of female researchers: A bibliometric study of Iceland. *Scientometrics*, 52, 29–43.
- Ley, T., & Hamilton, B. (2008). Gender gap in NIH grant applications. *Science*, 322, 1472–1474.
- Lindberg, S. M., Hyde, J. S., & Petersen, J. L. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, 136, 1123–1135.
- 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. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Journal of Personality and Social Psychology*, 74, 996–1009.
- Lippa, R. A. (2001). On deconstructing and reconstructing masculinity–femininity. *Journal of Research in Personality*, 35, 168–207.
- Lippa, R. A. (2010). Sex differences in personality traits and gender-related occupational preferences across 53 nations: Testing evolutionary and social-environmental theories. *Archives of Sexual Behavior*, 39, 619–636.
- Lohman, D. F., & Lakin, J. M. (2009). Consistencies in sex differences on the cognitive abilities test across countries, grades, test forms, and cohorts. *British Journal of Educational Psychology*, 79, 89–407.
- Long, J. S. (1992). Measures of sex differences in scientific productivity. *Social Forces*, 71, 159–178.
- Lortie, C. J., Aarssen, L., Budden, A. E., Koricheva, J., Leimu, R., & Tregenza, T. (2007). Publication bias and merit in ecology. *Oikos*, 116, 1247–1253.
- Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for math–science expertise. *Perspectives on Psychological Science*, 1, 316–345.
- Lubinski, D., Benbow, C. P., Shea, D. L., Eftekhari-Sanjani, H., & Halvorson, M. B. J. (2001). Men and women at promise for scientific excellence: Similarity not dissimilarity. *Psychological Science*, 12, 309–317.
- 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.
- Lundquist, J. H., Misra, J., & O'Meara, K. (2012). Parental leave usage by fathers and mothers at an American University. *Fathering*, 10, 337–363.
- Maliniak, D., Powers, R. M., & Walter, B. F. (2013, August). *The gender citation gap*. Paper presented at the 2013 Annual Meeting of the American Political Science Association, Chicago, IL. Retrieved from <http://ssrn.com/abstract=2303311>
- Manchester, C. F., & Barbezat, D. (2013). The effect of time use in explaining male-female productivity differences among economists. *Industrial Relations*, 52, 53–77.
- Manchester, C. F., Leslie, L. M., & Kramer, A. (2013). Is the clock still ticking? An evaluation of the consequences of stopping the tenure clock. *Industrial and Labor Relations Review*, 63, 1–36.
- Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O'Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Education Research*, 79, 1290–1326.
- Martinez, E. D., Botos, J., Dohoney, K. M., Geiman, T. M., Kolla, S. S., Olivera, A., . . . & Cohen-Fix, O. (2007). Falling off the academic bandwagon: Women are more likely to quit at the postdoc to principal investigator transition. *EMBO Reports*, 8, 977–981.
- Mason, M. A., Wolfinger, N. H., & Goulden, M. (2013). *Do babies matter: Gender and family in the ivory tower*. New Brunswick, NJ: Rutgers University Press.
- Massachusetts Institute of Technology. (1999). *A study of the status of women faculty in Science at MIT: How a Committee on Women Faculty came to be established by the Dean of the School of Science, what the Committee and the Dean learned and accomplished, and recommendations for the future*. Retrieved from <http://web.mit.edu/fnl/women/women.pdf>
- Massachusetts Institute of Technology. (2002). *Reports on the Committees on the Status of Women Faculty: Report of the school of humanities, arts, and social sciences*. Retrieved from <http://web.mit.edu/faculty/reports/sohass.html>
- Mavriplis, C., Heller, R. S., Beil, C., Dam, K., Yassinskaya, N., Shaw, M., & Sorensen, C. (2010). Mind the gap: Women in STEM career breaks. *Journal of Technology Management & Innovation*, 5, 140–151.
- Mayer, A. L., & Tikka, P. M. (2008). Family-friendly policies and gender bias in academia. *Journal of Higher Education Policy and Management*, 30, 363–374.

- McDowell, J. M., Singell, L. D., & Stater, M. (2006). Two to tango? Gender differences in the decisions to publish and coauthor. *Economic Inquiry*, 44, 153–168.
- McNeely, C. L., & Schintler, L. (2010, December 2–3). *Gender issues in scientific collaboration and workforce development: Implications for a federal policy research agenda*. Workshop on the Science of Science Measurement, U.S. Office of Science and Technology Policy, Washington, DC. Retrieved from <http://www.nsf.gov/sbe/sosp/workforce/mcneely.pdf>
- Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences*, 18, 37–45.
- Misra, J., Lundquist, J., Holmes, E. D., & Agiomavritis, S. (2011). Associate professors and gendered barriers to advancement. *Academe*, 97(1). Retrieved from <http://www.aaup.org/AAUP/pubsres/academe/2011/JF/Feat/misr.htm>
- Monosson, E. (Ed.). (2008). *Motherhood, the elephant in the laboratory: Women scientists speak out*. Ithaca, NY: Cornell University Press.
- Montpetit, E., Blais, A., & Foucault, M. (2008). What does it take for a Canadian political scientist to be cited? *Social Science Quarterly*, 89, 802–816.
- Moore, D. S., & Johnson, S. P. (2011). Mental rotation of dynamic, three-dimensional stimuli by 3-month-old infants. *Infancy*, 16, 435–445.
- Morgan, S. L., Gelbgeiser, D., & Weeden, K. (2013). Feeding the pipeline: Gender, occupational plans, and college major selection. *Social Science Research*. doi:10.1016/j.ssresearch.2013.03.008
- Moss-Racusin, C. (2012, September 21). Are science faculty biased against female students? [Web blog post]. Retrieved from <http://spsptalks.wordpress.com/2012/09/21/are-science-faculty-biased-against-female-students/>
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, USA*, 109, 16474–16479.
- Moss-Racusin, C. A., van der Toorn, J., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2014). Scientific diversity interventions. *Science*, 343, 615–616.
- Mullis, I. V. S., Martin, M. O., Fierros, E. G., Goldbcrg, A. L., & Stemler, S. E. (2000). *Gender differences in achievement: IEA's third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College. Retrieved from http://timss.bc.edu/timss1995i/timsspdf/t95_gender_all.pdf
- Mutz, R., Bornmann, L., & Daniel, H.-D. (2012). Does gender matter in grant peer review? An empirical investigation using the example of the Austrian Science Fund. *Zeitschrift fur Psychologie*, 220, 121–129. doi:10.1027/2151-2604/a000103
- National Institutes of Health. (2012). *Biomedical Research Workforce Working Group Report*. Retrieved from http://acd.od.nih.gov/Biomedical_research_wgreport.pdf
- National Research Council. (2001). *From scarcity to visibility: Gender differences in the careers of doctoral scientists and engineers*. Washington, DC: The National Academies Press.
- National Research Council. (2010). *Gender differences at critical transitions in the careers of science, engineering and mathematics faculty*. Washington, DC: The National Academies Press.
- National Research Council. (2013). *Seeking solutions: Maximizing American talent by advancing women of color in academia: Summary of a conference*. Washington, DC: The National Academies Press.
- National Science Foundation. (2004). *Gender differences in the careers of academic scientists and engineers* (NSF 04–323, Project Officer, Alan I. Rapoport). Arlington, VA: National Science Foundation.
- Nature Neuroscience. (2006, July). Women in neuroscience: A numbers game. *Nature Neuroscience*, 9(7), 853.
- Nettles, M. T., & Millett, C. M. (2006). *Three magic letters: Getting to the PhD*. Baltimore, MD: Johns Hopkins Press.
- Nettles, M. T., Perna, L. W., & Bradburn, E. M. (2000). *Salary, promotion, and tenure status of minority and women faculty in U.S. colleges and universities* (NCES 2000–361173). Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Nie, N., Golde, S., & Butler, D. (2009). *Education and verbal ability over time: Evidence from three multi-time sources*. Retrieved from http://www.learningace.com/doc/1391311/129f7bd3448db86d19dba0cae2553072/education_and_verbal_ability
- Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, 24, 129–144.
- 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.
- Penner, A. M. (2008). Gender differences in extreme mathematical achievement: An international perspective on biological and social factors. *American Journal of Sociology*, 114, 138–170.
- Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, 37, 239–253.
- Perez-Felkner, L., McDonald, S., Schneider, B., & Grogan, E. (2012). Female and male adolescents' subjective orientations to mathematics and the influence of those orientations on postsecondary majors. *Developmental Psychology*, 48, 1658–1673. doi:10.1037/a0027020
- Perna, L. W. (2001a). Sex and race differences in faculty tenure and promotion. *Research in Higher Education*, 42, 541–567.
- Perna, L. W. (2001b). Sex differences in faculty salaries: A cohort analysis. *Review of Higher Education*, 24, 283–307.
- Perna, L. W. (2005). Sex differences in faculty tenure and promotion: The contribution of family ties. *Research in Higher Education*, 46, 277–307.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811–832.
- Pohlhaus, J. R., Jiang, H., Wagner, R. M., Schaffer, W. T., & Pinn, V. W. (2011). Sex differences in application, success,

- and funding rates for NIH extramural programs. *Academic Medicine*, 86, 759–767.
- Pollack, E. (2013, October 3). Why are there still so few women in science? *The New York Times Magazine*. Retrieved from <http://www.nytimes.com/2013/10/06/magazine/why-are-there-still-so-few-women-in-science.html>
- Pope, D. G., & Sydnor, J. R. (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives*, 24, 95–108.
- Powell, K. (2007). Beyond the glass ceiling. *Nature*, 448, 99.
- Preston, A. E. (1994). Why have all the women gone? A study of exit from the science and engineering professions. *American Economic Review*, 84, 1446–1462.
- Preston, A. E. (2004). *Leaving science: Occupational exit from scientific careers*. New York, NY: Russell Sage Foundation.
- Prpić, K. (2002). Gender and productivity differentials in science. *Scientometrics*, 55, 27–58.
- 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 from http://rand.org/pubs/research_briefs/RB9147/RAND_RB9147.pdf
- Rask, K. N., & Bailey, E. M. (2002). Are faculty role models? Evidence from major choice in an undergraduate institution. *The Journal of Economic Education*, 33, 99–124.
- Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences, USA*, 111, 4403–4408.
- Rhoades, S. E., & Rhoades, C. H. (2012). Gender roles and infant/toddler care: Male and female professors on the tenure track. *Journal of Social, Evolutionary, and Cultural Psychology*, 6, 13–31.
- Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry to STEM college majors over time. *American Education Research Journal*, 49, 1048–1073.
- Rosenbloom, J. L., Ash, R. A., Dupont, B., & Coder, L. A. (2008). Why are there so few women in information technology? Assessing the role of personality in career choices. *Journal of Economic Psychology*, 29, 534–554.
- Sandberg, S. (2013). *Lean in: Women, work, and the will to lead*. New York, NY: Alfred A. Knopf.
- Saul, J. (2013). Implicit bias, stereotype threat, and women in philosophy. In K. Hutchison & F. Jenkins (Eds.), *Women in philosophy: What needs to change?* (pp. 39–60). New York, NY: Oxford University Press.
- Sax, L. J., Hagedorn, L. S., Arredondo, M., & Dicrisi, F. A., III. (2002). Faculty research productivity: Exploring the role of gender and family-related factors. *Research in Higher Education*, 43, 423–446.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440–452.
- Schwarzer, G., Freitag, C., & Schum, N. (2013). How crawling and manual object exploration are related to the mental rotation abilities of 9-month-old infants. *Frontiers in Psychology*, 4, 97–110. doi:10.3389/fpsyg.2013.00097
- Settles, I. H., Cortina, L. M., Malley, J., & Stewart, A. J. (2006). The climate for women in academic science: The good, the bad, and the changeable. *Psychology of Women Quarterly*, 30, 47–58.
- Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview Press.
- Sheltzer, J. M., & Smith, J. C. (2014). Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences, USA*. Advance online publication. doi:10.1073/pnas.1403334111
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703.
- Slyder, J. B., Stein, B. R., Sams, B. S., Walker, B. J., Feldhaus, J. J., & Copenheaver, C. A. (2011). Citation pattern and lifespan: A comparison of discipline, institution and individual. *Scientometrics*, 89, 955–966.
- Smart, S., & Waldfogel, J. (1996, February). *A citation-based test for discrimination at economics and finance journals* (NBER Working Paper No. 5460). Washington, DC: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w5460>
- Sonnert, G., & Fox, M. F. (2012). Women, men and academic performance in science and engineering: The gender difference in undergraduate grade point averages. *Journal of Higher Education*, 83, 73–101.
- Spelke, E. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, 60, 950–958.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28.
- Stack, S. (2004). Gender, children and research productivity. *Research in Higher Education*, 45, 891–920. doi:10.1007/s11162-004-5953-z
- Steele, C. M. (1997). A threat is in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629.
- Steinpreis, R., Anders, K., & 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.
- Stinebrickner, R., & Stinebrickner, T. R. (2013, June). *A major in science? Initial beliefs and final outcomes for college major and dropout* (NBER Working Paper 19165). Washington, DC: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w19165>
- Stinebrickner, T. R., & Stinebrickner, R. (2011, March). *Math or science? Using longitudinal expectations data to examine the process of choosing a college major* (NBER Working Paper 16869). Washington, DC: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w16869>
- Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within-nation assessment of ten years of PISA data. *PLoS ONE*, 8(3), e57988. Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0057988>
- 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., & 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. (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.
- Su, R., Rounds, J., & Armstrong, P. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135, 859–884.
- Swiatek, M. A., Lupkowski-Shoplik, 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.
- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, 105, 409–429.
- Symonds, M. R. E., Gemmell, N. J., Braisher, T. L., Gorringer, K. L., & Elgar, M. A. (2006). Gender differences in publication output: Towards an unbiased metric of research performance. *PLoS ONE*, 1(1), e127. Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000127>
- Tai, R. H., & Sadler, P. (2001). Gender differences in introductory undergraduate physics performance: University physics versus college physics in the USA. *International Journal of Science Education*, 23, 1017–1037.
- Thorndike, E. L. (1911). *Individuality*. New York, NY: Houghton & Mifflin.
- Toutkoushian, R. K., Bellas, M. L., & Moore, J. V. (2007). The interaction effects of gender, race, and marital status on faculty salaries. *The Journal of Higher Education*, 78, 572–601.
- Toutkoushian, R. K., & Conley, V. C. (2005). Progress for women in academe, but inequities persist: Evidence from NSOPF:99. *Research in Higher Education*, 46, 1–28.
- Tregenza, T. (2002). Gender bias in the refereeing process? *Trends in Ecology & Evolution*, 17, 349–351.
- Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse & Society*, 14, 191–220.
- Trower, C. A., & Bleak, J. L. (2004). *Study of new scholars. Gender: Statistical Report [Universities]*. Cambridge, MA: Harvard Graduate School of Education. Retrieved from <http://www.kuleuven.be/diversiteit/publicaties/Gender-statistical-report.pdf>
- Valian, V. (2014). Interests, gender, and science. *Perspectives in Psychological Science*, 9, 225–230. doi:10.1177/1745691613519109
- Valla, J., & Ceci, S. J. (2011). Can sex differences in science be tied to the long reach of prenatal hormones?: Brain organization theory, digit ratio (2D/4D), and sex differences in preferences and cognition. *Perspectives on Psychological Science*, 6, 134–146.
- Voyer, D., Voyer, 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.
- Vuoksima, E., Kaprio, J., Kremen, W. S., Hokkanen, L., Viken, R. J., Tuulio-Henriksson, A., & Rose, R. J. (2010). Having a male co-twin masculinizes mental rotation performance in females. *Psychological Science*, 21, 1069–1071.
- Wai, J., Cacchio, M., Putallaz, M., & Makel, M. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence*, 38, 412–423. doi:10.1016/j.intell.2010.04.006
- Wai, J., & Putallaz, M. (2011). The Flynn effect puzzle: A 30-year examination from the right tail of the ability distribution provides some missing pieces. *Intelligence*, 39, 443–455.
- Waldfogel, J. (1997). The effect of children on women's wages. *American Sociological Review*, 62, 209–217.
- Wang, M.-T., Eccles, J., & Kenney, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, 24, 770–775. doi:10.1177/0956797612458937
- Ward, K., & Wolf-Wendel, L. (2012). *Academic motherhood: How faculty manage work and family*. New Brunswick, NJ: Rutgers University Press.
- Wenneras, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, 387, 341–343.
- West, M. S., & Curtiss, J. W. (2006). *AAUP Faculty Gender Equity Indicators 2006*. Washington, DC: American Association of University Professors.
- Williams, C., & Meck, W. (1991). The organizational effects of gonadal steroids on sexually dimorphic spatial ability. *Psychoneuroendocrinology*, 16, 155–176.
- Williams, W. M., & Ceci, S. J. (2012). When scientists choose motherhood. *American Scientist*, 100, 138–145.
- Williams, W. M., & Ceci, S. J. (2014). *National experiment reveals 2-to-1 hiring preference for women in the STEM tenure-track*. Manuscript submitted for publication.
- Winslow, S. (2010). Gender inequality and time allocations among academic faculty. *Gender & Society*, 24, 769–793.
- Wolfinger, N., Mason, M. A., & Goulden, M. (2008). Problems in the pipeline: Gender, marriage, and fertility in the ivory tower. *Journal of Higher Education*, 79, 388–405.
- Wolfinger, N., Mason, M. A., & Goulden, M. (2009). Stay in the game: Gender, family formation, and alternative trajectories in the academic life course. *Social Forces*, 87, 1591–1621.
- Xie, Y., & Killewald, A. A. (2012). *Is American science in decline?* Cambridge, MA: Harvard University Press.
- Xie, Y., & Shauman, K. A. (1998). Sex differences in research productivity: New evidence about an old puzzle. *American Sociological Review*, 63, 847–870.
- Xie, Y., & Shauman, K. A. (2003). *Women in science: Career processes and outcomes*. Cambridge, MA: Harvard University Press.
- Xu, Y. J. (2008a). Faculty turnover: Discipline-specific attention is warranted. *Research in Higher Education*, 49, 40–61.
- Xu, Y. J. (2008b). Gender disparity in STEM disciplines: A study of faculty attrition and turnover intentions. *Research in Higher Education*, 49, 607–624.