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

Ian M. Handleya,1, Elizabeth R. Browna,b, Corinne A. Moss-Racusinc, and Jessi L. Smitha

aDepartment of Psychology, Montana State University, Bozeman, MT 59717-3440; bDepartment of Psychology, University of North Florida, Jacksonville, FL 32224; and cDepartment 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 biases 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 members 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.

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.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1To whom correspondence should be addressed. Email: handley@montana.edu.

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

www.pnas.org/cgi/doi/10.1073/pnas.1510649112

PNAS | October 27, 2015 | vol. 112 | no. 43 | 13201–13206

PNAS | October 27, 2015 | vol. 112 | no. 43 | 13201–13206
are more committed to STEM fields and men within non-STEM fields (e.g., English professors), assuming they men within STEM fields (e.g., physics professors) may feel more perceive greater threat toward their group (and defend against that exposes gender bias in STEM as more threatening than 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 threatened 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.

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) predict that men may more negatively evaluate research demonstrating gender bias against women in STEM (vs. non-STEM) fields, to whom a gender bias in STEM, but 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 men would evaluate research demonstrating no gender bias (belief inconsistent) more favorably than 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 first experiment supported hypothesis A, revealing a main effect of participant gender [$F(1, 197) = 9.85, P = 0.002, \eta^2_{\text{partial}} = 0.048$], 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 second experiment also supported hypothesis A, revealing a main effect of participant gender [$F(1, 174) = 6.08, P = 0.015, \eta^2_{\text{partial}} = 0.034$], 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 [$F(1, 174) = 5.19, P = 0.024, \eta^2_{\text{partial}} = 0.03$]. 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 [$F(1, 299) = 4.00, P = 0.046, \eta^2_{\text{partial}} = 0.013$] (all other $F < 1$). Although no simple-effect tests were significant (all $P > 0.10$), together, these results support the overall pattern predicted by hypothesis C, such 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 that 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 non-existent 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), whereas the reported 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, 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 focusses 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 research biases through a common representation of the process is biased (21, 42). Theoretically, however, there is reason to predict that gender biases in STEM, we suggest that is important for those in STEM to acknowledge 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 post-doctoral 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 their decision making (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-Westrick 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 woman. 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 peeks 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 were 18 y of age or older (M = 34.22; range = 18–78) 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 = 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 2, 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 = 18–78) opted to participate in the experiment and provided usable data. All participants engaged in the ~10-min experiment in exchange for $0.25.

In the experiment, invitation was randomized 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 $500 US 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 participant gender and affiliation. 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-Westervelt 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-Westervelt 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.).


