Science faculty's subtle gender biases favor male students

Corinne A. Moss-Racusin[a,b], John F. Dovidio[b], Victoria L. Brescoll[c], Mark J. Graham[a,d], and Jo Handelsman[a,1]

[a]Department of Molecular, Cellular and Developmental Biology, [b]Department of Psychology, [c]School of Management, and [d]Department of Psychiatry, Yale University, New Haven, CT 06520

Edited* by Shirley Tilghman, Princeton University, Princeton, NJ, and approved August 21, 2012 (received for review July 2, 2012)

Despite efforts to recruit and retain more women, a stark gender disparity persists within academic science. Abundant research has demonstrated gender bias in many demographic groups, but has yet to experimentally investigate whether science faculty exhibit a bias against female students that could contribute to the gender disparity in academic science. In a randomized double-blind study (*n* = 127), science faculty from research-intensive universities rated the application materials of a student—who was randomly assigned either a male or female name—for a laboratory manager position. Faculty participants rated the male applicant as significantly more competent and hireable than the (identical) female applicant. These participants also selected a higher starting salary and offered more career mentoring to the male applicant. The gender of the faculty participants did not affect responses, such that female and male faculty were equally likely to exhibit bias against the female student. Mediation analyses indicated that the female student was less likely to be hired because she was viewed as less competent. We also assessed faculty participants' preexisting subtle bias against women using a standard instrument and found that preexisting subtle bias against women played a moderating role, such that subtle bias against women was associated with less support for the female student, but was unrelated to reactions to the male student. These results suggest that interventions addressing faculty gender bias might advance the goal of increasing the participation of women in science.

diversity | lifestyle choices | science education | science workforce

A 2012 report from the President's Council of Advisors on Science and Technology indicates that training scientists and engineers at current rates will result in a deficit of 1,000,000 workers to meet United States workforce demands over the next decade (1). To help close this formidable gap, the report calls for the increased training and retention of women, who are starkly underrepresented within many fields of science, especially among the professoriate (2–4). Although the proportion of science degrees granted to women has increased (5), there is a persistent disparity between the number of women receiving PhDs and those hired as junior faculty (1–4). This gap suggests that the problem will not resolve itself solely by more generations of women moving through the academic pipeline but that instead, women's advancement within academic science may be actively impeded.

With evidence suggesting that biological sex differences in inherent aptitude for math and science are small or nonexistent (6–8), the efforts of many researchers and academic leaders to identify causes of the science gender disparity have focused instead on the life choices that may compete with women's pursuit of the most demanding positions. Some research suggests that these lifestyle choices (whether free or constrained) likely contribute to the gender imbalance (9–11), but because the majority of these studies are correlational, whether lifestyle factors are solely or primarily responsible remains unclear. Still, some researchers have argued that women's preference for nonscience disciplines and their tendency to take on a disproportionate amount of child- and family-care are the primary causes of the gender disparity in science (9–11), and that it "is not caused by discrimination in these domains" (10). This assertion has received substantial attention and generated significant debate among the scientific community, leading some to conclude that gender discrimination indeed does not exist nor contribute to the gender disparity within academic science (e.g., refs. 12 and 13).

Despite this controversy, experimental research testing for the presence and magnitude of gender discrimination in the biological and physical sciences has yet to be conducted. Although acknowledging that various lifestyle choices likely contribute to the gender imbalance in science (9–11), the present research is unique in investigating whether faculty gender bias exists within academic biological and physical sciences, and whether it might exert an independent effect on the gender disparity as students progress through the pipeline to careers in science. Specifically, the present experiment examined whether, given an equally qualified male and female student, science faculty members would show preferential evaluation and treatment of the male student to work in their laboratory. Although the correlational and related laboratory studies discussed below suggest that such bias is likely (contrary to previous arguments) (9–11), we know of no previous experiments that have tested for faculty bias against female students within academic science.

If faculty express gender biases, we are not suggesting that these biases are intentional or stem from a conscious desire to impede the progress of women in science. Past studies indicate that people's behavior is shaped by implicit or unintended biases, stemming from repeated exposure to pervasive cultural stereotypes (14) that portray women as less competent but simultaneously emphasize their warmth and likeability compared with men (15). Despite significant decreases in overt sexism over the last few decades (particularly among highly educated people) (16), these subtle gender biases are often still held by even the most egalitarian individuals (17), and are exhibited by both men and women (18). Given this body of work, we expected that female faculty would be just as likely as male faculty to express an unintended bias against female undergraduate science students. The fact that these prevalent biases often remain undetected highlights the need for an experimental investigation to determine whether they may be present within academic science and, if so, raise awareness of their potential impact.

Whether these gender biases operate in academic sciences remains an open question. On the one hand, although considerable research demonstrates gender bias in a variety of other domains (19–23), science faculty members may not exhibit this

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Author contributions: C.A.M.-R., J.F.D., V.L.B., M.J.G., and J.H. designed research; C.A.M.-R. performed research; C.A.M.-R. analyzed data; and C.A.M.-R., J.F.D., V.L.B., M.J.G., and J.H. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: jo.handelsman@yale.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1211286109/-/DCSupplemental.

bias because they have been rigorously trained to be objective. On the other hand, research demonstrates that people who value their objectivity and fairness are paradoxically particularly likely to fall prey to biases, in part because they are not on guard against subtle bias (24, 25). Thus, by investigating whether science faculty exhibit a bias that could contribute to the gender disparity within the fields of science, technology, engineering, and mathematics (in which objectivity is emphasized), the current study addressed critical theoretical and practical gaps in that it provided an experimental test of faculty discrimination against female students within academic science.

A number of lines of research suggest that such discrimination is likely. Science is robustly male gender-typed (26, 27), resources are inequitably distributed among men and women in many academic science settings (28), some undergraduate women perceive unequal treatment of the genders within science fields (29), and nonexperimental evidence suggests that gender bias is present in other fields (19). Some experimental evidence suggests that even though evaluators report liking women more than men (15), they judge women as less competent than men even when they have identical backgrounds (20). However, these studies used undergraduate students as participants (rather than experienced faculty members), and focused on performance domains outside of academic science, such as completing perceptual tasks (21), writing nonscience articles (22), and being evaluated for a corporate managerial position (23).

Thus, whether aspiring women scientists encounter discrimination from faculty members remains unknown. The formative predoctoral years are a critical window, because students' experiences at this juncture shape both their beliefs about their own abilities and subsequent persistence in science (30, 31). Therefore, we selected this career stage as the focus of the present study because it represents an opportunity to address issues that manifest immediately and also resurface much later, potentially contributing to the persistent faculty gender disparity (32, 33).

Current Study

In addition to determining whether faculty expressed a bias against female students, we also sought to identify the processes contributing to this bias. To do so, we investigated whether faculty members' perceptions of student competence would help to explain why they would be less likely to hire a female (relative to an identical male) student for a laboratory manager position. Additionally, we examined the role of faculty members' preexisting subtle bias against women. We reasoned that pervasive cultural messages regarding women's lack of competence in science could lead faculty members to hold gender-biased attitudes that might subtly affect their support for female (but not male) science students. These generalized, subtly biased attitudes toward women could impel faculty to judge equivalent students differently as a function of their gender.

The present study sought to test for differences in faculty perceptions and treatment of equally qualified men and women pursuing careers in science and, if such a bias were discovered, reveal its mechanisms and consequences within academic science. We focused on hiring for a laboratory manager position as the primary dependent variable of interest because it functions as a professional launching pad for subsequent opportunities. As secondary measures, which are related to hiring, we assessed: (*i*) perceived student competence; (*ii*) salary offers, which reflect the extent to which a student is valued for these competitive positions; and (*iii*) the extent to which the student was viewed as deserving of faculty mentoring.

Our hypotheses were that: Science faculty's perceptions and treatment of students would reveal a gender bias favoring male students in perceptions of competence and hireability, salary conferral, and willingness to mentor (hypothesis A); Faculty gender would not influence this gender bias (hypothesis B); Hiring discrimination against the female student would be mediated (i.e., explained) by faculty perceptions that a female student is less competent than an identical male student (hypothesis C); and Participants' preexisting subtle bias against women would moderate (i.e., impact) results, such that subtle bias against women would be negatively related to evaluations of the female student, but unrelated to evaluations of the male student (hypothesis D).

Results

A broad, nationwide sample of biology, chemistry, and physics professors ($n = 127$) evaluated the application materials of an undergraduate science student who had ostensibly applied for a science laboratory manager position. All participants received the same materials, which were randomly assigned either the name of a male ($n = 63$) or a female ($n = 64$) student; student gender was thus the only variable that differed between conditions. Using previously validated scales, participants rated the student's competence and hireability, as well as the amount of salary and amount of mentoring they would offer the student. Faculty participants believed that their feedback would be shared with the student they had rated (see *Materials and Methods* for details).

Student Gender Differences. The competence, hireability, salary conferral, and mentoring scales were each submitted to a two (student gender; male, female) × two (faculty gender; male, female) between-subjects ANOVA. In each case, the effect of student gender was significant (all $P < 0.01$), whereas the effect of faculty participant gender and their interaction was not (all $P > 0.19$). Tests of simple effects (all $d > 0.60$) indicated that faculty participants viewed the female student as less competent [$t(125) = 3.89$, $P < 0.001$] and less hireable [$t(125) = 4.22$, $P < 0.001$] than the identical male student (Fig. 1 and Table 1). Faculty participants also offered less career mentoring to the female student than to the male student [$t(125) = 3.77$, $P < 0.001$]. The mean starting salary offered the female student, $26,507.94, was significantly lower than that of $30,238.10 to the male student [$t(124) = 3.42$, $P < 0.01$] (Fig. 2). These results support hypothesis A.

In support of hypothesis B, faculty gender did not affect bias (Table 1). Tests of simple effects (all $d < 0.33$) indicated that female faculty participants did not rate the female student as more competent [$t(62) = 0.06$, $P = 0.95$] or hireable [$t(62) = 0.41$, $P = 0.69$] than did male faculty. Female faculty also did not offer more mentoring [$t(62) = 0.29$, $P = 0.77$] or a higher salary [$t(61) = 1.14$, $P = 0.26$] to the female student than did their male

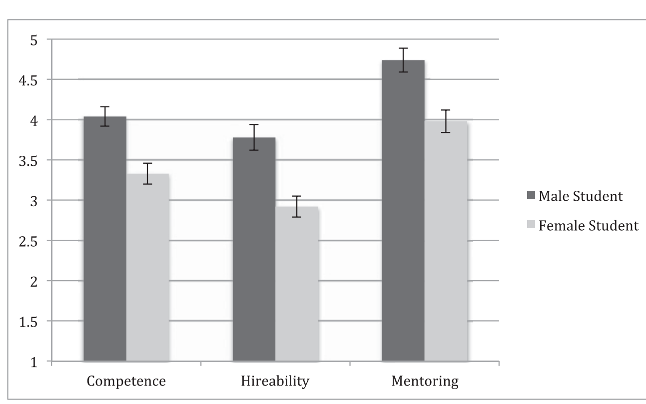

Fig. 1. Competence, hireability, and mentoring by student gender condition (collapsed across faculty gender). All student gender differences are significant ($P < 0.001$). Scales range from 1 to 7, with higher numbers reflecting a greater extent of each variable. Error bars represent SEs. $n_{\text{male student condition}} = 63$, $n_{\text{female student condition}} = 64$.

Table 1. Means for student competence, hireability, mentoring and salary conferral by student gender condition and faculty gender

	Male target student				Female target student				
	Male faculty		Female faculty		Male faculty		Female faculty		
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	*d*
Competence	4.01_a	(0.92)	4.1_a	(1.19)	3.33_b	(1.07)	3.32_b	(1.10)	0.71
Hireability	3.74_a	(1.24)	3.92_a	(1.27)	2.96_b	(1.13)	2.84_b	(0.84)	0.75
Mentoring	4.74_a	(1.11)	4.73_a	(1.31)	4.00_b	(1.21)	3.91_b	(0.91)	0.67
Salary	$30{,}520.83_a$	(5,764.86)	$29{,}333.33_a$	(4,952.15)	$27{,}111{,}11_b$	(6,948.58)	$25{,}000.00_b$	(7,965.56)	0.60

Scales for competence, hireability, and mentoring range from 1 to 7, with higher numbers reflecting a greater extent of each variable. The scale for salary conferral ranges from \$15,000 to \$50,000. Means with different subscripts within each row differ significantly ($P < 0.05$). Effect sizes (Cohen's *d*) represent target student gender differences (no faculty gender differences were significant, all $P > 0.14$). Positive effect sizes favor male students. Conventional small, medium, and large effect sizes for *d* are 0.20, 0.50, and 0.80, respectively (51). $n_{\text{male student condition}} = 63$, $n_{\text{female student condition}} = 64$. ***$P < 0.001$.

colleagues. In addition, faculty participants' scientific field, age, and tenure status had no effect (all $P > 0.53$). Thus, the bias appears pervasive among faculty and is not limited to a certain demographic subgroup.

Mediation and Moderation Analyses. Thus far, we have considered the results for competence, hireability, salary conferral, and mentoring separately to demonstrate the converging results across these individual measures. However, composite indices of measures that converge on an underlying construct are more statistically reliable, stable, and resistant to error than are each of the individual items (e.g., refs. 34 and 35). Consistent with this logic, the established approach to measuring the broad concept of target competence typically used in this type of gender bias research is to standardize and average the competence scale items and the salary conferral variable to create one composite competence index, and to use this stable convergent measure for all analyses (e.g., refs. 36 and 37). Because this approach obscures mean salary differences between targets, we chose to present salary as a distinct dependent variable up to this point, to enable a direct test of the potential discrepancy in salary offered to the male and female student targets. However, to rigorously examine the processes underscoring faculty gender bias, we reverted to standard practices at this point by averaging the standardized salary variable with the competence scale items to create a robust composite competence variable ($\alpha = 0.86$). This composite competence variable was used in all subsequent mediation and moderation analyses.

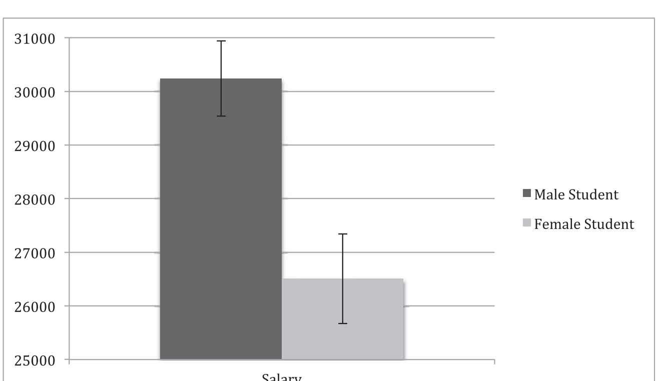

Fig. 2. Salary conferral by student gender condition (collapsed across faculty gender). The student gender difference is significant ($P < 0.01$). The scale ranges from \$15,000 to \$50,000. Error bars represent SEs. $n_{\text{male student condition}} = 63$, $n_{\text{female student condition}} = 64$.

Evidence emerged for hypothesis C, the predicted mediation (i.e., causal path; see *SI Materials and Methods: Additional Analyses* for more information on mediation and the results of additional mediation analyses). The initially significant impact of student gender on hireability ($\beta = -0.35$, $P < 0.001$) was reduced in magnitude and dropped to nonsignificance ($\beta = -0.10$, $P = 0.13$) after accounting for the impact of student composite competence (which was a strong predictor, $\beta = 0.69$, $P < 0.001$), Sobel's $Z = 3.94$, $P < 0.001$ (Fig. 3). This pattern of results provides evidence for full mediation, indicating that the female student was less likely to be hired than the identical male because she was viewed as less competent overall.

We also conducted moderation analysis (i.e., testing for factors that could amplify or attenuate the demonstrated effect) to determine the impact of faculty participants' preexisting subtle bias against women on faculty participants' perceptions and treatment of male and female science students (see *SI Materials and Methods: Additional Analyses* for more information on and the results of additional moderation analyses). For this purpose, we administered the Modern Sexism Scale (38), a well-validated instrument frequently used for this purpose (*SI Materials and Methods*). Consistent with our intentions, this scale measures unintentional negativity toward women, as contrasted with a more blatant form of conscious hostility toward women.

Results of multiple regression analyses indicated that participants' preexisting subtle bias against women significantly interacted with student gender to predict perceptions of student composite competence ($\beta = -0.39$, $P < 0.01$), hireability ($\beta = -0.31$, $P < 0.05$), and mentoring ($\beta = -0.55$, $P < 0.001$). To interpret these significant interactions, we examined the simple effects separately by student gender. Results revealed that the more preexisting subtle bias participants exhibited against women, the less composite competence ($\beta = -0.36$, $P < 0.01$) and hireability ($\beta = -0.39$, $P < 0.01$) they perceived in the female student, and the less mentoring ($\beta = -0.53$, $P < 0.001$) they were willing to offer her. In contrast, faculty participants' levels of preexisting subtle bias against women were unrelated to the perceptions of the male student's composite competence ($\beta = 0.16$, $P = 0.22$) and hireability ($\beta = 0.07$, $P = 0.59$), and the amount of mentoring ($\beta = 0.22$, $P = 0.09$) they were willing to offer him. [Although this effect is marginally significant, its direction suggests that faculty participants' preexisting subtle bias against women may actually have made them more inclined to mentor the male student relative to the female student (although this effect should be interpreted with caution because of its marginal significance).] Thus, it appears that faculty participants' preexisting subtle gender bias undermined support for the female student but was unrelated to perceptions and treatment of the male student. These findings support hypothesis D.

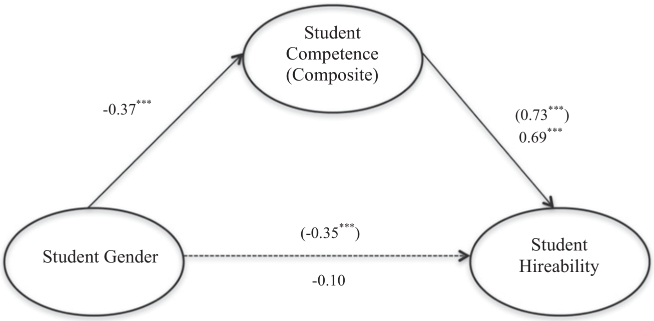

Fig. 3. Student gender difference hiring mediation. Values are standardized regression coefficients. The value in parentheses reflects a bivariate analysis. The dashed line represents the mediated path. The composite student competence variable consists of the averaged standardized salary variable and the competence scale items. Student gender is coded such that male = 0, female = 1. $n_{\text{male student condition}} = 63$, $n_{\text{female student condition}} = 64$. ***$P < 0.001$.

Finally, using a previously validated scale, we also measured how much faculty participants liked the student (see *SI Materials and Methods*). In keeping with a large body of literature (15), faculty participants reported liking the female (mean = 4.35, SD = 0.93) more than the male student [(mean = 3.91, SD = 0.1.08), $t(125) = -2.44$, $P < 0.05$]. However, consistent with this previous literature, liking the female student more than the male student did not translate into positive perceptions of her composite competence or material outcomes in the form of a job offer, an equitable salary, or valuable career mentoring. Moreover, only composite competence (and not likeability) helped to explain why the female student was less likely to be hired; in mediation analyses, student gender condition ($\beta = -0.48$, $P < 0.001$) remained a strong predictor of hireability along with likeability ($\beta = 0.60$, $P < 0.001$). These findings underscore the point that faculty participants did not exhibit outright hostility or dislike toward female students, but were instead affected by pervasive gender stereotypes, unintentionally downgrading the competence, hireability, salary, and mentoring of a female student compared with an identical male.

Discussion

The present study is unique in investigating subtle gender bias on the part of faculty in the biological and physical sciences. It therefore informs the debate on possible causes of the gender disparity in academic science by providing unique experimental evidence that science faculty of both genders exhibit bias against female undergraduates. As a controlled experiment, it fills a critical gap in the existing literature, which consisted only of experiments in other domains (with undergraduate students as participants) and correlational data that could not conclusively rule out the influence of other variables.

Our results revealed that both male and female faculty judged a female student to be less competent and less worthy of being hired than an identical male student, and also offered her a smaller starting salary and less career mentoring. Although the differences in ratings may be perceived as modest, the effect sizes were all moderate to large ($d = 0.60$–0.75). Thus, the current results suggest that subtle gender bias is important to address because it could translate into large real-world disadvantages in the judgment and treatment of female science students (39). Moreover, our mediation findings shed light on the processes responsible for this bias, suggesting that the female student was less likely to be hired than the male student because she was perceived as less competent. Additionally, moderation results indicated that faculty participants' preexisting subtle bias against women undermined their perceptions and treatment of the female (but not the male) student, further suggesting that chronic subtle biases may harm women within academic science. Use of a randomized controlled design and established practices from audit study methodology support the ecological validity and educational implications of our findings (*SI Materials and Methods*).

It is noteworthy that female faculty members were just as likely as their male colleagues to favor the male student. The fact that faculty members' bias was independent of their gender, scientific discipline, age, and tenure status suggests that it is likely unintentional, generated from widespread cultural stereotypes rather than a conscious intention to harm women (17). Additionally, the fact that faculty participants reported liking the female more than the male student further underscores the point that our results likely do not reflect faculty members' overt hostility toward women. Instead, despite expressing warmth toward emerging female scientists, faculty members of both genders appear to be affected by enduring cultural stereotypes about women's lack of science competence that translate into biases in student evaluation and mentoring.

Our careful selection of expert participants revealed gender discrimination among existing science faculty members who interact with students on a regular basis (*SI Materials and Methods: Subjects and Recruitment Strategy*). This method allowed for a high degree of ecological validity and generalizability relative to an approach using nonexpert participants, such as other undergraduates or lay people unfamiliar with laboratory manager job requirements and academic science mentoring (i.e., the participants in much psychological research on gender discrimination). The results presented here reinforce those of Stenpries, Anders, and Ritzke (40), the only other experiment we know of that recruited faculty participants. Because this previous experiment also indicated bias within academic science, its results raised serious concerns about the potential for faculty bias within the biological and physical sciences, casting further doubt on assertions (based on correlational data) that such biases do not exist (9–11). In the Steinpreis et al. experiment, psychologists were more likely to hire a psychology faculty job applicant when the applicant's curriculum vitae was assigned a male (rather than female) name (40). This previous work invited a study that would extend the finding to faculty in the biological and physical sciences and to reactions to undergraduates, whose competence was not already fairly established by accomplishments associated with the advanced career status of the faculty target group of the previous study. By providing this unique investigation of faculty bias against female students in biological and physical sciences, the present study extends past work to a critical early career stage, and to fields where women's underrepresentation remains stark (2–4).

Indeed, our findings raise concerns about the extent to which negative predoctoral experiences may shape women's subsequent decisions about persistence and career specialization. Following conventions established in classic experimental studies to create enough ambiguity to leave room for potentially biased responses (20, 23), the student applicants in the present research were described as qualified to succeed in academic science (i.e., having coauthored a publication after obtaining 2 y of research experience), but not irrefutably excellent. As such, they represented a majority of aspiring scientists, and were precisely the type of students most affected by faculty judgments and mentoring (see *SI Materials and Methods* for more discussion). Our results raise the possibility that not only do such women encounter biased judgments of their competence and hireability, but also receive less faculty encouragement and financial rewards than identical male counterparts. Because most students depend on feedback from their environments to calibrate their own worth (41), faculty's assessments of students' competence likely contribute to students' self-efficacy and goal setting as scientists, which may influence decisions much later in their careers. Likewise, inasmuch as the advice and mentoring that students receive affect their ambitions and choices, it is significant that the faculty in this study were less inclined to mentor women than men. This finding raises the possibility that women may opt out of academic science careers in part because of diminished competence judgments, rewards, and mentoring received in the early years of the careers. In sum, the predoctoral years represent a window during which students' experiences of faculty bias or encouragement are particularly likely to shape their persistence in academic science (30–33). Thus, the present study not only fills an important gap in the research literature, but also has critical implications for pressing social and educational issues associated with the gender disparity in science.

If women's decisions to leave science fields when or before they reach the faculty level are influenced by unequal treatment by undergraduate advisors, then existing efforts to create more flexible work settings (42) or increase women's identification with science (27) may not fully alleviate a critical underlying problem. Our results suggest that academic policies and mentoring interventions targeting undergraduate advisors could contribute to reducing the gender disparity. Future research should evaluate the efficacy of educating faculty and students about the existence and impact of bias within academia, an approach that has reduced racial bias among students (43). Educational efforts might address research on factors that attenuate gender bias in real-world settings, such as increasing women's self-monitoring (44). Our results also point to the importance of establishing objective, transparent student evaluation and admissions criteria to guard against observers' tendency to unintentionally use different standards when assessing women relative to men (45, 46). Without such actions, faculty bias against female undergraduates may continue to undermine meritocratic advancement, to the detriment of research and education.

Conclusions

The dearth of women within academic science reflects a significant wasted opportunity to benefit from the capabilities of our best potential scientists, whether male or female. Although women have begun to enter some science fields in greater numbers (5), their mere increased presence is not evidence of the absence of bias. Rather, some women may persist in academic science despite the damaging effects of unintended gender bias on the part of faculty. Similarly, it is not yet possible to conclude that the preferences for other fields and lifestyle choices (9–11) that lead many women to leave academic science (even after obtaining advanced degrees) are not themselves influenced by experiences of bias, at least to some degree. To the extent that faculty gender bias impedes women's full participation in science, it may undercut not only academic meritocracy, but also the expansion of the scientific workforce needed for the next decade's advancement of national competitiveness (1).

Materials and Methods

Participants. We recruited faculty participants from Biology, Chemistry, and Physics departments at three public and three private large, geographically diverse research-intensive universities in the United States, strategically selected for their representative characteristics (see *SI Materials and Methods* for more information on department selection). The demographics of the 127 respondents corresponded to both the averages for the selected departments and faculty at all United States research-intensive institutions, meeting the criteria for generalizability even from nonrandom samples (see *SI Materials and Methods* for more information on recruitment strategy and participant characteristics). Indeed, we were particularly careful to obtain a sample representative of the underlying population, because many past studies have demonstrated that when this is the case, respondents and nonrespondents typically do not differ on demographic characteristics and responses to focal variables (47).

Additionally, in keeping with recommended practices, we conducted an a priori power analysis before beginning data collection to determine the optimal sample size needed to detect effects without biasing results toward obtaining significance (*SI Materials and Methods: Subjects and Recruitment Strategy*) (48). Thus, although our sample size may appear small to some readers, it is important to note that we obtained the necessary power and representativeness to generalize from our results while purposefully avoiding an unnecessarily large sample that could have biased our results toward a false-positive type I error (48).

Procedure. Participants were asked to provide feedback on the materials of an undergraduate science student who stated their intention to go on to graduate school, and who had recently applied for a science laboratory manager position. Of importance, participants believed they were evaluating a real student who would subsequently receive the faculty participants' ratings as feedback to help their career development (see *SI Materials and Methods* for more information, and Fig. S1 for the full text of the cover story). Thus, the faculty participants' ratings were associated with definite consequences.

Following established practices, the laboratory manager application was designed to reflect high but slightly ambiguous competence, allowing for variability in participant responses (20, 23). In addition, a promising but still-nascent applicant is precisely the type of student whose persistence in academic science is most likely to be affected by faculty support or discouragement (30–33), rendering faculty reactions to such a student of particular interest for the present purposes. The materials were developed in consultation with a panel of academic science researchers (who had extensive experience hiring and supervising student research assistants) to ensure that they would be perceived as realistic (*SI Materials and Methods*). Results of a funneled debriefing (49) indicated that this was successful; no participant reported suspicions that the target was not an actual student who would receive their evaluation.

Participants were randomly assigned to one of two student gender conditions: application materials were attributed to either a male student (John, $n = 63$), or a female student (Jennifer, $n = 64$), two names that have been pretested as equivalent in likability and recognizeability (50). Thus, each participant saw only one set of materials, from either the male or female applicant (see Fig. S2 for the full text of the laboratory manager application and *SI Method and Materials* for more information on all materials). Because all other information was held constant between conditions, any differences in participants' responses are attributable to the gender of the student. Using validated scales, participants rated student competence, their own likelihood of hiring the student, selected an annual starting salary for the student, indicated how much career mentoring they would provide to such a student, and completed the Modern Sexism Scale.

ACKNOWLEDGMENTS. We thank faculty members from six anonymous universities for their involvement as participants; and Jessamina Blum, John Crosnick, Jennifer Frederick, Jaime Napier, Jojanneke van der Toorn, Tiffany Tsang, Tessa West, James Young, and two anonymous reviewers for valuable input. This research was supported by a grant from the Howard Hughes Medical Institute Professors Program (to J.H.).

PSYCHOLOGICAL AND COGNITIVE SCIENCES

1. President's Council of Advisors on Science and Technology (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics.* Available at http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_feb.pdf. (Accessed February 13, 2012).
2. Handelsman J, et al. (2005) Careers in science. More women in science. *Science* 309:1190–1191.
3. United States National Academy of Sciences (2007) *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering* (National Academies, Washington, DC).
4. National Science Foundation (2009) *Women, Minorities, and Persons with Disabilities in Science and Engineering* (National Science Foundation, Arlington).
5. Bell N (2010) *Graduate Enrollment and Degrees: 1999 to 2009* (Council of Graduate Schools, Washington, DC).
6. Halpern DF, et al. (2007) The science of gender differences in science and mathematics. *Psychol Sci* 8(1):1–51.
7. Hyde JS, Linn MC (2006) Diversity. Gender similarities in mathematics and science. *Science* 314:599–600.
8. Spelke ES (2005) Sex differences in intrinsic aptitude for mathematics and science?: A critical review. *Am Psychol* 60:950–958.
9. Ceci SJ, Williams WM (2010) Gender differences in math-intensive fields. *Curr Dir Psychol Sci* 19:275–279.
10. Ceci SJ, Williams WM (2011) Understanding current causes of women's underrepresentation in science. *Proc Natl Acad Sci USA* 108:3157–3162.

11. Ceci SJ, et al. (2011) Do subtle cues about belongingness constrain women's career choices? *Psychol Inq* 22:255–258.
12. Berezow AB (2011) *Gender discrimination in science is a myth.* Available at http://www.nationalreview.com/articles/256816/gender-discrimination-science-myth-alex-b-berezow?pg=2. (Accessed June 6, 2012).
13. Dickey Zakaib G (2011) Science gender gap probed. *Nature* 470:153.
14. Devine PG (1989) Stereotypes and prejudice: Their automatic and controlled components. *J Pers Soc Psychol* 56(1):5–18.
15. Eagly AH, Mladinic A (1994) Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *Eur Rev Soc Psychol* 5(1):1–35.
16. Spence JT, Hahn ED (1997) The attitudes toward women scale and attitude change in college students. *Psychol Women Q* 21(1):17–34.
17. Dovidio JF, Gaertner SL (2004) *Advances in Experimental Social Psychology*, ed Zanna MP (Elsevier, New York), pp 1–51.
18. Nosek BA, Banaji M, Greenwald AG (2002) Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dyn* 6(1):101–115.
19. Goldin C, Rouse C (2000) Orchestrating impartiality: The impact of "blind" auditions on female musicians. *Am Econ Rev* 90:715–741.
20. Foschi M (2000) Double standards for competence: Theory and research. *Annu Rev Sociol* 26(1):21–42.
21. Foschi M (1996) Double standards in the evaluation of men and women. *Soc Psychol Q* 59:237–254.
22. Goldberg P (1968) Are women prejudiced against women? *Transaction* 5(5):28–30.
23. Heilman ME, Wallen AS, Fuchs D, Tamkins MM (2004) Penalties for success: Reactions to women who succeed at male gender-typed tasks. *J Appl Psychol* 89:416–427.
24. Monin B, Miller DT (2001) Moral credentials and the expression of prejudice. *J Pers Soc Psychol* 81:33–43.
25. Uhlmann EL, Cohen GL (2007) "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organ Behav Hum Dec* 104:207–223.
26. Nosek BA, Banaji MR, Greenwald AG (2002) Math = male, me = female, therefore math not = me. *J Pers Soc Psychol* 83:44–59.
27. Dasgupta N (2011) Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychol Inq* 22:231–246.
28. The Massachusetts Institute of Technology (1999) *A study on the status of women faculty in science at MIT.* Available at http://web.mit.edu/fnl/women/women.html#The%20Study (Accessed February 13, 2012).
29. Steele J, James JB, Barnett RC (2002) Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. *Psychol Women Q* 26(1):46–50.
30. Lent RW, et al. (2001) The role of contextual supports and barriers in the choice of math/science educational options: A test of social cognitive hypotheses. *J Couns Psychol* 48:474–483.
31. Seymour E, Hewitt NM (1996) *Talking about Leaving: Why Undergraduates Leave the Sciences* (Westview, Boulder).
32. Gasiewski JA, et al. (2012) From gatekeeping to engagement: A multicontextual, mixed method study of student academic engagement in introductory STEM courses. *Res Higher Educ* 53:229–261.
33. Byars-Winston A, Gutierrez B, Topp S, Carnes M (2011) Integrating theory and practice to increase scientific workforce diversity: A framework for career development in graduate research training. *CBE Life Sci Educ* 10:357–367.
34. Campbell DT, Fiske DW (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 56:81–105.
35. Robins RW, Hendin HM, Trzesniewski KH (2001) Measuring global self esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Pers Soc Psychol Bull* 27(2):151–161.
36. Moss-Racusin CA, Rudman LA (2010) Disruptions in women's self-promotion: The backlash avoidance model. *Psychol Women Q* 34:186–202.
37. Rudman LA, Moss-Racusin CA, Glick P, Phelan JE (2012) *Advances in Experimental Social Psychology*, eds Devine P, Plant A (Elsevier, New York), pp 167–227.
38. Swim JK, Aikin KJ, Hall WS, Hunter BA (1995) Sexism and racism: Old-fashioned and modern prejudices. *J Pers Soc Psychol* 68:199–214.
39. Martell RF, Lane DM, Emrich C (1996) Male-female differences: A computer simulation. *Am Psychol* 51(2):157–158.
40. Steinpreis RE, Anders KA, Ritzke D (1999) The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles* 41:509–528.
41. Bandura A (1982) Self-efficacy mechanism in human agency. *Am Psychol* 37:122–147.
42. Scandura TA, Lankau MJ (1997) Relationships of gender, family responsibility and flexible work hours to organizational commitment and job satisfaction. *J Organ Behav* 18:377–391.
43. Rudman LA, Ashmore RD, Gary ML (2001) "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *J Pers Soc Psychol* 81:856–868.
44. O'Neill OA, O'Reilly CA (2011) Reducing the backlash effect: Self-monitoring and women's promotions. *J Occup Organ Psychol* 84:825–832.
45. Uhlmann EL, Cohen GL (2005) Constructed criteria: Redefining merit to justify discrimination. *Psychol Sci* 16:474–480.
46. Phelan JE, Moss-Racusin CA, Rudman LA (2008) Competent yet out in the cold: Shifting criteria for hiring reflects backlash toward agentic women. *Psychol Women Q* 32:406–413.
47. Holbrook AL, Krosnick JA, Pfent A (2007) *Advances in Telephone Survey Methodology*, eds Lepkowski JM, et al. (John Wiley & Sons, Hoboken), pp 499–528.
48. Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* 22:1359–1366.
49. Bargh JA, Chartrand TL (2000) *Handbook of Research Methods in Social and Personality Psychology*, eds Reis HT, Judd CM (Cambridge Univ Press, New York), pp 253–285.
50. Brescoll VL, Uhlmann EL (2008) Can angry women get ahead? Status conferral, gender, and workplace emotion expression. *Psychol Sci* 19:268–275.
51. Cohen J (1998) *Statistical Power Analysis for the Behavioral Sciences* (Erlbaum, Hillsdale), 2nd Ed.

Now Hiring! Empirically Testing a Three-Step Intervention to Increase Faculty Gender Diversity in STEM

JESSI L. SMITH, IAN M. HANDLEY, ALEXANDER V. ZALE, SARA RUSHING, AND MARTHA A. POTVIN

Workforce homogeneity limits creativity, discovery, and job satisfaction; nonetheless, the vast majority of university faculty in science, technology, engineering, and mathematics (STEM) fields are men. We conducted a randomized and controlled three-step faculty search intervention based in self-determination theory aimed at increasing the number of women faculty in STEM at one US university where increasing diversity had historically proved elusive. Results show that the numbers of women candidates considered for and offered tenure-track positions were significantly higher in the intervention groups compared with those in controls. Searches in the intervention were 6.3 times more likely to make an offer to a woman candidate, and women who were made an offer were 5.8 times more likely to accept the offer from an intervention search. Although the focus was on increasing women faculty within STEM, the intervention can be adapted to other scientific and academic communities to advance diversity along any dimension.

Keywords: diversity, women in science, science education and policy, behavioral science

A homogenous university faculty limits student and faculty creativity, discovery, and satisfaction (Page 2007, Apfelbaum et al. 2014), whereas diversity in science furthers social justice, expands workforce talent, and increases objectivity (Intemann 2009). However, university faculty are largely homogenous on the salient dimension of gender, because the majority of faculty at all ranks worldwide are men, especially within science, technology, engineering, and mathematics (STEM) fields (NSB 2012, European Commission 2013). For example, 68% to 89% of all academic grade C to grade A STEM personnel in the EU are men, and 81% of tenure-track STEM faculty at US public and land grant universities are men (European Commission 2013, Oklahoma State University 2013). Therefore, increasing gender diversity among STEM faculty is one straightforward way to enhance science education and scientific research innovation.

What is less straightforward are the reasons why STEM fields are male dominated and what can be done to enhance diversity. There is a tendency to blame “the pipeline” because few women candidates populate STEM-faculty search pools. It is true that fewer and fewer women advance at every transition point from secondary school to college to graduate study such that proportionally fewer women are qualified for STEM faculty positions than men (McCook 2011, NSB 2012). However, social psychological factors, such as implicit gender biases among university faculty and administrators that favor men in STEM, may inadvertently perpetuate homogeneity (Moss-Racusin et al. 2012, Shen 2013). Fortunately, educational programs could potentially actively counter this bias. What is more, search committees typically do not understand how to recruit and attract diverse candidates. For example, many assume that the competition for diverse candidates is fierce among institutions and therefore do not undertake efforts to broaden the pool of applicants. This scenario is consistent with social-judgment biases such as the false-consensus effect (Ross 1977), which occurs when people overestimate the extent to which others believe as they do. As a case in point, only 29% of white women who had won prestigious fellowships in the United States (Ford, Mellon, or Spencer fellows) received multiple tenure-track job offers for positions they desired; the majority of these women (71%) did not receive multiple offers or had limited choices among less than ideal offers (Smith DG et al. 1996). Acquiescence that universities cannot diversify their faculty is a form of system justification that ultimately maintains the homogenous status quo (Jost et al. 2004). Offering search committees concrete best-practice techniques to address these psychological considerations could potentially enhance diversity. Finally, search committees must understand that partner accommodations and other work–life integration

BioScience 65: 1084–1087.
doi:10.1093/biosci/biv138 Advance Access publication 10 October 2015

issues are central to recruiting women, because 83% of women scientists in academia have partners also in academic science (Schiebinger et al. 2008, Moors et al. 2014).

We designed an intervention to overcome these challenges. As Timothy Wilson noted in his 2006 *Science* article on the power of social psychological interventions, "Brief theory-based interventions that focus on people's construals can reap large benefits" (Wilson 2006). Intervening in the faculty search process is therefore one potential way to enhance the representation of women STEM faculty at an institution. Past intervention efforts to enhance gender diversity in academia focused mostly on the pipeline issue by supporting women students to perform well in, pursue, and persist in STEM (Hullerman and Harackiewicz 2009, Miyake et al. 2010, Moss-Racusin et al. 2012, Smith JL et al. 2013). One notable exception was a detailed case study of an ecology faculty search employing intuitive (albeit effort-intensive) gender-blind applicant tracking that achieved partial success (Jones and Urban 2013). Theory-driven, randomized control trials aimed at enhancing diversity are relatively rare in intervention research (Moss-Racusin et al. 2014), and few studies on the search process include faculty as participants (e.g., Stewart et al. 2004, Carnes et al. 2012, Fine et al. 2014). We designed and empirically tested an intervention guided by the tenets of self-determination theory (Deci and Ryan 2000) aimed at enhancing the recruitment processes for multiple and varied STEM-faculty search committees.

Self-determination theory (Deci and Ryan 1985, 2000) proposes that creativity, motivation, and performance thrive when three particular psychological needs are satisfied: to engage in opportunities for learning and mastery (*competency*), to have flexibility and control over processes and outcomes (*autonomy*), and to make meaningful connections with others (*relatedness*). Informed by this theory, we designed a three-step faculty search intervention to supplement the mandatory human resources (HR) training that would (1) enhance the *competence* of the search committee by delivering concrete strategies for conducting a broad applicant search in the form of a printed "faculty search toolkit," (2) enhance the *autonomy* of the search committee by showing them how to gain better control over possible unintentional biases in their decisionmaking through a 30-minute oral presentation by a faculty member on the role of implicit gender bias in skewing the candidate-screening and interview processes, and (3) enhance the *relatedness* of the search process more generally by both connecting the search committee with a peer faculty member who was supportive during the entire search process and by specifically connecting job finalists with a faculty "family advocate" totally independent from the search for a confidential 15-minute conversation. The faculty family advocate meetings were designed to meet all Equal Employment Opportunity rules by including all finalists, providing an overview of policies and practices without inquiring directly about a candidate's marital or family status, and maintaining the confidentiality of any information shared through the discussion of work–life related questions. Family-advocate conversations were in no way communicated to the search committee nor had any bearing on the hiring decision.

The search committees in the *no-intervention* (status-quo) condition received only the mandatory HR training. This brief in-person training was conducted by an assigned staff member from HR. The HR staff person provided a packet of handouts that outlined compliance issues (e.g., must have at least two people on every phone reference check) and procedure issues (e.g., how to submit paperwork for the web-posting of the vacancy advertisement). The HR training did include a brief overview of antidiscrimination law, including a handout with a list of protected classes and a list of questions committees were not allowed to ask. The emphasis on this part of the HR training was on avoiding discrimination lawsuits by treating everyone equally, akin to the colorblind or gender-blind notion that gender or race "should not and does not matter" (Neville et al. 2000, p. 60), which is limited (Bagenstos 2006) and may lead, however inadvertently, to greater bias (Richeson and Nussbaum 2004). More details on the intervention and no-intervention conditions, including the family advocate, are outlined in the supplemental method S1 section; materials and facilitator guides are also freely available at *www.montana.edu/nsfadvance/resources.html*.

Our hypothesis was that search committees randomly assigned to the intervention, compared with the no-intervention, as-usual search procedures, would have an increased number of women candidates considered for and offered tenure-track positions in STEM.

Methodology

Our experiment took place across a broad discipline of 23 STEM-faculty searches during one academic year at Montana State University (MSU), a Carnegie Foundation–ranked *Very High Research Activity* (VHR) university in the United States (see methods S1 for more details). At the time, the 235 STEM faculty at MSU were largely homogenous (81% men), making this a representative context that mirrored national faculty gender statistics (Oklahoma State University 2013) in which to test our intervention. Moreover, the rural setting of the university, its low salaries (lowest among the 102 VHR ranked universities; Curtis and Thornton 2014), and the lack of a medical school also posed recruitment challenges, allowing for a strong test of the intervention. Our research is the first to use STEM faculty as participants in a hypothesis-testing study on diversity faculty hiring.

Search committee chairs were identified and invited via email by a faculty peer to voluntarily participate in a supplemental training to coincide, if possible, with the mandatory human resource–search committee training, which all committees received (see supplemental methods and discussion S1). None refused to participate. The selection of a faculty peer to contact search committee chairs and to present the intervention material were intentional to increase participation (see discussion S1). Presenting material to each search committee separately ensured a small group setting meant to enhance engagement with the presentation.

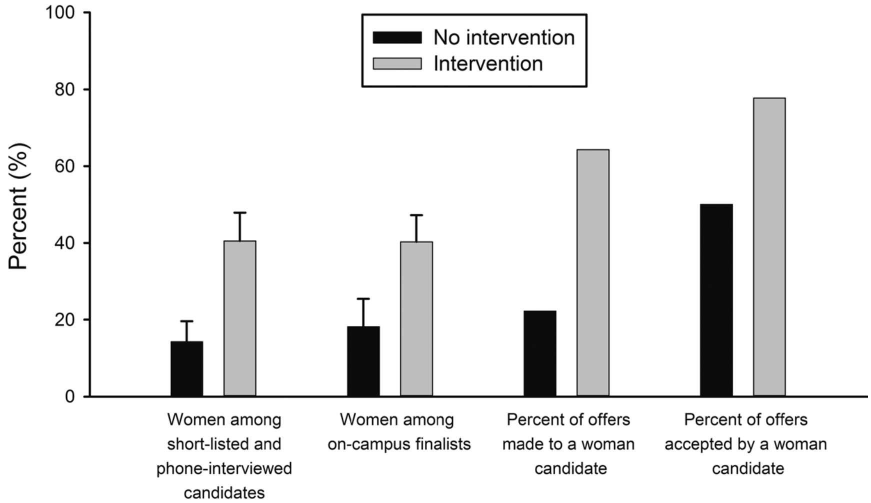

Figure 1. Mean percentages of women interviewed at two points in the science, technology, engineering, and mathematics (STEM) faculty search process and simple percentages of tenure-track job offers extended to and accepted by women, by intervention group. The error bars represent the standard error.

Results

The three-step intervention was successful. Among searches in the intervention condition, more applicants overall were short-listed and phone-interviewed (mean (M) = 9.5, standard error (SE) = 1.5) compared with those in the no-intervention condition (M = 4.7, SE = 1.3; Cohen's d = 0.99, $t(21)$ = 2.26, $p < .05$). Importantly, searches in the intervention condition phone-interviewed a significantly greater percentage of women applicants (M_{women} = 40.5%, SE = 7.4%) compared with searches in the no-intervention condition (M_{women} = 14.2%, SE = 5.4%; d = 1.16, $t(21)$ = 2.57, $p < .02$; figure 1), illustrating a large improvement in the representation of women on the short lists. Given that travel funding limits the number of finalists brought to campus for interviews in each search, no difference existed in the mean numbers of finalists brought to campus for interviews between searches in the intervention (M = 6.1, SE = 1.4) and no-intervention (M = 3.6, SE = 0.5; $p > .05$) groups. However, women made up a significantly greater percentage of on-campus interviewees for searches in the intervention group (M_{women} = 40.3%, SE = 6.9%) than in the no-intervention group (M_{women} = 18.2%, SE = 7.3%; d = 0.92, $t(21)$ = 2.12, $p < .05$), illustrating a large difference in the inclusion of women as finalists. Importantly, we ruled out alternative explanations and confirmed the effectiveness of our random assignment (see supplemental results and table S1).

Furthermore, 11 women were extended offers for tenure-track faculty positions—nine in the intervention condition and two in the no-intervention condition. Odds ratio statistics showed that a search in the intervention condition was 6.3 times more likely to make an offer to a woman candidate than a search in the no-intervention condition (d = 0.93; see figure 1). Moreover, women offered jobs were 5.8 times more likely to accept the offer from an intervention search (n = 7 accepted) than from a no-intervention search (n = 1 accepted; d = 0.80). The three-step intervention effectively increased the number of women hired as incoming STEM faculty at MSU. Subsequent application of our intervention to *all* STEM searches has continued this trend, with women representing precisely 50% of all STEM faculty hires with start dates in 2013–2014 academic year (n = 10 men and 10 women) and start dates in 2014–2015 academic year (n = 9 men and 9 women hired).

Conclusions

We tested a theory-derived three-step intervention that involved (1) a short presentation to search committees about overcoming the influence of unintentional (i.e., implicit) bias during the review process, (2) arming search committees with a guidebook on tactics for recruiting diverse candidates, and (3) providing access to a faculty family advocate who was unaffiliated with the search to confidentially discuss any work–life integration issues deemed appropriate by the candidates. The intervention measurably increased gender diversity among STEM faculty. Although the focus here was on increasing women faculty within STEM, the intervention can be adapted to other scientific and academic communities to advance diversity along any dimension.

Some pushback was experienced, as we expected, and a small number of male and female faculty expressed concerns that paying attention to gender diversity in STEM while conducting a faculty search was "lowering standards to fulfill a quota" (a sentiment that perfectly exemplifies gender bias). Indeed, a good next step would be to examine how faculty experience the intervention process itself (Moss-Racusin et al. 2014) versus the outcomes of the intervention as we reported here. For example, some faculty may believe that a focus on gender diversity is a form of reverse discrimination or that such a focus implies women are less competent and unable to make it on their own merits (Etzkowitz et al. 1994, Norton and Sommers 2011). Such mental frameworks probably have important ramifications for how people experience self-determination within what is perceived as a potentially threatening, high-stakes situation. Pushback notwithstanding, our brief three-step faculty search intervention was successful. We show that organizations can benefit from using psychological science to inform precise interventions. Although our data does not build on self-determination theory, it was inspired by and supports self-determination theory. Systematically testing theory through application can potentially contribute to theory-building in the future (e.g., Wilson 2006, Walton 2014). For example, future research could test which psychological need (competence, autonomy, or relatedness) was most essential to the success of the intervention and/or reveal the level at which it is important to foster psychological-need support, whether to the entire group (i.e., the search committee) or to an influential leader of the group (i.e., the search chair).

Worldwide, STEM funding agencies are investing heavily in diversifying the scientific workforce. As just two examples, the US National Science Foundation NSF ADVANCE-Institutional Transformation program and the European Commission genSET project have spent millions to bring about equality for women working in STEM. Our findings contribute to these important efforts. After all, a diverse faculty engenders social justice and betters the condition of underrepresented people working within STEM (Etzkowitz et al. 1994, Sekaquaptewa 2002). Diversity within STEM is essential for creating a thriving workplace and a learning environment replete with role models, diverse ways of thinking, and enhanced learning that elevates excellence and benefits scientific innovation, public health, and economic growth.

Acknowledgments

This research was supported by the National Science Foundation Grant no. 1208831. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The Montana Cooperative Fishery Research Unit is jointly sponsored by the US Geological Survey; Montana Fish, Wildlife, and Parks; Montana State University; and the US Fish and Wildlife Service. We are grateful for the assistance of Kelli J. Klebe, Rebecca Belou, and Theresa Marchwick in particular and thankful for the support of the entire ADVANCE Project TRACS team. Portions of these data were also presented at the American Association for the Advancement of Science Conference in February 2015, in San Jose, California.

Supplemental material

The supplemental material is available online at *http://bioscience.oxfordjournals.org/lookup/suppl/doi:10.1093/biosci/biv138/-/DC1*

References cited

Apfelbaum EP, Phillips K, Richeson JA. 2014. Rethinking the baseline in diversity research: Should we be explaining the effects of homogeneity? Perspectives in Psychological Science 9: 235–244.

Bagenstos SR. 2006. The structural turn and the limits of antidiscrimination law. California Law Review 94: 1–47.

Carnes M, Devine PG, Isaac C, Manwell LB, Ford CE, Byars-Winston A, Fine E, Sheridan JT. 2012. Promoting institutional change through bias literacy. Journal of Diversity in Higher Education 5: 63–77. doi:10.1037/a0028128

Curtis JW, Thornton S. 2014. Losing Focus: The Annual Report on the Economic Status of the Profession, 2013–2014. American Association of University Professors. (9 September 2015; *www.aaup.org/sites/default/files/files/2014%20salary%20report/zreport_0.pdf*)

Deci EL, Ryan RM. 1985. Intrinsic Motivation and Self-determination in Human Behavior. Plenum Press.

———. 2000. The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry 11: 227–268.

Etzkowitz H, Kemelgor C, Neuschatz M, Uzzi B, Alonzo J. 1994. The paradox of critical mass for women in science. Science 266: 51–54.

European Commission. 2013. She Figures 2012: Gender in Research and Innovation, Statistics and Indicators. Publications Office of the European Union. (9 September 2015; *http://ec.europa.eu/research/science-society/document_library/pdf_06/she-figures-2012_en.pdf*)

Fine E, Sheridan J, Carnes M, Handelsman J, Pribbenow C, Savoy J, Wendt A. 2014. Minimizing the influence of gender bias on the faculty search process. Pages 267–289 in Demos V, Berheide CW, Segal MT, eds. Gender Transformation in the Academy. (Advances in Gender Research, vol. 19). Emerald Group Publishing.

Hulleman CS, Harakiewicz JM. 2009. Promoting interest and performance in high school science classes. Science 326: 1410–1412.

Intemann K. 2009. Why diversity matters: Understanding and applying the diversity component of the National Science Foundation’s Broader Impacts Criterion. Social Epistemology 23: 249–266.

Jones CS, Urban MC. 2013. Promise and pitfalls of a gender-blind faculty search. BioScience 63: 611–612.

Jost JT, Banaji M, Nosek BN. 2004. A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. Political Psychology 25: 881–919.

McCook A. 2011. Education: Rethinking PhDs. Nature 472: 280–282.

Miyake A, et al. 2010. Reducing the gender achievement gap in college science: A classroom study of values affirmation. Science 330: 1234–1237.

Moors AC, Malley JE, Stewart AJ. 2014. My family matters: Gender and perceived support for family commitments and satisfaction in academia among postdocs and faculty in STEMM and non-STEMM fields. Psychology of Women Quarterly 38: 460–474.

Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J. 2012. Science faculty’s subtle gender biases favor male students. Proceedings of the National Academy of Sciences 109: 16474–16479.

Moss-Racusin CA, et al. 2014. Scientific diversity interventions. Science 343: 615–616.

Neville HA, Lilly RL, Duran G, Lee R, Browne L. 2000. Construction and initial validation of the Color-Blind Racial Attitudes Scale (CoBRAS). Journal of Counseling Psychology 47: 59–70.

Norton MI, Sommers SR. 2011. Whites see racism as a zero-sum game that they are now losing. Perspectives in Psychological Science 6: 215–218.

[NSB] National Science Board. 2012. Science and Engineering Indicators 2012. National Science Foundation Report no. NSB 12-01. (9 September 2015; *www.nsf.gov/statistics/seind12/*)

Oklahoma State University. 2013. The 2012–2013 Faculty Distribution Survey of Selected Criteria. Office of Institutional Research and Information Management, Oklahoma State University.

Page SE. 2007. The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies. Princeton University Press.

Richeson JA, Nussbaum RJ. 2004. The impact of multiculturalism versus color-blindness on racial bias. Journal of Experimental Social Psychology 40: 417–423.

Ross L. 1977. The false consensus effect: An egocentric bias in social perception and attribution processes. Journal of Experimental Social Psychology 13: 279–301.

Schiebinger L, Henderson AD, Gilmartin SK. 2008. Dual-career academic couples: What universities need to know. Michelle R. Clayman Institute for Gender Research, Stanford University. (9 September 2015; *http://gender.stanford.edu/sites/default/files/DualCareerFinal_0.pdf*)

Sekaquaptewa D, Thompson M. 2002. The differential effects of solo status on members of high- and low-status groups. Personality and Social Psychology Bulletin 28: 694–707.

Shen H. 2910. Inequality quantified: Mind the gender gap. Nature 495: 22–24.

Smith DG, Wolf LE, Busenberg BE. 1996. Achieving Faculty Diversity: Debunking the Myths. Association of American Colleges and Universities.

Smith JL, Lewis KL, Hawthorne L, Hodges SD. 2013. When trying hard isn’t natural: Women’s belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns. Personality and Social Psychology Bulletin 39: 131–143.

Stewart AJ, La Vaque-Manty D, Malley JE. 2004. Recruiting female faculty members in science and engineering: Preliminary evaluation of one intervention model. Journal of Women and Minorities in Science and Engineering 10: 361–375. doi:10.1615/JWomenMinorScienEng.v10.i4.50

Walton, G. M. The new science of wise psychological interventions. Current Directions in Psychological Science 23: 73–82.

Wilson T. 2006. The power of social psychological interventions. Science 313: 1251–1252.

Jessi L. Smith (jsismith@montana.edu) and Ian M. Handley are affiliated with the Department of Psychology at Montana State University, in Bozeman. Alexander V. Zale is affiliated with the US Geological Survey, the Montana Cooperative Fishery Research Unit, and the Department of Ecology at Montana State University, in Bozeman. Sara Rushing is affiliated with the Department of Political Science and Martha A. Potvin is with the Office of the Provost and the Department of Ecology at Montana State University, in Bozeman.

Quality of evidence revealing subtle gender biases in science is in the eye of the beholder

Ian M. Handley[a,1], Elizabeth R. Brown[a,b], Corinne A. Moss-Racusin[c], and Jessi L. Smith[a]

[a]Department of Psychology, Montana State University, Bozeman, MT 59717-3440; [b]Department of Psychology, University of North Florida, Jacksonville, FL 33224; and [c]Department of Psychology, Skidmore College, Saratoga Springs, NY 12866

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved September 16, 2015 (received for review May 31, 2015)

Scientists are trained to evaluate and interpret evidence without bias or subjectivity. Thus, growing evidence revealing a gender bias against women—or favoring men—within science, technology, engineering, and mathematics (STEM) settings is provocative and raises questions about the extent to which gender bias may contribute to women's underrepresentation within STEM fields. To the extent that research illustrating gender bias in STEM is viewed as convincing, the culture of science can begin to address the bias. However, are men and women equally receptive to this type of experimental evidence? This question was tested with three randomized, double-blind experiments—two involving samples from the general public (n = 205 and 303, respectively) and one involving a sample of university STEM and non-STEM faculty (n = 205). In all experiments, participants read an actual journal abstract reporting gender bias in a STEM context (or an altered abstract reporting no gender bias in experiment 3) and evaluated the overall quality of the research. Results across experiments showed that men evaluate the gender-bias research less favorably than women, and, of concern, this gender difference was especially prominent among STEM faculty (experiment 2). These results suggest a relative reluctance among men, especially faculty men within STEM, to accept evidence of gender biases in STEM. This finding is problematic because broadening the participation of underrepresented people in STEM, including women, necessarily requires a widespread willingness (particularly by those in the majority) to acknowledge that bias exists before transformation is possible.

gender bias | science workforce | diversity | science education | sexism

Objectivity is a fundamental value in the practice of science and is required to optimally assess one's own research findings, others' findings, and the merits of others' abilities and ideas (1). For example, when scientists evaluate data collected on a potentially controversial topic (such as climate change), they strive to set aside their own belief systems and instead focus solely on the strength of the data and conclusions warranted. Similarly, when scientists evaluate a resume for a laboratory-manager position or assess the importance of a conference submission, the gender of the applicant or author should be immaterial. If they are truly objective, scientists should focus only on the relevant criteria of applicant qualifications or research merit.

However, despite rigorous training in the objective evaluation of information and resultant values (2), people working and learning within the science, technology, engineering, and mathematics (STEM) community are still prone to the same subtle biases that subvert objectivity and distort accurate perceptions of scientific evidence by the general public (3, 4). We focus here on the robust gender biases documented repeatedly within the psychological literature (5–7). Some within the STEM community have turned to these methods and ideas as an explanation for the consistent underrepresentation of women in STEM fields (8, 9) and the undervaluation of these women and their work. Specifically, many scientists have systemically documented and reported (including in *PNAS*) a gender bias against women—or favoring men—in STEM contexts (10–17), including hiring decisions for a laboratory-manager position (10) and selection for a mathematical task (11), evaluations of conference abstracts (12), research citations (13), symposia-speaker invitations (14), postdoctoral employment (15), and tenure decisions (16). For example, Moss-Racusin et al. (10) conducted an experiment in which university science professors received the same application for a laboratory-manager position, either associated with a male or female name through random assignment. The results demonstrated that the science professors—regardless of their gender—evaluated the applicant more favorably if the applicant had a man's name compared to a woman's name. These findings mirror past results in which men and women psychology faculty participants evaluated an application from a faculty candidate with a woman's name less favorably than the identical application with a man's name (17). As another example, Knobloch-Westerwick et al. (12) found that graduate students evaluate science-related conference abstracts more positively when attributed to a male relative to a female author, particularly in male-gender-typed science fields. These biases are frequently unintentional (18–20), exhibited even by individuals who greatly value fairness and view themselves as objective (21). Indeed, gender biases often result from unconscious processes (22, 23) or manifest so subtly that they escape notice (24).

However unintentional or subtle, systematic gender bias favoring male scientists and their work could significantly hinder scientific progress and communication (12). In fact, the evidence for a gender bias in STEM suggests that our scientific community is not living up to its potential, because homogenous workforces (including the academic workplace) can deplete the creativity, discovery, and satisfaction of workers, faculty, and students (25–27). STEM fields are fairly homogeneously male; at 4-y US colleges, for example, an average of 71% of STEM faculty are men

Significance

Ever-growing empirical evidence documents a gender bias against women and their research—and favoring men—in science, technology, engineering, and mathematics (STEM) fields. Our research examined how receptive the scientific and public communities are to experimental evidence demonstrating this gender bias, which may contribute to women's underrepresentation within STEM. Results from our three experiments, using general-public and university faculty samples, demonstrated that men evaluate the quality of research unveiling this bias as less meritorious than do women. These findings may inform and fuel self-correction efforts within STEM to reduce gender bias, bolster objectivity and diversity in STEM workforces, and enhance discovery, education, and achievement.

Author contributions: I.M.H., E.R.B., C.A.M.-R., and J.L.S. designed research; E.R.B. and J.L.S. performed research; I.M.H., E.R.B., and J.L.S. analyzed data; and I.M.H., E.R.B., C.A.M.-R., and J.L.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: ihandley@montana.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1510649112/-/DCSupplemental.

(28). For these reasons, there is a growing call for broadening the participation of women (and other underrepresented groups) in STEM fields. For instance, the National Science Foundation (NSF) promoted inclusiveness as a core value in its 2014–18 strategic plan, continues to fund ADVANCE Institutional Transformation grants to broaden the participation of women faculty in STEM, and has created a directorate charged with broadening the participation of all underrepresented people within STEM. Similarly, the National Institutes of Health called for reducing subtle biases and broadening participation in STEM fields (29) and issued at least three large new requests for proposals to help accomplish this goal (30). Indeed, there are growing numbers of research studies, calls to action, strategic plans, and even resources to systematically document, understand, and hopefully ameliorate gender biases within STEM to create a thriving, diverse, and equitable scientific community (31–34). However, are people generally (e.g., taxpayers, voters, government officials, etc.) and STEM practitioners in particular "buying" the mounting evidence of these gender biases within the STEM community? Currently, to our knowledge, there is no experimental research examining how receptive or biased various individuals within the STEM and public communities are to research demonstrating gender bias that undermines women's participation within STEM. Thus, to address this question, our experimental research investigates potentially biased evaluations among the general public and STEM practitioners of evidence demonstrating gender biases against women/favoring men within STEM fields.

Of course, to ameliorate gender bias within STEM fields, it is not sufficient to simply herald findings demonstrating that STEM practitioners exhibit these biases. Indeed, there may well be another layer of bias such that men evaluate findings such as those reported by Moss-Racusin et al. (10) and Knobloch-Westerwick et al. (12) less favorably than women. In fact, a recent (nonexperimental) analysis of naturally occurring online comments written by readers of popular press articles covering the research of Moss-Racusin et al. (10) suggests that men were more likely than women to demonstrate negative reactions to experimental evidence of gender bias (35). Further, several lines of theorizing suggest that men may evaluate such research as less meritorious than would women (24, 36–42). Among these theories, Social Identity Theory (36–38) and related perspectives (39) posit that people are motivated to perceive their group favorably and defend that perception against threat, and that people within privileged groups often seek to retain and justify their privileged status (39). Men clearly hold an advantageous position within the sciences, because they represent the vast majority of STEM university faculty at all ranks, earn higher salaries controlling for rank and related factors (43), and on average receive more federal grant funding to support their research than their comparable women colleagues (44, 45). Indeed, growing evidence reveals an often invisible advantage for men, stemming in part from inequities against women in STEM, which can threaten that advantage (10, 12, 46, 47). That is, men might find the results reported by Moss-Racusin et al. (10) threatening, because remedying the gender bias in STEM fields could translate into favoring women over men, especially if one takes a zero-sum-gain perspective (47). Therefore, relative to women, men may devalue such evidence in an unintentional implicit effort (18–20) to retain their status as the majority group in STEM fields. However, some men might perceive research that exposes gender bias in STEM as more threatening than other men. According to Social identity Theory, individuals perceive greater threat toward their group (and defend against it) when they are highly committed to that group (37, 38). Thus, men within STEM fields (e.g., physics professors) may feel more threated by the research of Moss-Racusin et al. (10) than men within non-STEM fields (e.g., English professors), assuming they are more committed to STEM fields and men's status therein. Thus, men overall relative to women are likely to devalue research demonstrating bias against women in STEM, but this difference may be prominent among individuals within (and committed to) STEM fields, and weaker to nonexistent among individuals within non-STEM fields.

Beyond Social Identity Theory, other frameworks could predict a difference between men's and women's evaluation of research demonstrating bias against women in STEM, and, in fact, this difference might result from multiple factors. For instance, the predicted gender difference may also result from a confirmation bias such that people favorably evaluate information that is consistent with their beliefs, but unfavorably evaluate information that is inconsistent with their beliefs (48). A classic empirical example of confirmation bias showed that peer-reviewers were less favorable toward an essentially identical research manuscript when it was doctored to report results inconsistent with the reviewers' preferred theoretical viewpoint, but more favorable when it was doctored to report results consistent with the reviewers' preferred theoretical viewpoint (49). Add to this finding that there is compelling evidence that women faculty are more likely to view gender bias as a problem within their current working academic context (40), and it is possible that women may evaluate research demonstrating a gender bias (belief consistent) more favorably than men, but evaluate research demonstrating no gender bias (belief inconsistent) less favorably than men.

Current Research

We report three experiments designed to provide, to our knowledge, the first test for gender differences in the evaluation of scientific evidence demonstrating that individuals are biased against women within STEM contexts. In each experiment, men and women participants read via an online survey instrument an actual article abstract from a peer-reviewed scientific journal, accompanied by the date and title of the publication (see *Materials and Methods* for more details). Participants then evaluated their agreement with the authors' interpretation of the results, the importance of the research, and how well-written and favorable they found the quality of the abstract. These ratings were highly associated with one another and were averaged to create a measure of participants' overall evaluation of the abstract (for further details, see *SI Materials and Methods*, *Dependent Variables*). Globally, we predicted that male relative to female participants would evaluate the abstract less favorably when the abstract reported a gender bias against women in STEM (hypothesis A; experiments 1–3), and that this difference would be more prominent among participants in STEM (vs. non-STEM) fields, to whom a gender bias in STEM is most germane (hypothesis B; experiment 2). Further, we predicted that this gender difference would manifest for abstracts that reported a gender bias in STEM, but would reverse for abstracts that reported no gender bias in STEM (hypothesis C; experiment 3).

All experiments included 2 or more factors (some for exploratory purposes in Experiments 1 and 2; see *SI Materials and Methods* for more details), and thus we tested all hypotheses using between-groups factorial analyses of variance. Further, we calculated Cohen's d for each experiment to provide an index of strength for the predicted difference between men and women participants and to account for the unequal sample sizes between the genders. As per convention (50), effect sizes can range from small (d = 0.20), to medium (d = 0.50), to large (d = 0.80).

The first two experiments tested for participant-gender differences in the evaluation of the actual abstract written by Moss-Racusin et al. (10). As discussed above, Moss-Racusin et al. (10) produced experimental evidence that STEM faculty of both genders demonstrate a significant bias against an identical applicant with a female vs. male name. Although this gender bias was empirically demonstrated with a national sample, we predicted that men would be less receptive to these (and related) findings, and women more receptive. Our first experiment involved a general sample of US adults ($n = 205$) recruited online through Amazon's Mechanical Turk. Our second experiment involved a sample of professors ($n = 205$) from all STEM and non-STEM departments at a research-intensive university, allowing us to test whether the predicted gender difference in abstract evaluations is larger among individuals within STEM fields of study. A third experiment replicated the first two with a different abstract and is discussed in more detail below.

Results

Experiments 1 and 2. Results from our experiment 1 supported hypothesis A, revealing a main effect of participant gender [$F(1, 197) = 9.85$, $P = 0.002$, $\eta^2_{partial} = 0.048$], such that men ($M = 4.25$, $SD = 0.91$, $n = 146$) evaluated the research less favorably than women ($M = 4.66$, $SD = 0.93$, $n = 59$) in a general sample. Further, this effect was of moderate size ($d = 0.45$).

Results from our experiment 2 also supported hypothesis A, revealing a main effect of participant gender [$F(1, 174) = 6.08$, $P = 0.015$, $\eta^2_{partial} = 0.034$], such that male faculty evaluated the research less favorably ($M = 4.21$, $SD = 1.05$) than female faculty ($M = 4.65$, $SD = 1.19$; $d = 0.397$ [similar to experiment 1]). Thus, overall, experiments 1 and 2 provide converging evidence from multiple participant populations that men are less receptive than women—and by the same token, that women are more receptive than men—to experimental evidence of gender bias in STEM. Importantly, results from experiment 2 further reveal that this effect was qualified by a significant interaction between participant gender and field of study [$F(1, 174) = 5.19$, $P = 0.024$, $\eta^2_{partial} = 0.03$]. This interaction supported hypothesis B, because simple-effect tests confirmed that male faculty evaluated the research less favorably ($M = 4.02$, $SD = 0.988$, $n = 66$) than female faculty ($M = 4.80$, $SD = 1.14$, $n = 38$) in STEM fields [$F(1, 174) = 11.94$, $P < 0.001$], whereas male ($M = 4.55$, $SD = 1.09$, $n = 37$) and female ($M = 4.54$, $SD = 1.23$, $n = 49$) faculty reported comparable evaluations in non-STEM fields ($F < 1$). Further, the effect size for the observed gender difference was large within STEM departments ($d = 0.74$). Looking at this interaction another way, simple-effect tests demonstrated that men evaluated the research more negatively if they were in STEM than non-STEM departments [$F(1, 174) = 4.19$, $P = 0.042$], whereas the opposite trend was not statistically significant among female faculty [$F(1, 174) = 1.45$, $P = 0.23$]. Thus, it seems that men in STEM displayed harsher judgments of Moss-Racusin et al.'s (10) research, not that women in STEM exhibited more positive evaluations of it. The analysis revealed one other significant interaction that did not involve faculty gender (for further details, see *SI Additional Analyses*, *Experiment 2*). No other main effects or interactions reached significance (all other $F < 2.07$; $P > 0.15$). Finally, additional measures collected within a faculty survey (*SI Materials and Methods*, *Dependent Variables*) and analyses thereof provide suggestive evidence for a threat mechanism behind the effects (for the analyses and discussion, see *SI Additional Analyses*, *Experiment 2*).

Experiment 3. We predicted that, compared with women, men would be prone to more negative evaluations of research that demonstrates a gender bias against women (and favors men) in STEM, not just the specific research reported by Moss-Racusin et al. (10). Further, we predicted that, compared with men, women would be prone to more negative evaluations of research that demonstrates no gender bias against women in STEM. Thus, the gender effect seen in experiments 1 and 2 should replicate for a different abstract that also reports a gender bias, but reverse for an abstract that demonstrates no gender bias. Testing these predictions, we randomly assigned new participants to read either the original abstract published by Knobloch-Westerwick et al. (12) which reported a gender bias against women's (relative to men's) scientific conference submissions, or a version slightly altered to report no gender bias. These participants were recruited online through Amazon's Mechanical Turk ($n = 303$). Results indicated only a significant interaction between participant gender and abstract version [$F(1, 299) = 4.00$, $P = 0.046$, $\eta^2_{partial} = 0.013$] (all other $F < 1$). Although no simple-effect tests were significant (all $F < 2.69$, $P > 0.10$), together, these results support the overall pattern predicted by hypothesis C, such that that men evaluated the original (gender-bias exists) abstract less favorably ($M = 3.65$, $SD = 1.03$, $n = 78$) than did women ($M = 3.86$, $SD = 1.05$, $n = 74$; $d = 0.20$), whereas men evaluated the modified (no gender-bias exists) abstract more favorably ($M = 3.83$, $SD = 0.92$, $n = 84$) than did women ($M = 3.59$, $SD = 0.86$, $n = 67$; $d = 0.27$).

Discussion

There is now copious evidence that women are disadvantaged in STEM fields (51–53) and that this disadvantage may relate to gender stereotypes (11) and consequent biases against women (or favoring men) traversing the STEM pipeline (10–17). Of course, people should not passively accept such evidence, even if it appears in preeminent peer-reviewed journals (e.g., *Science*, *PNAS*, or *Nature*)—suggesting the quality of the research was sound. Ideally, especially within the STEM community, people should evaluate as objectively as possible the research producing such evidence, the resulting quality of the evidence, and the interpretation of that evidence.

However, the evidence from our three straightforward experiments indicates than men evaluate research that demonstrates bias against women in STEM less favorably than do women—or, that women evaluate it more favorably. Specifically, male relative to female participants (including university faculty) in experiments 1 and 2 assessed the quality of the research by Moss-Racusin et al. (10)—as presented simply through their actual abstract—as being lower. In addition, perhaps of greatest concern, this gender difference and accompanying effect size was large among faculty working within STEM fields (50) and nonexistent among faculty from non-STEM fields (experiment 2). Further, the overall gender difference observed in the first two experiments was replicated among participants in experiment 3 who read the true abstract of Knobloch-Westerwick et al. (12), which also reported a gender bias in STEM. However, this gender difference was reversed among participants who read an altered version purporting no gender bias in STEM.

The results from this third experiment are important for at least three reasons. First, they indicate that men relative to women do not uniquely disfavor the research of Moss-Racusin et al. (10), but research that reports a gender bias hindering women in STEM. Second, these results suggest that men do not generally evaluate research more harshly than women, as it might seem from the first two experiments (but see the results from non-STEM faculty in experiment 2). Rather, relative to women, men actually favor research suggesting there is no gender bias in STEM. Finally, the results indicate that individuals are likely to demonstrate a gender bias toward research pertaining to the mere topic of gender bias in STEM; men seem to disfavor (and women favor) research demonstrating a gender bias, but women seem to disfavor (and men favor) research demonstrating no gender bias. Of course, given that we cannot have a gender-free control condition, it is important to note that these biases are relative to the other gender; we cannot conclude that one gender is more biased than the other, just that individuals' judgments of research regarding gender bias in STEM is biased by their gender.

Critically, across three experiments, we uncovered a gender difference in the way people from the general public and STEM faculty evaluate the quality of research that demonstrates women's documented disadvantage in STEM fields: Men think the research is of lower quality, whereas women think the research is of higher quality. Why does this gender difference matter? For one, there are significant implications for the dissemination and impact of meritorious previous, current, and future research on gender bias in STEM fields. Foremost, our research suggests that men will relatively disfavor—and women will relatively favor—research demonstrating this bias. Given that men dominate STEM fields throughout industry and academia, scholars whose program of research focuses on demonstrating gender bias in STEM settings might experience undue challenges for publication, have lower chances of publication in top-tier outlets, experience greater challenges in receiving tenure, and overall have lower-than-warranted impact on the thinking, research, and practice of those in STEM fields. Such possibilities are highly problematic and call for additional research evaluating biased reactions to scientific evidence demonstrating gender and/or racial biases within STEM.

Second, because men represent the majority of individuals in STEM fields and yet are less likely than women to acknowledge biases against women in STEM, it may be challenging to fully embrace the numerous calls to broaden the participation of women and minorities in STEM. How can we successfully broaden the participation of women in STEM when the very research underscoring the need for this initiative is less valued by the majority group who dominate and maintain the culture of STEM? Intensifying the challenge, men hold an advantageous position in STEM fields and may feel threatened by research and efforts to "level the playing field" for women. Similarly, people often unintentionally exhibit in-group favoritism (54), wherein individuals engage in behaviors and allocate resources in ways that benefit members of their group (e.g., men unintentionally conferring advantage to other men).

Fortunately, there are current efforts in place to meet these challenges. For example, "Project Implicit" (https://implicit.harvard.edu/implicit/) provides workshops and talks to reveal the subtlety and implicitness of gender bias and considers how to foster a broader recognition of these biases and address them. Further, NSF funds ADVANCE-Institutional Transformation grants to specifically facilitate the increased participation of women in STEM and help transform academic cultures to foster equality and inclusivity. Shields et al. (55) created a "WAGES" game and accompanying discussion platform that effectively highlights male privilege and advantage among STEM faculty and helps reduce reactance to acknowledging this advantage (56). Finally, Moss-Racusin et al. have developed an evidence-based framework for creating, evaluating, and implementing diversity interventions designed to increase awareness of and reduce bias across STEM fields (31). Initial evidence reveals promising results for interventions adhering to these guidelines (31). These efforts, along with others that can help individuals actually acknowledge evidence demonstrating gender bias in STEM, are critical in bringing about change and increasing the participation of women in STEM.

Limitations and Future Directions

As with any research, ours is met with limitations. First, we did not directly test the potential mechanisms behind the reported gender effect. However, even before we understand exactly why men are less favorable than women toward research demonstrating a gender bias in STEM, we suggest that is important for the STEM community to know that this phenomenon exists. However, we uncovered evidence in experiment 2 suggesting that men in STEM found the abstract of Moss-Racusin et al. (10) threatening (*SI Additional Analyses*, *Experiment 2*), which may be one possible explanation for the results (37). In the future, researchers could test this possibility by including a direct measure of how threatening people find the implications of various research results and multiple measures of social identity. It is also worth investigating in future research whether the confirmation bias (48, 49) contributes to the reported gender effect by measuring people's beliefs about gender bias in STEM before reading research demonstrating that bias. We hope our findings will spark future research thoroughly investigating the mechanisms underscoring this effect. Second, we investigated individuals' evaluations of two abstracts reporting gender bias in STEM, specifically within the contexts of evaluating a laboratory-manager application and conference abstracts. It is worthwhile to investigate whether this bias furthermore generalizes to evaluations of research that demonstrates gender bias in other STEM contexts, such as disparities in funding, publication rates, faculty and postdoctoral applicants, talk invitations, tenure decisions, and so forth. Theoretically, however, there is reason to predict that gender biases toward such research would replicate our current findings. In fact, because these contexts suggest a bias against (or in favor of) one's direct peers and colleagues, it seems likely that gender-biased evaluations of this research would be even more prominent. For instance, STEM faculty might find threatening the possibility that they are biased regarding the quality of research from their female colleagues and prefer (likely implicitly) to find fault with the research rather than face that possibility.

Third, we investigated individuals' assessment of research quality after they read only an abstract. We chose an abstract as a reasonable basis for assessment because abstracts present key methods and findings, are indexed and available for free, and are often what people read to determine whether or not they will read the full article. Nonetheless, it is conceivable that the gender bias we uncovered is a short-lived reaction. Perhaps the bias would shrink or disappear after reading the full article or a longer synopsis of the research. However, there is ample reason to predict that the bias will actually strengthen as people receive greater amounts of information, because they will (unintentionally) process that information based on initial impressions and per their motivation to arrive at a particular conclusion (42, 48, 49). However, we encourage future research into this issue.

As a final point on limitations, our experiments took place on an Internet platform, either at the end of a faculty survey that offered US$5 or as a short 10-min experiment paying $0.25. Thus, it is possible that our participants were not highly motivated to think about the abstract and thus simply based their quality assessments on "gut reactions" resulting in part from unconscious biases. Perhaps our findings would not hold among highly motivated participants whose assessments might have actual bearing on the publication of the research described in the abstract (e.g., peer reviewers). This hypothesis is certainly a possibility that warrants future exploration. However, we note that greater motivation does not always result in greater objectivity. In fact, biases can influence people's judgments even more so when they are motivated to be accurate, particularly if they do not notice that their thought process is biased (21, 42).

Further research might also explore why our first two experiments did not replicate previous research demonstrating an overall bias favoring the research of men above women in STEM (*SI Additional Analyses*). In particular, Knobloch-Westerwick et al. (12) found that graduate students evaluate science-related conference abstracts more positively when attributed to a male (relative to female) author, particularly in male-gender-typed fields. However, we did not find that participants in experiment 1 and 2 favored the abstract written by Moss-Racusin et al. (10) more if they thought it was written by a man vs. a women. It is possible that participants in our first two experiments found the topic of gender bias within STEM "feminine," or perhaps only somewhat "scientific," thus decreasing the bias toward the author's gender. Future research might reveal that participants' perception of gender-bias research plays an important role in producing biases against women—and favoring men—who conduct such research.

Conclusion

Failures in objectivity are problematic to specific research projects, science generally, and receptivity to discovery. However, objectivity is threatened by a multitude of cognitive biases, including gender bias in STEM fields. Numerous experimental findings confirm the existence of this bias, and the research we present here peels back yet another level of bias: Men evaluate the research that confirms gender bias within STEM contexts as less meritorious than do women. We hope that our findings help inform and fuel self-correction efforts within STEM to reduce this bias, bolster objectivity, and diversify STEM workforces. After all, the success of these efforts can translate into greater STEM discovery, education, and achievement (57).

Materials and Methods

Participants. In experiments 1 and 3, participation was solicited from workers on Amazon's Mechanical Turk online job site, who could view our employment opportunity listed alongside other opportunities. In experiment 1, a total of 205 individuals (146 men and 59 women) from the United States who were 18 y of age or older (M = 30.13; range = 18–66) opted to participate in the experiment and provided usable data (for more details, see *SI Materials and Methods*, *Participants and Recruitment for Experiments 1 and 3*). In experiment 3, a total of 303 individuals (162 men and 141 women) from the United States who were 18 y of age or older (M = 34.22; range = 18–79) opted to participate in the experiment and provided usable data. All participants engaged in the ∼10-min experiment in exchange for $0.25.

In experiment 2, participation was first solicited from all tenure-track faculty at a research-intensive American university via an email from their university provost encouraging participation in a larger baseline faculty climate survey. The survey and experiment were conducted on an Internet platform, during which time 506 tenure-track faculty from this university received the email invitation to participate. A total of 268 of these faculty participated in the survey, and 205 of these faculty further elected to participate in our experiment at the end of the survey. The resulting sample included faculty from all departments at the university, from STEM departments (n = 116) and non-STEM departments (n = 89; for more details, see *SI Materials and Methods*, *Participants and Recruitment for Experiment 2*). All participants received a $5 coupon for a local coffee shop and, if they elected, were entered into a raffle for 1 of 50 possible $50US gift certificates for the campus bookstore.

Procedure. All procedures were approved by the Montana State University institutional review board. The three experiments were approximately identical, although the experiment stood alone in experiments 1 and 3 and followed a faculty climate survey in experiment 2. All participants completed the experiment using a personal or work computer and received experiment materials, provided informed consent, and provided responses through surveymonkey.com.

Participants in experiments 1 and 2 were first instructed to read the actual abstract from the Moss-Racusin et al. (10) paper, which was provided in full on a single screen. The abstract was accompanied by that paper's actual title, publication date, volume and issue number, first author's full name, keywords, and a fictitious DOI. Further, participants were randomly assigned to receive a version of the abstract that either identified the first authors' first name as "Karen" or "Brian," which previous research indicates are equally likable and common names in the United States (58). Independent from this manipulation, participants received a version of the abstract that identified the author as affiliated with either Yale University (Moss-Racusin's true affiliation at the time of the publication) or Iowa State University. After reading the abstract and affiliated information, participants were asked to provide ratings on several scales (adapted from scales commonly used to gauge attitude change and evaluations of persuasive materials) assessing the quality of the abstract and the research provided therein (for details, see *SI Materials and Methods*, *Dependent Variables*). Participants also provided demographic information, including their gender. Participants' responses were anonymous, but in experiment 2 their status as a STEM or non-STEM faculty member was identifiable using specialized codes. Overall, the research design allowed us to analyze participants' quality assessments of the Moss-Racusin et al. (10) research as a function of participant gender, author gender, author affiliation, and participants' STEM affiliation (experiment 2 only).

Participants in experiment 3 completed a similar procedure, with some key differences. First, participants were randomly assigned to read either the original version of the abstract by Knobloch-Westerwick et al. (12), which reported a gender bias, or a version slightly altered to report no gender differences. Second, the abstract was not accompanied by the author's name or affiliation (as was done in experiments 1 and 2). Otherwise, the procedures and dependent measures for this experiment are identical to those used in the previous experiments. This research design allowed us to analyze participants' quality assessments of the research by Knobloch-Westerwick et al. (10) as a function of participant gender and abstract version (reporting gender bias or no gender bias).

ACKNOWLEDGMENTS. We thank the social science research team (especially Rebecca Belou) and project management staff of ADVANCE Project TRACS (Transformation through Relatedness, Autonomy, and Competence Support) for their efforts. This work was supported in part by National Science Foundation Grant 1208831 (to J.L.S.).

1. Institute of Medicine, National Academy of Science, and National Academy of Engineering (1992) *Responsible Science: Ensuring the Integrity of the Research Process* (National Academy, Washington, DC), Vol 1. Available at www.nap.edu/openbook.php?isbn=0309047315. Accessed September 11, 2014.
2. de Melo-Martín I, Intemann K (2012) Interpreting evidence: why values can matter as much as science. *Perspect Biol Med* 55(1):59–70.
3. Carnes M, et al. (2012) Promoting institutional change through bias literacy. *J Divers High Educ* 5(2):63–77.
4. Gilovich T, Griffin D, Kahneman D, eds (2002) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge Univ Press, Cambridge, UK).
5. Al-Gazali L (2013) Remove social barriers. *Nature* 495:35–36.
6. Ceci SJ, Williams WM (2011) Understanding current causes of women's underrepresentation in science. *Proc Natl Acad Sci USA* 108(8):3157–3162.
7. Shen H (2013) Inequality quantified: Mind the gender gap. *Nature* 495(7439):22–24.
8. National Science Board (2014) *Science and Engineering Indicators 2014* (National Center for Science and Engineering Statistics, Arlington, VA). Available at www.nsf.gov/statistics/seind14/. Accessed September 11, 2014.
9. Snyder TD, Dillow SA, Hoffman CM (2009) *Digest of Education Statistics 2008* (U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, Washington, DC), NCES Publ No 2009-020. Available at nces.ed.gov/pubs2009/2009020.pdf. Accessed September 11, 2014.
10. Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J (2012) Science faculty's subtle gender biases favor male students. *Proc Natl Acad Sci USA* 109(41):16474–16479.
11. Reuben E, Sapienza P, Zingales L (2014) How stereotypes impair women's careers in science. *Proc Natl Acad Sci USA* 111(12):4403–4408.
12. Knobloch-Westerwick S, Glynn CJ, Huge M (2013) The Maltida effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Sci Commun* 35(5):603–625.
13. Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR (2013) Bibliometrics: Global gender disparities in science. *Nature* 504(7479):211–213.
14. Schroeder J, et al. (2013) Fewer invited talks by women in evolutionary biology symposia. *J Evol Biol* 26(9):2063–2069.
15. Sheltzer JM, Smith JC (2014) Elite male faculty in the life sciences employ fewer women. *Proc Natl Acad Sci USA* 111(28):10107–10112.
16. Jaschik S (2014) Productivity or Sexism? *Inside Higher Education*. Available at https://www.insidehighered.com/news/2014/08/18/study-raises-questions-about-why-women-are-less-likely-men-earn-tenure-research. Accessed September 11, 2014.
17. Steinpreis RE, Anders KA, Ritzke D (1999) The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles* 41(7/8):509–528.
18. Devine PG (1989) Stereotypes and prejudice: Their automatic and controlled components. *J Pers Soc Psychol* 56(1):5–18.
19. Greenwald AG, Banaji MR (1995) Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychol Rev* 102(1):4–27.
20. Swim JK, Scott ED, Sechrist GB, Campbell B, Stangor C (2003) The role of intent and harm in judgments of prejudice and discrimination. *J Pers Soc Psychol* 84(5):944–959.
21. Uhlmann EL, Cohen GL (2007) "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organ Behav Hum Decis Process* 104(7):207–223.
22. Baron AS, Banaji MR (2006) The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychol Sci* 17(1):53–58.
23. Knutson KM, Mah L, Manly CF, Grafman J (2007) Neural correlates of automatic beliefs about gender and race. *Hum Brain Mapp* 28(10):915–930.
24. Becker JC, Swim JK (2011) Seeing the unseen: Attention to daily encounters with sexism as way to reduce sexist beliefs. *Psychol Women Q* 35(2):227–242.
25. Apfelbaum EP, Phillips KW, Richeson JA (2014) Rethinking the baseline in diversity research: Should we be explaining the effects of homogeneity? *Perspect Psychol Sci* 9(3):235–244.
26. Page SE (2007) Making the difference: Applying the logic of diversity. *Acad Manage Perspect* 21(4):6–20.
27. Freeman RB, Huang W (2014) Collaborating with People Like Me: Ethnic Co-Authorship Within the US. (National Bureau of Economic Research, Cambridge, MA), NBER Working Paper No 19905. Available at www.nber.org/papers/w19905. Accessed September 11, 2014.
28. National Science Board (2012) *Science and Engineering Indicators 2012* (National Center for Science and Engineering Statistics, Arlington, VA). Available at www.nsf.gov/statistics/seind12/. Accessed September 11, 2014.
29. Tabak LA, Collins FS (2011) Sociology. Weaving a richer tapestry in biomedical science. *Science* 333(6045):940–941.

30. National Institutes of Health (2014) Enhancing the Diversity of the NIH-Funded Workforce (National Institutes of Health, Bethesda). Available at commonfund.nih.gov/diversity/Initiatives. Accessed September 11, 2014.
31. Moss-Racusin CA, et al. (2014) Social science. Scientific diversity interventions. *Science* 343(6171):615–616.
32. US Department of Education (2014) *Science, Technology, Engineering and Math: Education for Global Leadership*. Available at www.ed.gov/stem. Accessed September 11, 2014.
33. Obama BH (2013) State of the Union Address. Available at https://www.whitehouse.gov/the-press-office/2013/02/12/remarks-president-state-union-address. Accessed September 11, 2014.
34. President's Council of Advisors on Science and Technology (2012) Engage to Excel: Producing One Million Additional College Graduates with degrees in Science, Technology, Engineering, and Mathematics (Executive Office of the President, Washington, DC). Available at https://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_feb.pdf. Accessed September 11, 2014.
35. Moss-Racusin CA, Molenda AK, Cramer CR (2015) Can evidence impact attitudes? Public reactions to evidence of gender bias in STEM fields. *Psychol Women Q* 39(2):194–209.
36. Tajfel H, Turner JC (1986) The social identity theory of intergroup behaviour. *Psychology of Intergroup Relations*, eds Worchel S, Austin WG (Nelson-Hall, Chicago), pp 7–24.
37. Ellemers N, Spears R, Doosje B (2002) Self and social identity. *Annu Rev Psychol* 53(1):161–186.
38. Schmitt MT, Branscombe NR (2001) The good, the bad, and the manly: Threats to one's prototypicality and evaluations of fellow in-group members. *J Exp Soc Psychol* 37(6):510–517.
39. Jost JT, Banaji MR, Nosek BA (2004) A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Polit Psychol* 25(6):881–919.
40. Ecklund EH, Lincoln AE, Tansey C (2012) Gender segregation in elite academic science. *Gend Soc* 26(5):693–717.
41. Festinger L (1957) *A Theory of Cognitive Dissonance* (Stanford Univ Press, Stanford, CA).
42. Kunda Z (1990) The case for motivated reasoning. *Psychol Bull* 108(3):480–498.
43. Curtis JW (2010) Faculty Salary Equity: Still a Gender Gap? *On Campus with Women* 39(1). Available at archive.aacu.org/ocww/volume39_1/feature.cfm?section=2. Accessed August 20, 2014.
44. Boyle PJ, Smith LK, Cooper NJ, Williams KS, O'Connor H (2015) Gender balance: Women are funded more fairly in social science. *Nature* 525(7568):181–183.
45. Wenneras C, Wold A (1997) Nepotism and sexism in peer-review. *Nature* 387(6631):341–343.
46. McIntosh P (1988) White Privilege: Unpacking the Invisible Knapsack. *Peace and Freedom*. Available at nationalseedproject.org/white-privilege-unpacking-the-invisible-knapsack. Accessed September 25, 2015.
47. Norton MI, Sommers SR (2011) Whites see racism as a zero-sum game that they are now losing. *Perspect Psychol Sci* 6(3):215–218.
48. Lord CG, Ross L, Lepper MR (1979) Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *J Pers Soc Psychol* 37(11):2098–2109.
49. Mahoney MJ (1977) Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognit Ther Res* 1(2):161–175.
50. Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences* (Erlbaum, Hillsdale, NJ).
51. American Association of University Women (2010) Why So Few? Women in Science, Technology, Engineering, and Mathematics (American Association of University Women, Washington, DC). Available at www.aauw.org/resource/why-so-few-women-in-science-technology-engineering-mathematics/. Accessed September 11, 2014.
52. Ginther DK, Kahn S (2006) Does Science Promote Women? Evidence from Academia 1973-2001 (National Bureau of Economic Research, Cambridge, MA), NBER Working Paper No 12691. Available at www.nber.org/papers/w12691. Accessed September 11, 2014.
53. National Science Foundation (2010) *Women, Minorities, and Persons with Disabilities in Science and Engineering* (National Center for Science and Engineering Statistics, Arlington, VA). Available at www.nsf.gov/statistics/2015/nsf15311/. Accessed September 11, 2014.
54. Brewer MB (1999) The psychology of prejudice: Ingroup love and outgroup hate. *J Soc Issues* 55(3):429–444.
55. Shields SA, Zawadzki MJ, Johnson RN (2011) The impact of a workshop activity for gender equity simulation in the academy (WAGES-Academic) in demonstrating cumulative effects of gender bias. *J Divers High Educ* 4(2):120–129.
56. Zawadzki MJ, Danube CL, Shields SA (2012) How to talk about gender inequity in the workplace: Using WAGES as an experiential learning tool to reduce reactance and promote self-efficacy. *Sex Roles* 67(11-12):605–616.
57. Hong L, Page SE (2004) Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc Natl Acad Sci USA* 101(46):16385–16389.
58. Kasof J (1993) Sex bias in the naming of stimulus persons. *Psychol Bull* 113(1):140–163.
59. Mackie DM, Worth LT, Asuncion AG (1990) Processing of persuasive in-group messages. *J Pers Soc Psychol* 58(5):812–822.
60. Maitner AT, Mackie DM, Claypool HM, Crisp RJ (2010) Identity salience moderates processing of group-relevant information. *J Exp Soc Psychol* 46(2):441–444.

Supporting Information

Handley et al. 10.1073/pnas.1510649112

SI Materials and Methods

Participants and Recruitment for Experiments 1 and 3. Participation was elicited from workers on Amazon's Mechanical Turk online job site, who could view our employment opportunity (titled "What do REAL people think about science research results?") listed alongside other opportunities.

A total of 205 individuals opted to participate in experiment 1 and provided usable data, which was active in March 2014. Originally, 218 individuals participated in the experiment, but 9 were excluded from data analysis because they failed one or more attention-check items (e.g., "If you are reading, respond 'very much' to this question;" "If you are reading, respond 'not at all' to this question"), 2 because they reported being under 18 y of age, and 2 because they did not specify their gender. Ultimately, 146 men and 59 women from the United States who were 18 y of age or older ($M = 30.13$; range = 18–66) were retained for analysis. Of this general sample, 68.12% reported their race as "white," and 51 individuals reported they were currently college students.

A total of 303 individuals opted to participate in experiment 3 and provided usable data, which was active in November 2014. Originally, 321 individuals participated in the experiment, but 12 were excluded from data analysis because they failed one or more attention-check items, 2 because they reported being under 18 y of age or did not specify an age, 1 because they did not specify their gender, and 7 because they reported they had read the abstract before (some participants met multiple exclusion criteria). Ultimately, 162 men and 141 women from the United States who were 18 y of age or older ($M = 34.22$; range = 18–79),were retained for analysis. Of this general sample, 73.93% reported their race as "white," and 55 individuals reported they were currently college students.

Participants and Recruitment for Experiment 2. Participation was initially elicited on November 4, 2013, from all 506 tenure-track faculty at a research-intensive university via an email from their university provost encouraging participation in a larger faculty climate survey. That same day, our research team emailed all tenure-track faculty a message that explained the nature and importance of the survey, contained an informed consent form for faculty to read, explained the compensation faculty would receive for their participation, and contained a link to the survey and experiment, which was hosted on surveymonkey.com. This email included a unique identification code for each person, which preserved respondents' anonymity and confidentiality, but allowed us to trace the faculty's home department. In this way, we could determine whether faculty resided in STEM, including Social and Behavioral Sciences, departments (i.e., Agricultural Economics and Economics, Animal and Range Sciences, Cell Biology and Neuroscience, Center for Biofilm Engineering, Chemical and Biological Engineering, Chemistry, Civil Engineering, Computer Science, Earth Science, Ecology, Electrical Engineering, Industrial and Management Engineering, Institute on Ecosystems, Immunology and Infectious Diseases, Land Resources and Environmental Science, Mathematical Sciences, Mechanical and Industrial Engineering, Microbiology, Native American Studies, Physics, Plant Sciences, Political Science, Psychology, and Sociology and Anthropology) or non-STEM departments (i.e., Agricultural Education, Art, Nursing, Education, English, Film and Photography, Health and Human Development, History and Philosophy, Honor's Program, Liberal Studies, Modern Languages and Literature, Science Education, Music, and University Studies). All faculty who did not participate as of November 18 received a reminder email, which also contained a link to the survey and experiment and their unique identification code. The survey was closed on the evening of November 22.

Ultimately, 286 faculty participated in the unrelated survey, and 205 (40.5% of faculty) further elected to participate in our experiment at the end of the survey. Of these, 111 (54%) were men and 94 were women. Further, as specified above, 116 faculty were categorized as residing in STEM departments and 89 as residing in non-STEM departments. A comparable ratio of faculty from STEM (116/289 or 40.1%) and non-STEM (89/217 or 41.0%) departments completed the experiment. Participants indicated their race as white/Caucasian (86.3%), Asian (2%), Hispanic/Latino (1%), Native American (0.5%), or mixed (0.5%), or they opted not to report these data (9.8%). Further, participants reported their faculty rank as assistant (43.9%), associate (27.8%), or full (26.3%), or they did not specify (2%). Participants' ages ranged from 27 to 73 y ($M = 47.35$), and they had worked in their current position between 0 and 35 y ($M = 10.51$). The demographics of our sample closely match the population of professors from this university (which is 64% male and 90.9% white/Caucasian), although assistant professors were somewhat overrepresented in our sample relative to the university population (assistant, associate, and full ranks comprise 29.3%, 32.1%, and 38.6% of professors, respectively). Aside from rank, perhaps, we can reasonably infer that there were no systematic biases influencing individuals' decisions to participate in the experiment. That is, the results from this sample likely generalize to the population of faculty under investigation.

Procedure for Experiment 1. For experiment 1, once participants clicked on the title for our experiment on Amazon's Mechanical Turk, they encountered the following short paragraph: "In the scientific world, peer experts judge the quality of research and decide whether or not to publish it, fund it, or discard it. But what do everyday people think about these articles that get published? We are conducting an academic survey about people's opinions about different types of research that was published back in the last few years. You will be asked to read a very brief research summary and then answer a few questions about your judgments as non-experts about this research. There is no right or wrong answer and we realize you don't have all the information or background. But just like in the scientific world, many judgments are made on whether something is quality science or not after just reading a short abstract summary. So to create that experience for you, we ask that you just provide your overall reaction as best you can even with the limited information. You will also be asked to provide demographic information about yourself. Select the link below to complete the survey." Participants were also reminded that they would receive $0.25 in exchange for submitting the job "hit." Participants then accepted the hit and opened up the survey in a separate tab or window. After consenting to participate, participants were given a summary of the experiment that they read before accepting the hit and then were asked, "Please read the following abstract from a 2012 published research study then provide your opinion with the items below." Next, participants viewed the abstract written by Moss-Racusin et al. (10), the first author's name and affiliation, and keywords, as described in the main text, and participants then provided their opinions about the abstract using scale ratings (*SI Materials and Methods*, *Dependent Variables*). Once they began the survey, participants learned that they could skip over any questions or task that they wished, ensuring that our procedures were not coercive. Participants then completed demographic information, were debriefed regarding the purpose of the experiment, and were compensated \$0.25 for their time.

Procedure for Experiment 2. For experiment 2, once participants followed the link to the survey website, they first read information about the faculty climate survey and the types of tasks and questions they would encounter. Participants were also reminded that they would receive a \$5 coupon from a local coffee shop for completing the survey and would be entered into a raffle to win 1 of 50 gift certificates form the campus bookstore (worth \$50). Once they advanced to the survey, participants further learned that they could skip over any question or task they wished. This option resulted in several participants providing only partial data for the experiment (addressed in *SI Additional Analyses*, *Experiment 2*). The faculty climate survey took ~15 min to complete and primarily contained questions about the university work environment, which were independent from the reported experiment.

Just after the survey, participants were asked to "Please read the following abstract from a 2012 published research study then provide your opinion with the items below." They then viewed the same abstract and associated information as in experiment 1 and evaluated that abstract using the same scale ratings. Finally, participants entered their unique code and could print off a coupon in compensation for their participation.

Procedure for Experiment 3. The procedures for experiment 3 were identical to experiment 1, with a few minor, but important, differences. First, participants were randomly assigned to read either the original version of the Knobloch-Westerwick et al. (12) abstract, which reported a gender bias (e.g., "Publications from male authors were associated with greater scientific quality, in particular if the topic was male-typed") or a version slightly altered to report no gender differences (e.g., "Publications from male and female authors were associated with comparable scientific quality, even if the topic was male-typed"). Second, unlike in experiments 1 and 2, the abstract was not accompanied by the author's name or affiliation. Otherwise, the procedures and dependent measures for this experiment were identical to those used in the previous experiments. At the end of the experiment, participants completed demographic information, were debriefed regarding the purpose of the experiment, and were compensated \$0.25 for their time.

Dependent Variables. After reading the abstract, participants in all experiments reported their evaluation of the abstract and research using measures adapted from those commonly used to gauge attitude change and evaluations of persuasive materials (59, 60). Specifically, on scales from 1 (*not at all*) to 6 (*very much*), participants responded to the following four questions or statements: "To what extent do you agree with the interpretation of the research results?" "To what extent are the findings of this research important?" "To what extent was the abstract well written?" and "Overall, my evaluation of this abstract is favorable." These four responses demonstrated high internal consistency in all experiments (Cronbach's $\alpha = 0.84$, 0.89, and 0.78 in experiments 1, 2, and 3, respectively) and were therefore averaged to measure participants' perceived quality of the research.

For experiment 2 only, participants completed a faculty climate survey before the experiment, which included items assessing the extent to which faculty felt that they had been personally discriminated against due to their gender. Specifically, on scales from 1 (*strongly disagree*) to 7 (*strongly agree*), participants responded to the following three statements: "I have personally been a victim of gender discrimination," "I consider myself a person who has been deprived of opportunities because of my gender," and "Prejudice against my gender group has not affected me personally" (the latter of which was reverse-scored). These three responses demonstrated high internal consistency (Cronbach's $\alpha = 0.87$) and were therefore averaged to measure participants' personal experience of gender discrimination.

SI Additional Analyses

Experiment 1. For the primary measure, author gender and affiliation alone did not influence evaluations, and neither did any two-way interactions among factors (all $P > 0.3$). However, the analysis revealed a nonpredicted and significant interaction among participant gender, author gender, and author affiliation [$F(1, 197) = 18.13$; $P < 0.001$]. Consistent with the theme of this work, we describe this interaction in terms of gender differences at each combination of author gender and affiliation. When the abstract author was supposedly a man from Iowa State University, male participants rated the abstract as being of higher quality ($M = 4.57$, SD = 0.787) than did women ($M = 4.26$, SD = 0.893), whereas when the abstract author was supposedly a woman from Iowa State University, female participants rated the abstract as being of higher quality ($M = 5.03$, SD = 0.713) than did men ($M = 3.89$, SD = 1.13). Thus, when the author was supposedly affiliated with Iowa State University, all participants seemed to demonstrate a gender bias in favor of their own gender; women had higher ratings for a female author, and men gave higher ratings for a male author. However, when the abstract author was supposedly a man from Yale University, female participants instead rated the abstract as being of higher quality ($M = 5.02$, SD = 0.874) than did men ($M = 4.13$, SD = 0.897), whereas when the abstract author was supposedly a woman from Yale University, female participants reported ratings of the abstract ($M = 4.38$, $SD = 1.031$) that were equivalent to those of men ($M = 4.38$, $SD = 0.697$). Interestingly, when evaluating research from Yale that reveals gender bias, it seems that women demonstrated the greatest bias against women (or favoring men) authors.

There are at least two important notes regarding this interaction between participant gender, author gender, and author affiliation. First, this interaction was not observed in the second experiment among university faculty. Thus, although this interaction is certainly interesting, we withhold focusing too much on this result until it is replicated in future research. This result was not predicted or replicated and may be spurious. Second, if this interaction pattern does replicate in future research, this finding may indicate that the lay public and scientific community manifest bias toward research uncovering gender bias differently under different conditions. Within scientific communities, perhaps the gender bias against such research is unaffected by author gender or affiliation. However, in the lay public, the gender bias is more complex and context-dependent. Ultimately, it is important to understand failures in objectivity among the scientific community, as well as the public, regarding research demonstrating gender bias in STEM. After all, it is often the nonscientists (the public, government officials, bureaucrats, nonprofit organizations, special-interest groups, etc.) that drive the funding opportunities so critical to scientific progress and discovery.

Experiment 2. In addition to the predicted effects reported in the paper, the primary analysis also revealed a significant interaction among field of study, author gender, and author affiliation [$F(1, 174) = 8.07$; $P < 0.01$]. The interaction pattern indicated that faculty in STEM evaluated the abstract written by a man more favorably if the author was from Yale (vs. Iowa State), but the abstract written by a woman more favorably if the author was from Iowa State (vs. Yale), whereas the opposite pattern manifested among non-STEM faculty.

Additionally, we conducted the analysis again, removing fields of study associated with the social and behavioral sciences (i.e., Agricultural Economics and Economics, Native American Studies, Political Science, Psychology, and Sociology and Anthropology) from the analysis entirely. Given that the classification of some of these fields as STEM might vary depending on who one consults, we wanted to confirm that the key results held comparing STEM to non-STEM fields, even excluding the social and behavioral sciences. Indeed, this analysis, too, revealed the predicted significant main effect of gender [$F(1, 156) = 8.30$, $P = 0.005$] and the predicted significant interaction between gender and field of study [$F(1, 156) = 7.31$, $P = 0.008$].

Further, given that there was a somewhat disproportionate representation of assistant professors in our sample, we investigated whether our results held accounting for faculty rank. To do this analysis, we collapsed across the author's gender and affiliation (including all factors created several conditions with only one participant's response) and conducted an analysis with faculty gender, field of study, and faculty rank as factors (four participants did not report their rank and were therefore not included in this analysis). Like the primary analysis, this analysis revealed a significant main effect of gender [$F(1, 174) = 6.04$; $P = 0.015$] and a significant interaction between gender and field of study [$F(1, 174) = 5.27$; $P = 0.023$]. Therefore, the original results hold while controlling for faculty rank. No other main effects or interactions reached significance (all other $F < 2.43$; $P > 0.09$).

Of note, several participants in experiment 2 elected to skip some of our four measures. Of the full 205 participants, 190 completed all four measures—which were averaged for the primary analyses. Thus, we examined how well our predicted findings held examining each measure independently. Critically, there was a significant main effect of participant gender for three of the four measures. Relative to female faculty, male faculty agreed less with the interpretations of the research [$n = 199$, $F(1, 183) = 6.66$, $P = 0.011$], evaluated the research findings as less important [$n = 202$, $F(1, 186) = 7.00$, $P = 0.009$], evaluated the abstract as less well written [$n = 196$, $F(1, 181) = 4.67$, $P = 0.032$], and overall evaluated the abstract less favorably [$n = 201$, $F(1, 185) = 3.45$, $P = 0.065$)].

Additionally, the pattern of means for the interaction between participant gender and their STEM status for each of these measures was identical to that observed for the primary analysis. However, the omnibus test of this interaction was significant for participants' ratings of how important they evaluated the research findings [$F(1, 186) = 8.31$, $P = 0.004$], how well written they found the abstract [$F(1, 181) = 4.22$, $P = 0.041$], and their overall favorability toward the abstract [$F(1, 185) = 9.80$, $P = 0.002$], but not for their assessment of how much they agreed with the interpretations of the research [$F(1, 183) = 1.55$, $P = 0.21$]. Nonetheless, as in the primary analysis, simple-effect tests for all measures revealed that male faculty reported less favorable evaluations than female faculty in STEM departments (all $F > 7.91$ and < 17.14; all $P < 0.005$), but comparable evaluations within non-STEM departments (all $F < 1$). Overall, then, the critical findings for the primary measure hold well when looking at each individual measure.

Finally, although we did not design experiment 2 to specifically investigate potential mechanisms behind these effects, especially regarding the interaction, some data within a faculty survey (completed just before our experiment) allowed us to explore the possibility that these effects were related to perceptions of threat. Specifically, faculty rated the extent to which they felt they had been personally discriminated against due to their gender (*SI Materials and Methods*, *Dependent Variables*). We reasoned that the greater men's experience of gender discrimination (the more they feel women have had an unjust advantage at men's expense), the more threatening they should find research demonstrating an actual bias against women in STEM. After all, men who have experienced gender discrimination may harbor concern that such research could promote future "reverse" discrimination against men in STEM. Further, assuming men in STEM are more committed to (or identify with) STEM than men in non-STEM fields, Social Identity Theory (36, 37) predicts that the experience of threat should predominantly manifest among men in STEM. Indeed, there was a negative correlation between the personal experience of gender discrimination and evaluations of the abstract only among men in STEM. The more male faculty in STEM felt they experienced gender discrimination, the less favorably they evaluated the abstract [$r(63) = -0.404$; $P = 0.001$]. This same correlation among non-STEM men was positive but nonsignificant, [$r(34) = 0.157$; $P = 0.367$]. Among women, results yielded a significant correlation within non-STEM fields [$r\ (48) = 0.35$; $P = 0.014$], but no correlation within STEM fields [$r\ (36) = 0.262$; $P = 0.118$]. However, these correlations would not indicate anything about threat because the results of Moss-Racusin et al. (10) affirm women's experience with gender discrimination.

Together, these two correlations among men in STEM and non-STEM are consistent with Social Identity Theory and our assumption that men in STEM identify more with STEM than do non-STEM men and likely perceived the abstract as more threatening. However, the gender-discrimination measure did not mediate the effects found for the abstract evaluation. To test for possible effects, we subjected the gender-discrimination measure to an analysis of variance with gender and field of study as factors (participants completed this measure before reading the abstract, making the factors associated with the abstract inconsequential). Importantly, this analysis revealed a significant main effect of gender such that women experienced greater gender discrimination than men [$F(1, 194) = 16.87$; $P < 0.001$], indicating that the construct was valid. However, this analysis revealed no interaction [$F(1, 194) = 1.77$; $P > 0.18$], meaning this construct did not mediate our primary results. This finding is not necessarily surprising, however, given that the gender-discrimination measure was not designed to directly measure the extent to which participants find the results of Moss-Racusin et al. (10) to be threatening. Overall, then, the correlation evidence is only suggestive, and we encourage future research to explore this and other possible mechanisms behind our effect.