



# Men's Overpersistence and the Gender Gap in Science and Mathematics

Andrew M. Penner<sup>1</sup> and Robb Willer<sup>2</sup>

Socius: Sociological Research for a Dynamic World  
 Volume 5: 1–17  
 © The Author(s) 2019  
 Article reuse guidelines:  
[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)  
 DOI: 10.1177/2378023118821836  
[srd.sagepub.com](http://srd.sagepub.com)



## Abstract

Large and long-standing gaps exist in the gender composition of science, technology, engineering, and mathematics (STEM) fields. Abundant research has sought to explain these gaps, typically focusing on women, though these gaps result from the decisions of men as well as women. Here we study gender differences in STEM persistence with a focus on men's choices, finding that men persist in these domains even where opting out could lead to greater material payoffs. Study 1 employed a novel experimental paradigm for measuring "overpersistence," finding that undergraduate men chose mathematics questions over verbal questions at higher rates than undergraduate women on a test in which mathematics questions were substantially more difficult than verbal questions and participants were paid for correct answers. Study 2 analyzed data from a nationally representative longitudinal survey, finding that men are more likely than women to retake college STEM courses after failing them and that men's STEM retaking after failure is as likely to lead to lower later life earnings as to higher earnings. Finally, in Study 3, we used a survey-embedded experiment to examine the intervening factors driving men's overpersistence in a diverse sample of adults. Integrating prior theoretical work, we find evidence for a model in which cultural stereotypes of male superiority in mathematics lead men both to be more confident in and identify more with the mathematics domain, factors that in turn lead men to pursue math to a greater extent than women.

## Keywords

gender, mathematics, science, education

While gender inequality in many domains has declined significantly over the past several decades, women remain significantly underrepresented in many mathematics- and science-intensive fields (National Academy of Sciences 2006). Some researchers have even suggested that the underrepresentation of women in science, technology, engineering, and mathematics (STEM) fields is a "final frontier" of occupational gender inequality (Xie and Shauman 2003), and research on this issue has garnered considerable attention (e.g., Barres 2006; Williams and Ceci 2015). Given concerns around gender equality in this domain, a considerable body of research has sought to understand how women can be encouraged to enter and persist in these areas in greater numbers (see e.g., National Academy of Sciences 2006).

In seeking to understand how women's participation in STEM fields can be increased, however, research has focused primarily on women's failure to enter and persist in STEM fields as the primary factor driving unequal representations in these fields (cf. Eccles 1987). By contrast, prior research has paid much less attention to men's decisions to join and persist in STEM fields, implicitly assuming that men's

behavior in this domain is normal and desirable while women's behavior is aberrant, requiring careful study and explanation. As a result, this work neglects that men's persistence also contributes to unequal representation in STEM fields. The belief that men's math- and science-related behavior is normal and sound is consistent with stereotypic views of men, especially those engaged in mathematics- and science-intensive fields of study, as rational, cerebral, and intellectual (Andersen 2001).

Here we propose that gender gaps in STEM fields are produced not only by women "underpersisting," refraining from an activity even when doing so is likely to bring less material success, but also by men "overpersisting," or engaging in an activity even when doing so is likely to bring less material

<sup>1</sup>University of California, Irvine, CA, USA

<sup>2</sup>Stanford University, Stanford, CA, USA

## Corresponding Author:

Andrew M. Penner, Department of Sociology, University of California, Irvine, 3151 Social Science Plaza, Irvine, 92697-5100, USA.  
 Email: [penner@uci.edu](mailto:penner@uci.edu)



success.<sup>1</sup> We reason that men, buttressed by cultural beliefs that mathematics and science are male domains (e.g., Fennema and Sherman 1977), are more likely than women to choose to enter and persist in STEM fields even when doing so leads to materially suboptimal outcomes.

Following a long tradition in feminist scholarship that critiques the acceptance of male behavior as normative (Beauvoir 2010; Stage and Maple 1996), we provide evidence that men's pursuit of mathematics is not always strictly optimizing of material payoffs. Instead, men's mathematics-related behavior is driven by gendered self-perceptions conditioned by cultural beliefs about men's relative competence in STEM fields in the same way women's choices are shaped by cultural beliefs about women's STEM competence.

Research and policy in STEM fields imply that gender differences in STEM persistence result from female underpersistence. Work in this area suggests that women's underpersistence should be brought into line with men's persistence, characterizing women as "leaking" out of the STEM pipeline (Alper 1993; Griffith 2010; Ma 2011) and highlighting the "obstacles to women's persistence" (Blair et al. 2017:15). While prior research has documented men's greater affinity for math- and science-related fields, it has not addressed whether men's pursuit of these fields optimizes material outcomes or whether it is simply greater than women's rates. We provide a novel experimental paradigm that allows us to address this question by creating contexts where persisting in math is and is not optimal for achieving material success,<sup>2</sup> allowing us to demonstrate that men's choices in this field are not necessarily optimal. In addition to providing an existence proof of male overpersistence in a controlled experimental setting, we also explore whether male overpersistence is reflected in men's real-world STEM choices.

## Previous Research

A substantial body of research has documented significant gaps in gender representation in most STEM fields. Although in fields like biology women now receive 53 percent of PhDs (Snyder and Dillow 2012), they remain underrepresented in fields like mathematics and statistics (where they receive 29

percent of PhDs), physical sciences (32 percent), and engineering (22 percent). A variety of explanations for the underrepresentation of women in STEM fields have been advanced, and research on STEM training often portrays advancement in these fields in terms of progress through the "STEM pipeline."<sup>3</sup> This metaphor emphasizes the sequential nature of training in mathematics and other STEM fields, conceptualizing the STEM training process as a progression through a series of stages, each of which must be successfully navigated before advancing to the next. In this framework, the gender differences observed in representation in STEM fields are typically attributed to women's failure to persist, or put differently, the idea that women "leak out" of the pipeline at various points (Alper 1993). Although this research typically views men's persistence as normative, we suggest that men's behavior may also contribute to the gender gap in STEM. Just as women are often seen as underpersisting, failing to pursue careers in science and mathematics despite sufficient qualifications, we hypothesize here that men may often overpersist, choosing STEM even when doing so is likely to lead to less academic and professional success.

Two perspectives within research on the gender gap in STEM representation provide insight on men's persistence in STEM fields. First, Correll's (2001) research on gender-biased self-assessments suggests that men may make decisions regarding STEM persistence based on inflated views of their STEM abilities. Alternatively, men may persist because they perceive STEM fields as masculine, with STEM persistence being a way to enact a masculine gender identity in these domains (Charles and Bradley 2009). In the following, we briefly review these two perspectives in turn.

## The Role of Self-Perceptions

One potential mechanism that could help us understand why boys are more likely to choose mathematics-related fields than girls centers on biases in self-perceived mathematical competence. According to this line of argumentation, people's perceptions of their competence are more important than their actual competence in determining their choices about whether to persist in STEM fields. That is, if people

<sup>1</sup>We focus on material success because we can clearly identify contexts in which the material payoffs vary by choosing STEM or non-STEM options. In theory, our reasoning can be extended to nonpecuniary motivations as well, though these are likely more difficult to assess. In practice, as we note in the following, participants in our experiments report being overwhelmingly motivated by pecuniary considerations.

<sup>2</sup>Research suggests that individuals often enter and exit the STEM pipeline multiple times in their educational and professional careers (Xie and Shauman 2003). Thus, we can think of the STEM pipeline as a series of decision points at which people choose STEM or non-STEM options. We follow the literature in referring to these choices as "persistence" (e.g., Ackerman, Kanfer, and Beier 2013; Griffith 2010; Price 2010).

<sup>3</sup>The literature on gender differences in STEM fields is extensive, including many important factors affecting both supply- and demand-side considerations. We focus narrowly on the literature around self-assessments and gender identity as we believe that it provides insights that are important for understanding men's STEM choices. There are undoubtedly other important factors, including national gender schemas (Riegle-Crumb 2005), school structures (Ayalon and Livneh 2013), single sex schooling (Legewie and DiPrete 2014), peers (Frank et al. 2008), teacher attitudes (Beilock et al. 2010), cultural mismatch (Seron et al. 2016), and discrimination (Li 2012), among others. For a more comprehensive review of gender differences in STEM fields, see the National Academy of Sciences (2006) report.

think that they lack the skills necessary to be successful in a STEM field, they are unlikely to persist regardless of their actual level of competence. Likewise, if people believe that they have the skills necessary to be successful, they may persist even in the face of mounting evidence that they lack the skills needed. Correll (2001) finds evidence of a gap between actual and perceived competence, showing that controlling for actual achievement, boys generally rate themselves better at math than girls do. However, it remains an open question whether gender differences in self-perceptions of competence are caused by boys overrating themselves, girls underrating themselves, or some combination of the two processes. Given that Correll (2001) shows that girls are more sensitive to feedback in creating their mathematical self-concept, we might expect boys' self-conceptions to be less affected by past performance than girls', lending credence to the idea that men may overpersist in STEM fields.

General processes underlie gender differences in mathematical self-assessment, as Correll (2004) shows that this pattern of gender differences in assessments can exist in domains beyond STEM where gender differences are perceived to exist. Using a test in a fictitious skill domain with manipulated feedback, Correll (2004) found that when participants were told that there is a male advantage in a skill area, men tended to assess themselves as more competent. In contrast, when participants were told that gender differences did not exist in the skill area, no differences in self-assessments were found. This suggests that gender differences in self-assessments are influenced by individuals' beliefs about how people of their gender generally perform in a given area.

Just as gendered self-assessments of competence could lead women to not pursue STEM fields proportional to their actual competence, corresponding biased self-assessments in men could lead them to pursue these domains beyond what their competence would justify. While Correll (2004) does not link gender differences in self-assessments with behavioral measures of persistence, she shows that differences in self-assessments in a domain lead in turn to gender differences in aspirations in the domain even when individuals receive the same feedback about their performance. This provides strong evidence that perceived gender differences actually create gender differences in both self-assessments and aspirations independent of individual performance feedback. Thus, we might expect men to be strongly oriented toward persistence in STEM fields because they believe that math is a domain in which men outperform women regardless of their actual likelihood of success.

### *Doing Gender by Doing Math*

International research on the sex segregation of college majors suggests another explanation for the divergent choices of women and men. Charles and Bradley (2009) underscore the importance of understanding the role of self-expression in gender segregation, arguing that gender differences in

college majors are driven in part by students choosing their majors not just for instrumental reasons (e.g., getting a job) but also as acts of self-expression intended to project a particular identity. Drawing on research highlighting the conception of STEM fields as more appropriate for men (cf. Fennema and Sherman 1977), Charles and Bradley (2009) note that gender segregation tends to be more extreme in contexts where self-expression is more highly valued. They suggest that this is driven by gender essentialist ideologies, which in combination with the emphasis on self-expression found in advanced industrial societies result in students expressing their gendered selves in part through their choice of major.

This perspective emphasizes the role of stereotypes linking mathematics and masculinity in academic domains. Stage and Maple (1996), for example, interview women who note that math is “the macho of the intellectual world” (p. 29) and that in mathematics graduate programs they were asked to “be like a man” (p. 36). Among other things, “being like a man” in these contexts means working in isolation (as opposed to in a group) and neglecting other relationships and responsibilities. Likewise, Andersen (2001) argues that objectivity, rationality, and a scientific approach are typically conceptualized as male-typed characteristics by both men and women, suggesting that stereotypic views of men and scientists are highly overlapping.

Research on the masculine conception of mathematics and the expressive nature of educational choices may help us understand men's persistence in STEM domains (Charles and Bradley 2009). Just as researchers have cited a close link between cultural conceptions of men and STEM fields as a reason why women might opt out of these fields in order to behave in a gender-consistent manner, the same link may lead men to opt in to them. According to this perspective, men's choices to pursue mathematics and other STEM fields can be understood not only as reflecting instrumental calculations regarding the costs and benefits of STEM activities but also as expressive enactments of a masculine gender identity in an academic context. That is, to the degree that men are passionately devoted to math and other STEM fields (cf. Blair-Loy and Cech 2017), this can potentially be understood not only as a commitment to the field of study or career but also as expressive of their commitment to a masculine gender identity. Importantly, Cech (2013) notes that even where gendered self-expressive motivations drive choices, the individuals making these choices may not understand them as gendered but simply as acts of self-expression. Taken together, this research suggests that when men persist in STEM fields, these choices may—consciously or unconsciously—reflect the value these men place on expressing their masculinity in this domain. Further, when men receive feedback that their STEM performance is insufficient, it is possible that this feedback is interpreted not as confirmation that they do not belong but rather as a challenge to their masculinity that stimulates greater investment in masculinized domains (cf. Willer et al. 2013).

Prior research explaining gender differences in STEM has highlighted the importance of both self-confidence and gendered self-expression, but it is not yet known how these mechanisms may be related. We explore these questions in our third study, where we test different ways in which the key factors presented here—gendered stereotypes of STEM competence, gender differences in STEM confidence, and gender differences in STEM identification—may interrelate in giving rise to gender differences in STEM persistence. It could be that either gender differences in confidence or in domain identification solely explain gender differences in STEM persistence. Alternatively, these factors could interact such that the confluence of high confidence and domain identification drives men to persist at high levels while low levels of both lead women to opt out. Alternatively, these factors might work in parallel, serving as two independent paths explaining gender differences. Additionally, it is important to consider what role gender stereotypes play. Theoretical treatments of both the confidence and identity accounts typically posit that both mechanisms begin with perceptions of widely held stereotypes regarding gender differences in STEM competence, but it has not yet been established if these stereotypes drive these particular mechanisms. We explore all this in Study 3 by testing different ways in which these factors—perceptions of stereotypic differences, gender differences in competence, and gender differences in identification—may interrelate to give rise to STEM differences in persistence. By testing how confidence and identity may work together in a single analysis, this paper addresses Cech's (2013) call for researchers to conduct social psychological research to understand the relationship between the self-expressive and confidence-based explanations of gender differences in STEM.

### *Analytic Strategy*

In examining the role that men's behaviors play in gender gaps in STEM persistence, we focus on three research questions:

*Research Question 1:* Do men overpersist more often than women in math-related academic activities?

*Research Question 2:* Are men more likely than women to persist following failure in STEM classes in American colleges, and how often does this persistence lead to greater professional rewards?

*Research Question 3:* Why do men overpersist?

To examine our first research question, we develop a novel experimental paradigm designed to measure overpersistence in mathematics. In Study 1, we use this experimental setting to test whether undergraduate men choose mathematics more than women in a setting configured such that opting for nonmathematics activities would almost certainly lead to greater success. We examine women's and men's decisions to answer either mathematics or verbal questions when mathematics questions are much more difficult and students are paid for the number of questions they answer correctly. As

we discuss in the following, manipulation checks and supplementary analyses confirm that this experimental paradigm allows us to establish whether men ever overpersist in math, choosing math in contexts where it does not allow them to achieve greater material success (their stated goal), and whether they do so more often than women.

To address our second research question, we examine whether similar processes may operate in college students' course-taking. Specifically, in Study 2, we use data from a nationally representative longitudinal sample to examine women and men's likelihood of retaking any college STEM class after failing it. While these data do not allow us to precisely measure payoffs associated with different choices, we find that men are more likely to retake STEM classes after failing, that this behavior is not on average associated with higher later-life earnings, and that a substantial proportion of men retaking STEM courses have later-life earnings lower than similar men who did not retake STEM courses after failing.

To investigate our third research question, in Study 3 we return to the experimental framework used in Study 1, examining the roles of math confidence and identification in explaining why men choose math more than women using a sample of participants recruited from an online panel. We find evidence that men's overpersistence is driven by gender stereotypes that shape men's relative confidence in and identification with the mathematics domain.

It is difficult to assess precisely when STEM course retaking represents over- or underpersistence with observational data. Payoffs for mathematics persistence are likely to vary across individuals based on labor market conditions, unobserved skill levels, and unmeasured academic and professional consequences. Thus, our experiments are advantageous in that they allow us to fix the payoffs for mathematics persistence to be relatively high or low. We examine STEM course-taking persistence following course failure to see if analogous patterns exist in field settings where decisions have higher stakes and have longer-term consequences. While this study cannot assess with certainty whether students' persistence following course failure was ultimately advantageous or not, it can help us assess whether the results of the experiments are externally valid.

We view our findings as an existence proof of male overpersistence. Establishing whether a given act of persistence is overpersistence, underpersistence, or appropriate persistence necessitates contexts in which the costs and benefits are clearly defined, ideally to both the researcher and those making the persistence decisions. While the experimental framework we employ in Studies 1 and 3 allows us to do this, and Study 2 provides congruent observational evidence suggesting that these processes operate beyond our experimental framework, as an existence proof, we are ultimately unable to establish the degree to which the gender differences observed in STEM representation result from men's overpersistence. But in highlighting the existence of male overpersistence, our findings underscore the need for research on gender differences in STEM persistence to consider both

women's and men's STEM persistence to fully understand the dynamics underlying gender differences in representation in these fields.

### *Study 1: An Experimental Study of Overpersistence in Mathematics*

Our first experimental study explores our first research question, testing whether men choose mathematics in a setting where doing so is very likely to lead to them being less materially successful. To do this, we developed a new experimental paradigm for measuring such overpersistence. Students were given a 10-question test and were informed that they would earn a dollar for every question they answered correctly. Before each question, study participants chose whether to answer a math or verbal question. To orient them to the test, participants were given representative questions before beginning. Participants were randomly assigned to one of two conditions. In the difficult math condition, participants could choose between extremely difficult mathematics questions taken from the GRE subject test in mathematics or moderately difficult verbal questions taken from an SAT verbal section. We also included an easy math condition in which participants could choose between very easy mathematics questions taken from a test of basic mathematics skills given to elementary school teachers or the same, moderately difficult verbal questions. While our primary interest in Study 1 was whether men would choose more mathematics questions in a setting where doing so would likely yield worse test performance (the difficult math condition), we included the easy math condition to explore whether the gender difference in choosing math questions would appear more generally, including when choosing math questions was likely to be beneficial. We predicted that men would be more likely than women to select math questions and that they would do so even in the difficult math condition, where selecting math would very likely lead to less success on the task. To ensure that men and women were similarly motivated to earn as much money as possible, we also asked students questions about the degree to which they were motivated by money and other considerations.

## **Method**

### *Design and Participants*

The study features a 2 (participants were men/women)  $\times$  2 (mathematics questions were very difficult/very easy) experimental design. In all, 190 undergraduate students (81 men, 109 women) took part in the study.

### *Procedure*

Students at a large selective public university were recruited via announcements in a large undergraduate sociology class advertising payment and class credit for participation in an

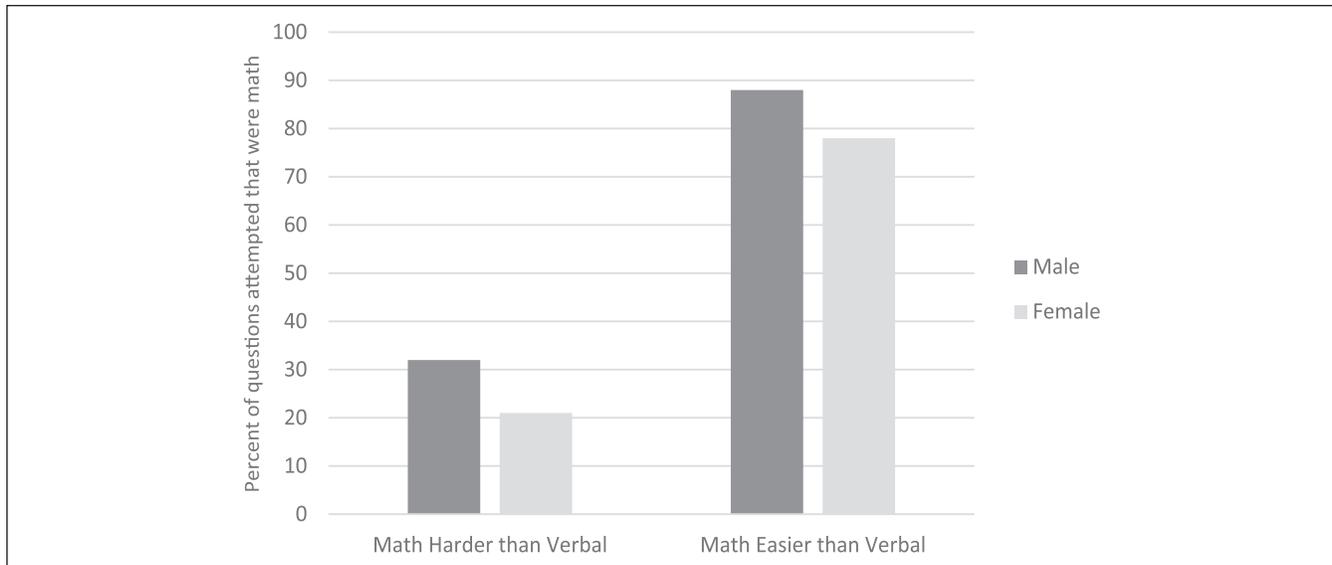
“Academic Choice Study.” After reporting to the lab, participants were seated at one of several cubicles where they completed the study on a computer terminal. The computer program informed participants that they would take a 10-question test and that before each question they would choose whether they would like a math or verbal question. They were told that they would receive one dollar for each question that they answered correctly and were shown two representative examples of both the math and verbal questions to illustrate their relative difficulty before beginning the test.

Participants were randomly assigned to either the difficult math condition or the easy math condition. In the difficult math condition, the math questions came from a practice GRE math subject test and were among the items test takers had least frequently answered correctly. In the easy math condition, the math questions were among the easiest questions on a practice California Basic Educational Skills Test (CBEST), a test given to prospective California elementary school teachers to assess competence in upper elementary school math. In both conditions, the verbal questions came from an official practice SAT and were items that high school students had answered correctly approximately 50 percent of the time.<sup>4</sup> All math and verbal questions were multiple choice with five possible responses. After answering each question, students were informed whether they had answered correctly and how much money they had earned to that point and were asked whether they would like to answer a math or verbal question next.

Following the test, participants completed a poststudy questionnaire measuring how important several considerations were to them in choosing what kinds of questions to answer on the test (“earn as much money as possible,” “finish the test quickly,” “enjoy the test,” “challenge yourself,” “show determination,” and “be persistent”), with participants indicating the degree to which each was important using a scale ranging from 0 to 100. These items allow us to address potential concerns that men and women might differ in their motivations for choosing between math and verbal questions. That is, even though we provided monetary incentives to perform well on the test, it is possible that participants had other goals that influenced their choices of what sort of questions to attempt.<sup>5</sup>

<sup>4</sup>To ensure that students did not choose mathematics questions because they had a more compact form (and thus potentially the appearance of a lighter cognitive load), we used only analogy-based vocabulary items for our verbal questions (e.g., we did not use verbal questions that would have required participants to read a paragraph to answer questions).

<sup>5</sup>As 99 percent of our sample were native English speakers, we believe that it is unlikely that students chose math questions because of a lack of familiarity or comfort with English.



**Figure 1.** Percentage of questions attempted in Study 1 that were math.

## Results and Discussion

Overall, study participants answered the difficult math questions correctly just 12 percent of the time, much less often than the easy math questions, which were answered correctly 85 percent of the time ( $t = 24.65, p < .001$ ). Verbal questions were answered correctly 40 percent of the time, significantly more often than the difficult math questions ( $t = 7.34, p < .001$ ) and significantly less often than the easy math questions ( $t = 6.74, p < .001$ ). Male participants correctly answered slightly more difficult math (14.3 percent) and easy math (86.0 percent) questions than female participants (10.7 percent and 84.4 percent, respectively), though neither of these differences is statistically significant ( $t = .76, .40, p = .45, .69$ ). Male participants also answered slightly more verbal questions correctly than female participants (42.8 percent compared to 38.5 percent), though this difference is again not statistically significant ( $t = .93, p = .35$ ). Thus, we concluded that we had successfully constructed the tests in such a way that the difficult math questions were rarely answered correctly, the easy math questions were very often answered correctly, and the frequency of correctly answering the verbal questions was between these two, with no gender differences in the likelihood of answering any of the items correctly.

Turning to our predictions regarding gender and math, we explored whether men and women differed in how frequently they chose to answer mathematics versus verbal questions in the study. We first looked at choices in the study as a whole, conducting an ANOVA of the effects of participant's gender and experimental condition on the percentage of math questions participants attempted. That model reveals significant effects of experimental condition,  $F(1, 186) = 196.03, p < .001$ , indicating that overall, participants attempted significantly

more mathematics questions in the easy mathematics condition. More relevant to our hypotheses, we also found a main effect of participant's gender,  $F(1, 186) = 6.85, p = .01$ , reflecting the fact that men tended to choose more mathematics questions than women did in the study. We found no significant interaction of gender and experimental condition,  $F(1, 186) = 0.02, p = .89$ , indicating that the gender gap in math questions attempted did not vary by condition. These results show that across the study as a whole, men chose to complete more mathematics questions than women.

Next we looked at results within each experimental condition specifically. As shown in Figure 1, male participants chose to answer math questions more often (32.3 percent) than female participants (21.2 percent) ( $t = 2.07, p = .04$ ) in the difficult math condition. This result is consistent with our prediction that men would choose mathematics questions more often than women even when those questions were quite difficult, meaning that choosing them was likely to lead to worse test performance.<sup>6</sup> We also tested whether a gender gap emerged in the easy mathematics condition, finding that men there also chose more mathematics questions (87.6

<sup>6</sup>To ensure that the gender differences that we observe are not being driven by respondents' choices on the first few questions, we conduct supplementary analyses in which we interact gender and question number. We operationalize question number in three different ways: (1) a linear term, (2) a dummy variable for the first and second half of the test, and (3) dummy variables for each question number (using an  $F$  test to test for the joint significance of these interaction terms). As none of our tests for the interaction of the question number and gender are significant, we conclude that the gender differences we observe do not vary significantly across the test.

percent) than women (77.7 percent), though this effect is only marginally significant ( $t = 1.67, p = .10$ ).

In sum, in Study 1, we found that male participants were significantly more likely to choose mathematics questions than their female counterparts even when the mathematics questions were substantially more difficult than alternative, verbal questions. Participants' self-reported motivations for choosing math versus verbal questions in the study do not account for these differences: Both male and female participants rated earning as much money as possible as more important than the other motivations (all  $ps < .001$ ), and the only significant gender difference in motivation was that men placed greater importance on earning as much money as possible ( $t = 2.28, p = .02$ ). Thus, both men and women's primary concern in choosing questions was to perform well on the test and maximize their earnings in the study, and gender differences in other motivations are unlikely to account for the gender gap in choosing difficult math questions that we observe. If anything, since men reported being more concerned with earning money on this task, based on reported motivations alone, men would be expected to choose fewer math questions than women when those questions were very difficult. Yet, consistent with our argument, we find that a gender difference exists in both conditions when choosing math led to earning more money and when it led to earning less. The condition where choosing math leads to greater material success is congruent with researchers' typical assumptions regarding the payoffs of STEM, and in this context, men choose more math questions than women, a familiar pattern in which women are said to underpersist. But we also find evidence that men choose math questions even in a context where this does not lead to greater material success, a less familiar pattern in which men could be said to overpersist.

### *Study 2: College STEM Course Persistence in the United States*

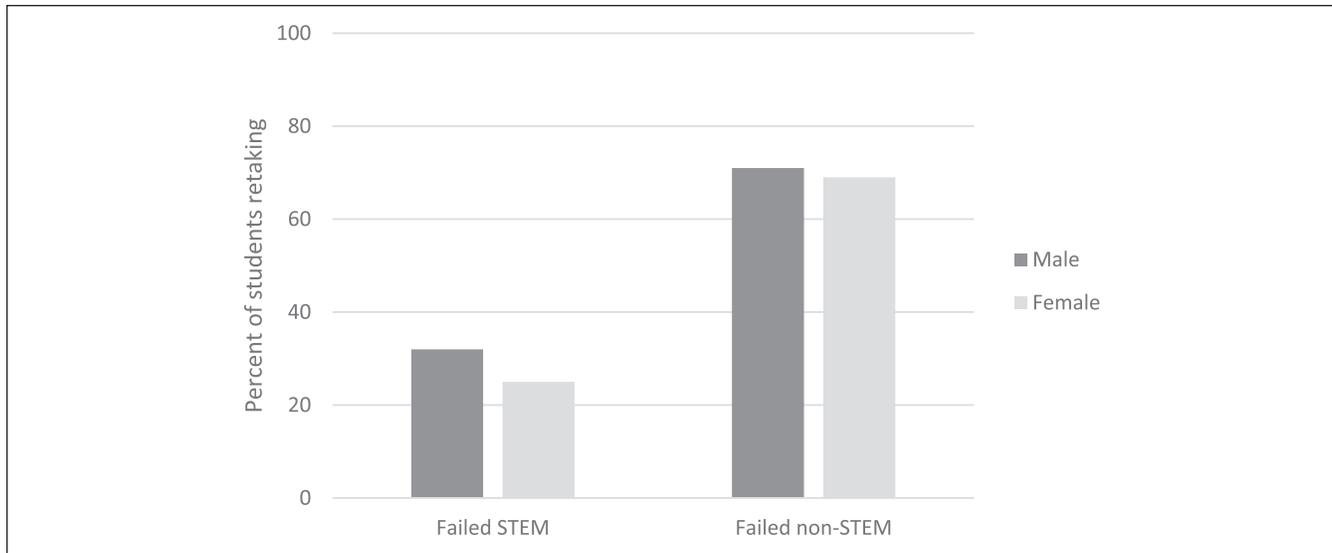
Although Study 1 allowed us to establish the existence of overpersistence by carefully crafting a situation in which mathematics persistence was likely to lead participants to be less successful, it is difficult to know whether the findings are generalizable to other tasks and contexts. By contrast, in observational data, it is much more difficult to determine whether persistence led to a less desirable outcome than opting out would have. However, analysis of observational data allows us to examine contexts that have important consequences for students' futures and can also provide insight into whether the pattern we observe might hold in policy-relevant arenas beyond the lab. Thus, Study 2 was designed to explore our second research question by examining whether we find evidence of overpersistence in college course-taking and whether this is more common among men. We seek to identify a context in which taking STEM classes might not be associated with advantages later in life and thus examine whether women and men retake STEM courses at similar rates after failing.

To address our second research question, we examine transcript data from the Post-secondary Educational Transcript Study (PETS) of the National Education Longitudinal Study (NELS) of 1988. NELS follows a nationally representative cohort of eighth graders beginning in the 1987–1988 school year, and the PETS data were collected in 2000, when most respondents were between 26 and 27 years old. The PETS data include information on over 8,000 students and contain transcripts from roughly 3,200 institutions (see Adelman, Daniel, and Berkovits 2003 for more information on the PETS sample).<sup>7</sup> We focus on the 2,620 students who failed STEM courses and examine the likelihood that students subsequently attempted to retake the class that they failed.<sup>8</sup> However, to confirm that our results are not reflective of responses to course failure more generally, we also examine the likelihood of subsequently retaking courses after failing non-STEM courses. For both STEM and non-STEM courses, we estimate a series of logistic regression models predicting whether students who fail a course subsequently retake it. In addition to gender (our independent variable of interest), our models include controls for students' GPA from their high school transcript; a series of dummy variables for the most advanced math course they took in high school; standardized test scores from high school math, science, English, and history tests; a standardized parental socioeconomic status composite index created by NELS that includes parental education, occupation, and income information; self-reported race; and indices measuring math and verbal confidence (created following Correll 2001).<sup>9</sup>

<sup>7</sup>National Education Longitudinal Study (NELS) analyses use survey weights and account for the stratified sampling frame. As NELS is a nationally representative sample, we do not have sufficient numbers of students at the same institutions to include fixed effects for universities. Our population of interest consists of students who failed STEM classes and are thus faced with choices regarding whether to retake the class; as such, we do not attempt to account for the selection processes determining who fails. Research examining calculus failure finds no evidence of gender differences in the likelihood of failing introductory calculus (Sanabria and Penner 2017).

<sup>8</sup>For all classes, we utilize the NELS definition of failing a class as receiving a noncredit bearing grade (e.g., D or F).

<sup>9</sup>Math and verbal confidence were each measured via students' degree of agreement with three statements (e.g., "I have always done well in math" for math confidence; "I learn things quickly in English" for verbal confidence) on 6-point scales ranging from false to true (cf. Correll 2001). We created composite measures of math and verbal confidence by averaging the items from each battery. The reliability of these scales was high (Cronbach's alpha = .91 and .85, respectively). We would have preferred to use measures of confidence that were collected closer in time to when students were making decisions about whether to retake courses; however, these questions were only asked when students were in 10th grade. Ideally, we would also have been able to examine other mechanisms, such as gender identification and math identification, but measures of these factors were not available in any wave of the NELS.



**Figure 2.** Percent of students who retook STEM and non-STEM courses after initially failing (Study 2).

The longitudinal nature of these data allows us to examine whether there are differences in income (measured eight years after high school) that are associated with having retaken STEM coursework in college. This is important as students who chose to retake a STEM class after failing had a relatively high likelihood of passing (64 percent), and while we find no gender differences in the likelihood of passing among the students retaking classes ( $p = .98$ ), it is possible that retaking a STEM course after failing is associated with positive longer-term outcomes, in which case course retaking should not be conceptualized as overpersistence. To better understand how likely students are to benefit from retaking the STEM (and non-STEM) courses they fail, we use propensity score matching (Rosenbaum and Rubin 1983) to examine not only the average difference in logged income between students who did and did not retake courses after failing but also the probability that students who retook courses had an income higher than their propensity score matched counterfactual. Intuitively, we can think of this as matching each individual who failed and then retook a STEM class with an observationally similar (i.e., similar high school GPA, math course-taking, test scores, parental background, and math and verbal confidence levels) student who failed a STEM class but decided not to retake the course. We then examine the percentage of cases where the student who retook the course had a higher, lower, or similar income relative to their propensity score matched counterpart. While we are unable to account for unobserved differences between students who did and did not retake STEM classes, this analysis provides some indication as to whether and how often students experience longer-term benefits from retaking courses.

Like introducing controls into regression analyses, propensity score matching seeks to account for observable differences between groups (in this case, students who did and did not retake courses after failing). Propensity score matching

accounts for observable differences by estimating the likelihood that students who failed courses retook them, and it uses this estimate to match individuals who retook courses after failing with other individuals who had similar likelihoods of retaking but did not retake the class after failing. These observationally similar non-retakers provide an estimate of the counterfactual outcome for the students who did retake the course. All models were estimated separately by gender. Given our interest in whether retaking STEM courses may indicate overpersistence, we focus on the treatment on the treated (TOT) estimate and compare students who retook a course to a matched student who did not retake the course.<sup>10</sup>

It is unclear whether one should expect retaking courses after initially failing to be associated with better or worse longer-term outcomes. On the one hand, students who retake courses are displaying persistence and determination, qualities associated with positive academic and professional outcomes (cf. Duckworth et al. 2007). Alternatively, retaking courses after failing might indicate a failure in adaptive goal disengagement (cf. Heckhausen 1997) and thus might be associated with negative longer-term outcomes. It is also possible that retaking might be beneficial for some students and not for others, in which case, retaking a course after failing might constitute overpersistence for some students but not others.

## Results and Discussion

Figure 2 reports the percentage of men and women who retook STEM (and non-STEM) courses after initially failing them. We found that among the 2,620 students who failed a

<sup>10</sup>We do not control for occupation in these models because we view occupation as a mechanism through which course retaking might affect income.

**Table 1.** Results from Logistic Regression Models Predicting Retaking STEM and Non-STEM Courses among Students Who Previously Failed (Study 2).

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
	STEM	STEM	STEM	STEM	Non-STEM	Non-STEM	Non-STEM	Non-STEM
Male	.31*	.33*	.29*	.32*	-.04	.02	.01	.12
Asian	.31	.34	.35	.35	-.04	-.01	-.06	.03
Hispanic	.23	.16	.18	.17	.23	.08	.09	.10
Black	.24	.29	.28	.26	.54*	.21	.18	.28
Native American	-.33	-.40	-.45	-.18	-.28	-.32	-.36	-.25
Socioeconomic status	.22	.21	.22	.22	.17	.14	.14	.08
Math test		.02	-.01	-.01		-.02	-.02	-.01
Science test		.00	.01	.00		.00	.00	.01
English test			-.01	-.01			.00	.00
History test			.00	.00			.00	-.01
High school GPA			-.15	-.15			-.22	-.21
Math confidence				-.04				.01
Verbal confidence				.07				.01
Highest high school math course		X	X	X		X	X	X
N	2,600	2,520	2,510	2,420	3,710	3,590	3,580	3,440

Note: Models 1 through 4 present coefficients from logistic regression models predicting the odds of having retaken a STEM class among students who have ever failed a STEM class. Models 5 through 8 present coefficients from logistic regression models predicting the odds of having retaken a non-STEM class among students who have ever failed a non-STEM class.

\* $p < .05$ .

STEM class, 32 percent of men retook the class at a later date compared to 25 percent of women. This difference is statistically significant ( $p = .02$ ). To rule out that men are simply more likely to retake classes that they fail across all subjects, we also examined whether this pattern exists in non-STEM courses. Here we did not find a statistically significant gender difference ( $p = .53$ ); men retake non-STEM classes after failing 71 percent of the time compared to 69 percent for women. Comparing STEM and non-STEM course retaking is important because women and men could face differences in the costs and benefits of retaking a course after failing that are common across all courses (e.g., women may not retake courses after failing because they had long-term family responsibilities that were unlikely to change from term to term, while men might be more inclined to retake if their failure was due to factors in their control, like insufficient time spent studying). Finding a different pattern of results for retaking STEM and non-STEM courses suggests that men’s decisions to retake STEM courses are not solely attributable to broader gender differences in why students fail courses.

We build on these results by estimating logistic regression models predicting the odds of retaking STEM and non-STEM courses with a variety of controls to examine whether any of these variables can help us understand these gender differences. Model 1 in Table 1 analyzes gender differences in students’ odds of retaking STEM courses net of race and family SES. Results from this model indicate that girls have lower odds of retaking STEM courses. In Model 2, we introduce math and science test scores, finding that the gender difference in STEM course retaking persists when controlling for measures of math and science achievement.

In Model 3, we additionally introduce controls for English and history test scores to address concerns that women might be less likely to retake STEM classes in part because they have stronger skills in non-STEM fields than men (cf. Wang, Eccles, and Kenny 2013), which could make the opportunity costs of retaking a STEM course higher. We find that the gender difference in STEM retaking is robust to controlling for these measures. Finally, in Model 4, we introduce controls for math and verbal confidence, finding that controlling for math and verbal confidence does not eliminate the gender difference in course retaking behavior.<sup>11</sup>

Models 5 through 8 of Table 1 replicate Models 1 through 4 but present results from models predicting the odds of retaking a non-STEM course after failing. Importantly, we did not find statistically significant gender differences in the odds of retaking non-STEM courses in any of these models, suggesting that the gender differences observed are unique to STEM courses.

Although these course-taking data preclude the more complete understanding of the costs and benefits provided by our experimental framework, we can nonetheless examine whether there were any differences in longer-term earnings between those who did and did not retake STEM courses after failing. Table 2 presents results from propensity score matching analyses examining the earnings differences associated with

<sup>11</sup>The primary coefficient of interest examining men’s course retaking behavior in Table 1 does not vary significantly across Models 1 through 4 or Models 5 through 8. Estimating the same models on a consistent sample (i.e., using the Model 4 sample for Models 1–4 and the Model 8 sample for Models 5–8) yields a similar conclusion.

**Table 2.** Long-term Income Differences Associated with Retaking Classes Relative to Propensity Score Matched Counterfactual among Students Who Failed Classes (Study 2).

	STEM		Non-STEM	
	Female	Male	Female	Male
Panel A. Average income difference between retakers and matched counterfactuals				
Retakers	.70	.13	.00	.13
N	990	1,070	1,460	1,440
Panel B. Retakers' income relative to matched counterfactual (%)				
Retakers earning less than 90%	43.1	44.8	44.9	44.8
Retakers earning between 90% and 110%	7.9	9.9	11.5	10.8
Retakers earning greater than 110%	49.1	45.4	43.6	44.3

Note: No differences were statistically significant at the .05 level. Propensity score models adjust for race, a parental socioeconomic status composite measure, high school test scores, high school GPA, the highest math course taken in high school, and math and verbal confidence. Average income differences report exponentiated coefficients from models predicting logged wage, which can be interpreted as proportional differences.

retaking STEM (and non-STEM) courses for men and women. The first two columns present information about the men and women who retake STEM classes after failing them, and the second set of columns reports results from non-STEM classes. Panel A presents the average income difference between those who did and did not retake a course after failing. We find that men who retook a STEM course after failing earned on average 12 percent more than men who did not retake the class, but this difference is not statistically significant ( $p = .52$ ). By contrast, women who retook a STEM class after failing on average earned 70 percent more than women who did not, a marginally significant difference ( $p = .052$ ). Neither men nor women who retook non-STEM courses earned significantly more on average than their matched counterparts who did not.

While these average differences are informative, they do not provide information about the proportion of men and women who might have earned more had they not retaken the course. To better understand how many students who retake a failed course end up with worse outcomes than would be expected if they had not retaken the course, we examine the proportion of students who earned at least 10 percent less than their propensity score–matched counterfactual; we also examine the percentage who earned at least 10 percent more than their propensity score matched counterfactual. This information is given in Panel B of Table 2. Among men who failed and retook STEM courses, we see that 44.8 percent earned at least 10 percent less than their propensity score–matched counterfactual. This is effectively the same as the 45.4 percent of students who earned at least 10 percent more than their counterfactual, suggesting that men who retook STEM courses after failing were essentially as likely to earn less than their counterfactual as they were to earn more.<sup>12</sup> By contrast, among women who failed a STEM

course, 49.1 percent of students who retook the course earned at least 10 percent more than their matched counterfactual compared with 43.1 percent who earned at least 10 percent less, though this difference was also not statistically significant. The longer-term pay differences associated with retaking non-STEM courses are similar for men and women, with results suggesting that both are as likely to earn more as to earn less by retaking a failed non-STEM course.

Study 2 suggests that men's greater persistence in STEM domains—originally documented for mathematics persistence in Study 1—is present in college-level course-taking. Importantly, we found that men who retook STEM courses did not receive higher earnings, though we did find evidence suggesting that women who persisted in the face of failure in a STEM course may have benefited. These findings are consistent with past research suggesting women may benefit by persisting in STEM fields (cf. Oh and Lewis 2011) as well as the notion that men's persistence may not yield the desired benefits. Further, our analyses show that a substantial proportion (45 percent) of men who retook STEM classes after failing had lower incomes than similar men who failed STEM classes but did not retake the course. We argue that examining how students respond to failing a course provides insight into gender differences in persistence in STEM fields. In particular, it is noteworthy that we found evidence that men were more likely to retake STEM classes after initially failing them but that they did not exhibit the same pattern for non-STEM classes. Finding that men persisted in STEM classes more than women but retook non-STEM courses at similar rates as women suggests that men's STEM behavior differs from their behavior in other fields.

We found no evidence that levels of self-reported math confidence mediated the link between gender and persistence (cf. Correll 2001). This was perhaps in part because math confidence was measured in high school, several years before students' subsequent persistence following failure in college courses. In addition, the NELS data set does not contain any measure of other potential mediating variables, such as gender identification or mathematics domain identification. Thus, to investigate our third research question, we designed an additional study to address

<sup>12</sup>Supplemental analyses find that of the men earning within 10 percent of their counterfactual, slightly more earn less than their matched counterfactual so that overall, men who retake STEM classes are, if anything, slightly more likely to earn less (49.0 percent) than they are to earn more (48.7 percent).

these shortcomings and examine potential mediating processes by more extensively measuring our hypothesized mediating variables and importantly, assess these potential mediating variables at a closer point in time to our measure of persistence.

### *Study 3: An Analysis of Factors Driving Men's Overpersistence*

Our first two studies offer evidence that men tend to persist more in STEM fields than women even in the face of failure and in an experimental setting where such choices are designed to lead to worse performance. However, these studies offer little insight on the factors that might explain these differences. Previously, we reviewed prior research on two theoretical mechanisms that might account for men's behavior in STEM fields, both of which are thought to result from widely held cultural beliefs about men's superiority in STEM fields. One mechanism argues that men's greater STEM persistence is driven by their greater confidence in their STEM abilities (cf. Correll 2004), while the other views STEM persistence by men as a result of their enacting a male gender identity (Charles and Bradley 2009).

Study 3 was designed to assess whether one or both of these mechanisms might drive the gender gaps that we observe. In this study, we return to the experimental paradigm of Study 1. We added several survey measures to the study to assess possible mechanisms for persistence in this paradigm, including measures of participants' confidence in their math abilities, gender identification, mathematics domain identification, and the perception that mathematics is a domain in which men necessarily perform better than women.<sup>13</sup> We also sought to establish that the gender difference found in Study 1 would be obtained in a more diverse sample by running the study with a sample of participants recruited online that was more heterogeneous with respect to age, education, and income than Study 1's student sample. Further, to ensure that choosing mathematics questions would correspond to overpersistence, we increased the difference in difficulty between the verbal and mathematics questions by using a new set of verbal questions that were substantially easier than those used in Study 1.

## **Method**

### *Design and Participants*

The design of the study was the same as Study 1 with a few exceptions: (1) We no longer included a condition of the

<sup>13</sup>As participants were asked questions about their math abilities, attitudes, and identity, this study may have primed these constructs, affecting participants' choices. As such, we view Study 1, which does not feature these questions, as being more definitive in establishing the existence of male overpersistence, while Study 3 (which finds similar patterns as Study 1) allows us to understand the correlates of this behavior.

study in which the math questions were substantially easier than the verbal questions. Instead, all participants were given a test in which they could choose between very difficult math questions or very easy verbal questions. (2) The study was conducted online using a sample of participants recruited via an advertisement posted on Amazon Mechanical Turk (AMT).<sup>14</sup> Additionally, (3) several additional survey measures were added to the study.

In all, 852 participants (398 men, 454 women) took part in the study.<sup>15</sup> All participants were US residents ranging in age from 18 to 73 ( $M = 33.87$  years,  $SD = 11.69$ ). Six hundred and forty-one participants identified as white (75.2 percent), 72 as Asian (8.5 percent), 60 as Latino (7.0 percent), 41 as black (4.8 percent), and 38 indicated another or mixed race (4.5 percent). The median respondent earned between \$35,001 and \$50,000, and 48.6 percent of respondents had at least a college degree.

### *Procedure*

The study was conducted online. Participants were recruited to an "Academic Choice Study" via an advertisement posted on AMT promising a small base payment plus the opportunity to win a \$100 bonus. Participants first completed a demographic questionnaire followed by a series of survey batteries. Levels of math identification were assessed via participants' degree of agreement on scales ranging from 1 (strongly disagree) to 7 (strongly agree) with a series of three statements (e.g., "My math ability is an important reflection of who I am") adapted from a standard measure of identification (Luhtanen and Crocker 1992). Reliability for this scale was high (Cronbach's  $\alpha = .86$ ), so these items were averaged to create a composite. Gender identification was measured via agreement with statements designed to fit with the participant's reported gender (e.g., "Being a man [woman] is unimportant to my sense of what kind of person I am" [reverse-coded]) adapted from the same standard measure of identification. Reliability for the items of this composite were moderate among men and women (Cronbach's  $\alpha$ 's = .68, .75, respectively). We measured math confidence via participants' average agreement on the

<sup>14</sup>Amazon Mechanical Turk (AMT) is an online marketplace in which users complete short computer-based tasks for money. AMT is used as a source for research subjects in political science, psychology, economics, and sociology. The worker pool is demographically diverse but not representative of the US population. Studies find that the quality of data collected via the platform is typically equal to or greater than data collected in person or via other online sources (Buhrmester, Kwang, and Gosling 2011; Weinberg, Freese, and McElhatten 2014).

<sup>15</sup>This number and all analyses exclude 33 participants who incorrectly answered either of the two items included to assess whether participants were paying attention (e.g., "For this response, answer 'Agree'"). All included participants answered both of these items correctly. Exclusion of these participants had no substantive effect on results presented here.

same 7-point scales with a series of three statements (e.g., “I am good at math”); Cronbach’s  $\alpha = .98$ ).

We also measured participants’ levels of risk-taking via degree of agreement on 7-point scales with five statements (e.g., “I like to take chances”) culled from a standard battery (Patenaude and Laufersweiller-Dwyer 2002; Cronbach’s  $\alpha = .94$ ). Finally, we assessed the extent to which participants endorsed a view that math is a male domain by their average agreement on 7-point scales with a series of 12 statements from Fennema and Sherman’s (1976) Mathematics as a Male Domain scale (e.g., “I would have more faith in the answer for a math problem solved by a man than a woman,” “It’s hard to believe that a woman could be a genius at mathematics”); Cronbach’s  $\alpha = .91$ ).

Following completion of these survey items, participants were introduced to the test. As in Study 1, participants were told that they would complete 10 questions and that before each they could choose whether the next question would be math or verbal. Participants were told that for every question they got right, they would receive an entry in a raffle drawing for a \$100 prize. Past research finds that raffle-based incentive systems of this sort are motivating to participants (e.g., Feinberg, Willer, and Keltner 2012). As in Study 1, following the test, participants completed a post-study questionnaire, including a measure of how important a series of considerations was to them in choosing what kinds of questions to answer on the test (“earn as many raffle tickets as possible,” “finish the test quickly,” “enjoy the test,” “challenge yourself,” “show determination,” and “be persistent”). Consistent with Study 1, participants overwhelmingly reported being motivated by the pecuniary payoff (in this case, raffle tickets).

The mathematics questions were again selected from the GRE math subject test and were very difficult. We constructed the verbal questions ourselves to be very easy (e.g., “The dance instructor \_\_\_\_\_ upon her best student the honor of teaching the class for the week she was absent. Answer options: (A) Tailored, (B) Bestowed, (C) Retracted, (D) Created, (E) Manipulated”).<sup>16</sup> As in Study 1, all math and verbal questions were five-item multiple choice questions. Before beginning the test, participants were shown four example questions (two each of math and verbal) representative of the difficulty of each set of questions. Following the test, participants were thanked and paid a flat rate for their participation. The bonus drawing was conducted after completion of the study, and the winner was paid the additional amount.

<sup>16</sup>As in Study 1, we sought to ensure that the mathematics questions did not appear to have a lighter cognitive load. As such, we used only sentence completion items for verbal questions and did not use any verbal questions that would have required participants to read a paragraph to answer questions based on the text they read.

## Results and Discussion

We first analyzed participants’ test performance. Overall, math questions were rarely answered correctly (15.9 percent), even less often than would be expected from random guessing, as in Study 1. Participants correctly answered verbal questions (63.1 percent) significantly more often ( $t = 19.20, p < .001$ ). Male participants correctly answered slightly more math (16.6 percent) and slightly fewer verbal (61.7 percent) questions than females (15.1 percent and 64.3 percent, respectively), though neither of these differences was significant ( $t = .66, 1.17, p = .51, .24$ ).

Turning next to choices to answer math or verbal questions, overall participants chose verbal questions (84.1 percent) much more often than math questions (15.8 percent) ( $t = 47.46, p < .001$ ).<sup>17</sup> We found that male participants chose to answer significantly more math questions (18.3 percent) than female participants (13.6 percent) ( $t = 3.33, p = .001$ ), consistent with our hypothesis and the results of Study 1.<sup>18</sup>

We found gender differences for all of the survey batteries we measured. Male participants reported significantly greater math confidence ( $M = 4.59$ ) than female participants ( $M = 4.05$ ) ( $t = 4.47, p < .001$ ). Males reported significantly greater identification with the math domain ( $M = 3.91$ ) than females ( $M = 3.60$ ) ( $t = 2.87, p < .01$ ). Male participants also reported a greater perception that math is a male domain ( $M = 2.17$ ) than females ( $M = 1.85$ ) ( $t = 5.25, p < .001$ ). Finally, we found that women in the study reported higher levels of gender identification ( $M = 5.54$ ) than men ( $M = 5.34$ ) ( $t = 2.51, p = .01$ ).

Next we assess our theoretical arguments. First, we tested whether math confidence, found to be significantly higher among men in our study, might mediate the gender difference in persistence. Model 1 of Table 3 gives results from our baseline ordinary least squares (OLS) model of the association between gender (with males coded as 1) and number of math questions attempted. As previously described, we found that men attempted significantly more math questions than women. Model 2 adds our measure of math confidence. Results for this model show that math confidence is a highly significant predictor of the number of math questions that respondents attempted. Further, the significance and magnitude of the effect of gender is greatly diminished in this model, suggesting that the gender difference in persistence operated at least partly through levels of confidence.

<sup>17</sup>Due to a computer error, in seven instances (representing .08 percent of 8,520 participant choices) participants advanced to the next question in the test without making a choice to answer a math or verbal question. All results are robust to dropping these cases.

<sup>18</sup>Note that although men reported significantly greater risk-taking ( $M = 4.10$ ) than women ( $M = 3.62, t = 5.53, p < .001$ ), the effect of gender on the number of math questions participants chose to answer remained highly significant ( $B = .421, p < .01$ ) and in the same direction after controlling for the battery measure of risk-taking ( $B = .118, p = .04$ ).

**Table 3.** Results of Ordinary Least Squares Models Analyzing Variables That Mediate and Moderate the Effect of Gender on Math Persistence in Study 3.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
	Persistence	Persistence	Persistence	Persistence	Persistence	Persistence	Confidence	Identification	Persistence
Male	.48**	.29*	.32	.35*	.29*	-.25	-.69*	-.44	-.001
Math confidence		.35***			.21***				.20***
Gender identification			-.14						
Male × Gender identification			.02						
Math identification				.40***	.24***				.24***
Math as a male domain						-.14	-.35**	-.02	-.06
Male × Math as a male domain						.35*	.62***	.35**	.14
Constant	1.36***	-.05	2.12***	-.06	-.36	1.61***	4.69***	3.64***	-.22
N	852	852	852	852	852	852	852	852	852

Note: Models 1 through 6 and Model 9 present unstandardized coefficients from ordinary least squares regression models predicting math persistence. Models 7 and 8 present unstandardized coefficients from ordinary least squares regression models predicting math confidence and math identification, respectively. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

We next tested the argument that men’s STEM persistence can be thought of as an identity-based process in which men perform masculinity by participating in a male-typed field. If this argument is true, and assuming that individuals who identify more strongly with their genders are more motivated to behave in gender identity-consistent ways, then we would expect a greater gender gap in math persistence among those participants who identify more strongly with their respective genders. We tested this in Model 3 by adding to the baseline model terms for gender identification and the interaction between gender and gender identification. Results for Model 3 showed that none of the terms in the model testing this account were statistically significant. Further, we analyzed whether gender identification and number of math questions attempted were correlated among both male and female participants separately, finding no correlation among men ( $r = -.06, p = .24$ ) and only a marginally significant correlation among women ( $r = -.08, p = .06$ ) in the study. Thus, these analyses offered no evidence that men in the study who identified more strongly as male answered more mathematics questions.

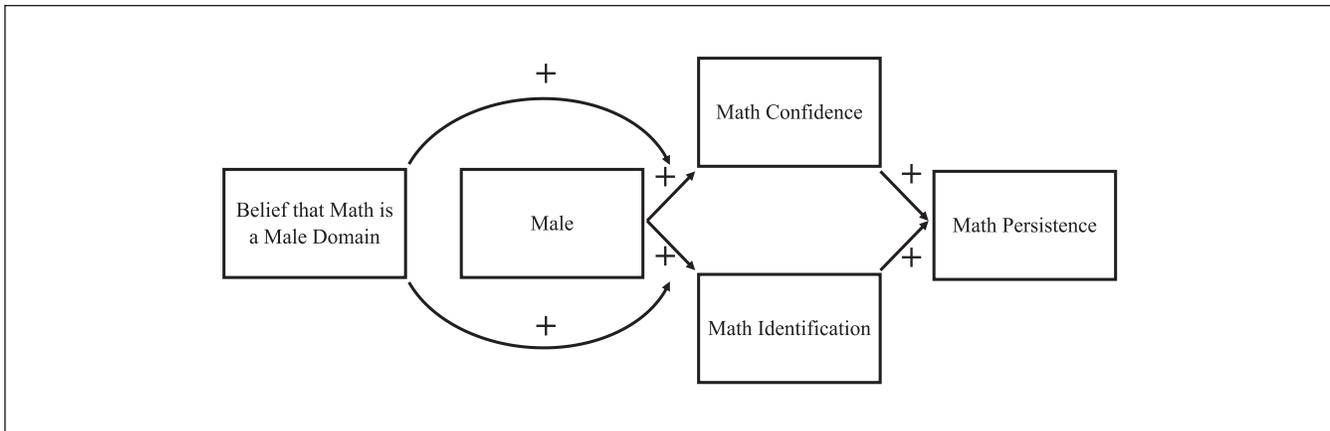
However, there are other ways in which men’s math persistence could reflect an enactment of masculinity. Cech (2013) notes, for example, that gendered self-expression may operate through self-conceptions that do not appear gendered to individuals. This suggests that men might identify with masculine domains and enact masculinity as an expression of this identification. If this is the case, we would expect activity in a gendered domain like math to be driven in part by identification with that gendered domain. Consistent with this reasoning, we found previously that men exhibited significantly higher levels of math identification. To assess whether the gender difference in math identification partly drove the gender difference in math persistence that we observed, in Model 4, we added the measure of mathematics identification to the baseline model. Results of this model

show that math identification is also a highly significant predictor of the number of math questions attempted, and its inclusion reduced the effect of gender on math persistence.

Models 2 and 4 give evidence that both math confidence and math identification partially mediate the link between gender and math persistence in the study. But are these distinct causal paths? To assess this, we estimated a model including terms for both math confidence and math identification. The results of Model 5 support the view that both factors significantly and independently mediate the gender difference in math questions attempted in the study. Note also that gender remains a significant, though weaker, predictor in this model, suggesting that these factors do not completely account for effect of gender in the model.

Both the confidence and identification theoretical mechanisms are related to the notion that mathematics is seen as a domain in which men are generally superior to women as this view is a likely cause of men’s greater confidence in and identification with the mathematics domain. If endorsement of the belief in male math superiority plays a role in gender differences in persistence, then we would expect a greater gender difference in persistence among those participants who hold the view more strongly. To test this, we estimated a model in which we added terms for the belief that math is a male domain as well as the interaction of gender and the degree of this belief to the baseline model. Results of Model 6 show that this interaction term was significant. The effect of gender on the number of math questions participants attempted was greater among participants with greater belief that math is a male domain.

Taken together with the aforementioned analyses, we thus find that the effect of gender on persistence (1) operates through both math identification and math confidence and (2) is moderated by the degree of belief that math is a male domain. However, these mechanisms are presumably inter-related as the belief that math is a male domain is thought to



**Figure 3.** Integrated theoretical model from Study 3.

precipitate higher levels of confidence and domain identification among men. We trace these causal relationships in the theoretical diagram of Figure 3. To test whether mathematics confidence resulted in part from endorsement of the cultural view that men are superior to women at mathematics, we next estimated a model in which terms for gender, the belief that math is a male domain, and their interaction predicted participants' mathematics confidence (Model 7). Results showed a significant interaction between gender and the belief that math is a male domain such that the gender gap in mathematics confidence was greater among participants who more strongly endorsed the view that math is a male domain. Model 8 gives the corresponding model with math identification as the outcome variable. Here we found a similar significant interaction effect. The gender difference in mathematics identification was greater among those participants who more strongly endorsed the view that math is a male domain.

Finally, we estimated a model testing this integrated “mediated moderation” model. In Model 9, we add terms for math confidence and identification to Model 6. If the effect of believing that math is a male domain on the gender difference in persistence operates through levels of math confidence and identification, then in this model we would expect the confidence and identification terms to be significant but the interaction of gender and the belief that math is a male domain to be insignificant (as its effects are mediated by confidence and identification). This is exactly the pattern we find, offering evidence for the integrated theoretical model of Figure 3.<sup>19</sup>

These results advance our understanding of the overpersistence effect in several ways. First, they offer evidence

about the mediating processes driving the tendency for men to overpersist more than women in this setting. Here we found that men's greater math confidence and mathematics domain identification partly accounted for men's greater persistence. Second, we found that the gender gap in persistence was greater among those who endorsed the cultural belief that men are generally better at mathematics than women. Third, our analysis further supports a novel, integrated model of the theoretical arguments based on confidence and gender performance that motivated our research. Together, our results suggest that cultural stereotypes about gender and math ability lead men to develop greater confidence in and identification with the mathematics domain and that these factors influenced their overpersistence in STEM fields.

## General Discussion

The dominant approach to explaining gender gaps in the STEM pipeline views women's failure to persist in these fields as aberrant while treating men's behavior as normal. Although evidence exists that women underpersist in these fields, opting out when further effort could lead to success, here we sought to evaluate the opposite possibility: that men—buttressed by cultural beliefs that they are highly competent in STEM fields—often overpersist in these domains, continuing even when opting out could lead to greater success. If this were the case, it would suggest that the gender gap in STEM fields is driven by the behavior of both men and women, each of whom are influenced by cultural beliefs about gender and abilities in academic domains.

upper limit = .22) and identification (lower limit = .03, upper limit = .16) did not include zero, indicating both variables independently mediated the moderating effect of belief in math as a male domain on number of math questions attempted. This test is preferable to classical tests in this context as it allows us to formally test for conditional indirect effects using bootstrapping rather than relying on asymptotic test statistics in small samples.

<sup>19</sup>We formally test our mediated moderation model using a bootstrapping analysis of mediated moderation (Preacher, Rucker, and Hayes 2007). Results of this analysis indicated that the 95 percent confidence interval for both confidence (lower limit = .05,

We presented three research questions motivating our empirical studies:

*Research Question 1:* Do men overpersist more than women in math-related academic activities?

*Research Question 2:* Are men more likely than women to persist following failure in STEM classes in American colleges, and how often does this persistence lead to greater professional rewards?

*Research Question 3:* Why do men overpersist?

Here we review our findings regarding each of these questions in turn.

It is difficult to know when persistence in STEM activities represents overpersistence, underpersistence, or an appropriate amount of persistence to maximize material success. Thus, to rigorously assess whether men ever overpersist in STEM fields, in Study 1, we developed an experimental setting in which choosing math would almost certainly lead to less material success. In this setting, we found that men were more likely to attempt extremely difficult math questions than women, who opted to answer much easier verbal questions instead. Importantly, men's behavior was inconsistent with their stated goals and as such would appear to be an odd exemplar to encourage women to emulate.

Analyses of course-taking patterns sought to establish the external validity of this pattern and answer our second research question. In Study 2, we examine course-taking using nationally representative longitudinal data, finding that men were more likely than women to retake STEM courses (but not non-STEM courses) after failure. We find no average differences in long-term income for men and found that among men who retook these courses, 45 percent had longer-term incomes that were at least 10 percent lower than their matched counterfactual who did not retake these courses. Overall, male students who retook STEM courses after failure were as likely to earn less than their matched counterfactual as they were to earn more. These findings offer evidence that STEM persistence for men does not necessarily lead to greater long-term success and suggest that overpersistence may be relatively common in male college students' course-taking decisions.

Finally, Study 3 sought to understand the factors driving gender differences in overpersistence. Returning to the experimental paradigm used in Study 1, we again found that men were more likely to choose mathematics questions even when opting for verbal questions would have led to greater success. We found evidence that men's greater confidence in and identification with the domain of mathematics partly explained their greater persistence. The effects of confidence and domain identification also mediated the moderating role of believing that men are superior at math. Thus, men were more likely to overpersist the more they reported believing that men are superior to women in math, and this effect was driven by their greater confidence in and identification with the mathematics domain. This pattern of results is consistent

with past work suggesting that gendered conceptions of mathematics shape not only levels of confidence in different domains but also domain identification, supporting the integrated theoretical model we discuss previously.

Across the three studies, we present evidence for male overpersistence that is diverse with respect to method (using a laboratory experiment, a survey experiment, and transcript data) and population (college students and a diverse convenience sample from the general population). While the diversity of this evidence suggests that the phenomenon we describe is robust, we view our contribution primarily as an existence proof of male overpersistence in STEM fields. We are limited in how much we can say at this stage about the relative frequency of this behavior in different field settings. However, the moderation finding from Study 3 suggests that overpersistence will be more common in cultures and organizational settings with pronounced stereotypes regarding the superior competence of men in STEM fields. Further, the phenomenon is likely less common in settings where STEM activities are substantially easier than non-STEM activities inasmuch as men and women alike should rarely opt out of STEM in such settings, leading to a small or nonexistent persistence gap by gender. Conversely, the phenomenon is likely more common in settings where STEM activities are relatively more difficult than non-STEM, thus encouraging at least some people to opt out, but not impossibly difficult, so that not everyone drops out.

Our findings have important implications for theoretical understandings regarding the gender gap in STEM fields. While gender differences are defined by both women and men's behavior, research on gender differences in STEM persistence typically focuses on how to increase women's persistence so that it mirrors men's. We call attention to the role that men's choices play in creating gender differences in STEM persistence, highlighting that men's STEM persistence levels are driven in part by cultural stereotypes of male superiority in STEM fields. In doing so, we bring confidence and identity—two largely separate lines of inquiry that are often used to understand women's underpersistence in STEM fields—together into an integrated theoretical framework for understanding men's STEM behavior.

Our results also have important implications for policies seeking to create gender parity in STEM fields. Research on gender segregation in the labor force more broadly has not only highlighted the importance of encouraging women to enter male-dominated fields but has argued that realizing gender equality requires also encouraging men to pursue careers in female-dominated fields (England 2010). Presumably because it is driven by concerns around the shortage of qualified workers in the STEM workforce, research on gender differences in STEM fields has focused on encouraging women to enter STEM fields but has not addressed questions around men's STEM persistence. Our results suggest that efforts to increase women's representation in STEM fields by addressing cultural stereotypes are likely to not only increase

women's participation but also affect men's decisions, leading to less overpersistence by men. Thus, from the perspective of creating gender equality in STEM representation, addressing the gendered stereotypes surrounding STEM might effectively work on creating gender balance both by encouraging more women to pursue these fields as well as encouraging men to enter other fields instead of overpersisting in STEM fields. To the degree, however, that STEM policies are seeking to increase the rates of both men and women entering STEM fields, our results suggest that gains in the persistence of women from addressing stereotypes may be partially offset by a loss of men. This is a departure from previous work, which often argues for policies aimed at increasing the representation of women in STEM fields as a way to boost the total number of college graduates with STEM training (see e.g., National Academy of Sciences 2006).

## Conclusion

Central to feminist critiques of the practice of social science is a call for scholars to move beyond approaches in which men are implicitly portrayed as normal, a reference category from which women depart. This traditional approach implies men's behavior is logical and intuitive while treating women's behavior as aberrant and in need of explanation. Consistent with this, the academic literature on STEM persistence has overwhelmingly focused on why so many women exit the STEM pipeline. There is no work of which we are aware that seeks to understand why so many men have remained in the STEM pipeline.

Here we propose that men's choices to persist in the STEM pipeline contribute to gender differences in STEM fields and that their greater persistence manifests even in settings where it leads to less success. Further, we sought to explain why men might overpersist more than women, suggesting that this effect is driven by gendered norms portraying science as a field in which men are superior, boosting men's confidence in and identification with mathematics. In light of these findings, we argue that researchers should conceptualize gender differences in STEM fields as not only the result of female underpersistence but also of male overpersistence. Our findings suggest that if we are to fully understand gender inequality in representation in STEM fields, we must understand the choices made by both women and men in these areas.

## Acknowledgments

Parts of this paper were presented at the annual meeting of the Sociology of Education Association (Pacific Grove, February 2012), the annual meeting of the American Sociological Association (Denver, August 2012), the biennial meeting of the Society for Research in Child Development (Seattle, April 2013), the annual meeting of the American Educational Research Association (San Francisco, April 2013), the fall meeting of the Association for Public Policy Analysis and Management (Miami, November 2015), and in seminars at University of California, Irvine's School of Education (October 2012), University of California, Santa Barbara's

Broom Center for Demography (October 2012), University of California, Riverside's Department of Sociology (October 2012), Princeton University's Office of Population Research (March 2013), New York University's Department of Sociology (November 2013), and the University of Michigan's Institute for Social Research (November 2013). Several meeting and seminar participants made useful comments for which we are grateful. We are also grateful to Shelley Correll for helpful comments and discussions, and to Chrystal Redekopp for research assistance. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Eunice Kennedy Shriver National Institute of Child Health and Human Development, Grant/Award Number: K01HD073319.

## References

- Ackerman, Phillip L., Ruth Kanfer, and Margaret E. Beier. 2013. "Trait Complex, Cognitive Ability, and Domain Knowledge Predictors of Baccalaureate Success, STEM Persistence, and Gender Differences." *Journal of Educational Psychology* 105(3):911–27.
- Adelman, Clifford, Bruce Daniel, and Ilona Berkovits. 2003. *Postsecondary Attainment, Attendance, Curriculum, and Performance: Selected Results from the NELS:88/2000 Postsecondary Education Transcript Study (PETS), 2000*. Washington, DC: National Center for Education Statistics.
- Alper, Joe. 1993. "The Pipeline Is Leaking Women All the Way along." *Science* 260(5106):409–11.
- Andersen, Heine. 2001. "Gender Inequality and Paradigms in the Social Sciences." *Social Science Information* 40:265–89.
- Ayalon, Hanna, and Idit Livneh. 2013. "Educational Standardization and Gender Differences in Mathematics Achievement: A Comparative Study." *Social Science Research* 42:432–45.
- Barres, Ben A. 2006. "Does Gender Matter?" *Nature* 442(7099): 133–36.
- Beauvoir, Simone de. 2010. *The Second Sex* (Trans. Constance Borde and Sheila Malovany-Chevallier). New York: Alfred A. Knopf.
- Beilock, Sian L., Elizabeth A. Gunderson, Gerardo Ramirez, and Susan C. Levine. 2010. "Female Teachers' Math Anxiety Affects Girls' Math Achievement." *Proceedings of the National Academy of Sciences* 107(5):1860–63.
- Blair, Elizabeth E., Rebecca B. Miller, Maria Ong, and Yevgeniya V. Zastavker. 2017. "Undergraduate STEM Instructors' Teacher Identities and Discourses on Student Gender Expression and Equity." *Journal of Engineering Education* 106(1):14–43.
- Blair-Loy, Mary, and Erin A. Cech. 2017. "Demands and Devotion: Cultural Meanings of Work and Overload among Women Researchers and Professionals in Science and Technology Industries." *Sociological Forum* 32(1):5–27.
- Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, yet High-quality, Data?" *Perspectives on Psychological Science* 6(1):3–5.
- Cech, Erin A. 2013. "The Self-expressive Edge of Occupational Sex Segregation." *American Journal of Sociology* 119(3):747–89.

- Charles, Maria, and Karen Bradley. 2009. "Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries." *American Journal of Sociology* 114:924–76.
- Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-assessments." *American Journal of Sociology* 106:1691–730.
- Correll, Shelley J. 2004. "Constraints into Preferences: Gender, Status, and Emerging Career Aspirations." *American Sociological Review* 69:93–113.
- Duckworth, Angela L., Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. 2007. "Grit: Perseverance and Passion for Long-term Goals." *Journal of Personality and Social Psychology* 92(6):1087–101.
- Eccles, Jacquelynne S. 1987. "Gender Roles and Women's Achievement-related Decisions." *Psychology of Women Quarterly* 11(2):135–72.
- England, Paula. 2010. "The Gender Revolution: Uneven and Stalled." *Gender & Society* 24(2):149–66.
- Feinberg, Matthew, Robb Willer, and Dacher Keltner. 2012. "Flustered and Faithful: Embarrassment as a Signal of Prosocial Behavior." *Journal of Personality and Social Psychology* 102:81–97.
- Fennema, E., and Julia A. Sherman. 1976. "Fennema-Sherman Mathematics Attitude Scales: Instruments Designed to Measure Attitudes toward the Learning of Mathematics by Females and Males." *Catalog of Selected Documents in Psychology* 6:31 (Ms. No. 1225).
- Fennema, Elizabeth, and Julia Sherman. 1977. "Sex-related Differences in Mathematics Achievement, Spatial Visualization and Affective Factors." *American Educational Research Journal* 14:51–71.
- Frank, Kenneth A., Chandra Muller, Kathryn S. Schiller, Catherine Riegle-Crumb, Anna Strassmann Mueller, Robert Crosnoe, and Jennifer Pearson. "The Social Dynamics of Mathematics Course-taking in High School." *American Journal of Sociology* 113:1645–96.
- Griffith, Amanda L. 2010. "Persistence of Women and Minorities in STEM Field Majors: Is It the School That Matters?" *Economics of Education Review* 29:911–22.
- Heckhausen, Jutta. 1997. "Developmental Regulation across Adulthood: Primary and Secondary Control of Age-related Challenges." *Developmental Psychology* 33(1):176–87.
- Legewie, Joscha, and Thomas A. DiPrete. 2014. "The High School Environment and the Gender Gap in Science and Engineering." *Sociology of Education* 87: 259–80.
- Li, Danielle. 2012. "Essays on the Organization of Science and Education." PhD Thesis, MIT.
- Luhtanen, Riia, and Jennifer Crocker. 1992. "A Collective Self-esteem Scale: Self-evaluation of One's Social Identity." *Personality and Social Psychology Bulletin* 18(3):302–18.
- Ma, Yingyi. 2011. "Gender Differences in the Paths Leading to a STEM Baccalaureate." *Social Science Quarterly* 92(5):1169–90.
- National Academy of Sciences. 2006. *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering*. Washington, DC: National Academy Press.
- Oh, Seong Soo, and Gregory B. Lewis. 2011. "Stemming Inequality? Employment and Pay of Female and Minority Scientists and Engineers." *The Social Science Journal* 48(2):397–403.
- Patenaude, Allan L., and Deborah L. Laufersweiller-Dwyer. 2002. *Arkansas Comprehensive Substance Abuse Treatment Program: Process Evaluation of the Modified Therapeutic Community*. Report prepared for the U.S. Department of Justice, Washington, DC.
- Preacher, Kristopher J., Derek D. Rucker, and Andrew F. Hayes. 2007. "Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions." *Multivariate Behavioral Research* 42(1):185–227.
- Price, Joshua. 2010. "The Effect of Instructor Race and Gender on Student Persistence in STEM Fields." *Economics of Education Review* 29(6):901–10.
- Riegle-Crumb, C. 2005. "The Cross-national Context of the Gender Gap in Math and Science." Pp. 227–43 in *The Social Organization of Schools*, edited by L. Hedges and B. Schneider. New York: Russell Sage Press.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55.
- Sanabria, Tanya, and Andrew M. Penner. 2017. "Weeded out: Gendered Responses to Failing Calculus." *Social Sciences* 6(47):1–14.
- Seron, Carroll, Susan S. Sibley, Erin Cech, and Brian Rubineau. 2016. "Persistence is Cultural: Professional Socialization and the Reproduction of Sex Segregation." *Work and Occupations* 43(2):178–214.
- Snyder, Thomas D., and Sally A. Dillow. 2012. *Digest of Education Statistics 2011*. Washington, DC: National Center for Education Statistics.
- Stage, Frances K., and Sue A. Maple. 1996. "Incompatible Goals: Narratives of Graduate Women in the Mathematics Pipeline." *American Educational Research Journal* 33:23–51.
- Wang, Ming-Te, Jacquelynne S. Eccles, and Sarah Kenny. 2013. "Not Lack of Ability but More Choice Individual and Gender Differences in Choice of Careers in Science, Technology, Engineering, and Mathematics." *Psychological Science* 24:770–75.
- Weinberg, Jill D, Jeremy Freese, and David McElhattan. 2014. "Comparing Data Characteristics and Results of an Online Factorial Survey between a Population-based and a Crowdsourced Sample." *Sociological Science* 1:292–310.
- Willer, Robb, Christabel Rogalin, Bridget Conlon, and Michael T. Wojnowicz. 2013. "Overdoing Gender: A Test of the Masculine Overcompensation Thesis." *American Journal of Sociology* 118:980–1022.
- Williams, Wendy M., and Stephen J. Ceci. 2015. "National Hiring Experiments Reveal 2: 1 Faculty Preference for Women on STEM Tenure Track." *Proceedings of the National Academy of Sciences* 112(17):5360–65.
- Xie, Yu, and Kimberlee A. Shauman. 2003. *Women in Science*. Cambridge, MA: Harvard University Press.

### Author Biographies

**Andrew M. Penner** is professor of sociology at the University of California, Irvine. His research focuses on inequality, social categorization, and educational policy.

**Robb Willer** is a professor of sociology and the director of the Polarization and Social Change Laboratory at Stanford University. He studies a variety of topics, including politics, morality, status hierarchies, morality, and cooperation. His research has appeared in the *American Sociological Review*, *American Journal of Sociology*, *Journal of Personality and Social Psychology*, *Proceedings of the National Academy of Sciences*, and elsewhere.