Sex Roles
https://doi.org/10.1007/s11199-018-0902-z

ORIGINAL ARTICLE

# Gender Bias Produces Gender Gaps in STEM Engagement

Corinne A. Moss-Racusin[1] · Christina Sanzari[1] · Nava Caluori[1] · Helena Rabasco[1]



**Abstract**
We explored whether the existence of gender bias causes gender gaps in STEM engagement. In Experiment 1 ($n = 322$), U.S. women projected less sense of belonging, positivity toward, and aspirations to participate in STEM than did men when exposed to the reality of STEM gender bias. These gender differences disappeared when participants were told that STEM exhibits gender equality, suggesting that gender bias produces STEM gender gaps. Experiment 2 ($n = 429$) explored whether results generalized to a specific STEM department, and whether organizational efforts to mitigate gender bias might shrink gender gaps. U.S. women exposed to a biased chemistry department anticipated more discrimination and projected less sense of belonging, positive attitudes and trust and comfort than did men. These gender differences vanished when participants read about an unbiased department, again suggesting that gender bias promotes STEM gender gaps. Further, moderated mediation analyses suggested that in the presence of gender bias (but not gender equality), women projected less positive attitudes and trust and comfort than did men *because* they experienced less sense of belonging and anticipated more discrimination. Results were largely unaffected by whether departments completed a diversity training, suggesting that knowledge of diversity initiatives alone cannot close STEM gender gaps.

**Keywords** Gender bias · STEM · Gender gap · Sexism · Diversity training

Most scholars agree that there is a gender disparity across science, technology, engineering, and mathematics (STEM) fields. For example, recent estimates indicate that women comprise only 28.4% of the science and engineering workforce in the United States (National Science Foundation 2017). However, there is ongoing debate about the causal factors responsible for the STEM gender gap. Researchers have identifying numerous contributing variables, including but not limited to, differences in men's and women's lifestyle "choices" (whether free or constrained; Ceci and Williams 2011; Williams and Ceci 2012), endorsement of communal goals (which are often viewed as incompatible with STEM work; Diekman et al. 2017), access to appropriate role models and mentors (Stout et al. 2011), social identity threat (Murphy et al. 2007), stereotypes casting STEM as male gender-typed (Nosek et al. 2009), and masculine cultures that undermine women's sense of belonging (Cheryan et al. 2009).

However, the potential causal impact of gender bias itself on the STEM gender disparity remains largely unexplored. Clarifying the role of gender bias is critical; if gender bias has no impact on men's and women's STEM outcomes, then current policies aimed at stamping out bias reflect wasted resources and missed opportunities (Ceci et al. 2014; Ceci and Williams 2011). However, if gender bias does contribute to women's underrepresentation, then ignoring its causal role may allow current gender gaps to linger or expand. Thus, in the current research, we sought to provide the first (to our knowledge) experimental test of the causal role of field and department-level gender bias in producing U.S. gender gaps in STEM engagement, and we also explored the extent to which institutional diversity efforts aimed at mitigating bias may close these gaps.

## Consequences of STEM Gender Bias

A growing body of evidence reveals gender biases favoring men across STEM fields. For example, women in STEM

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s11199-018-0902-z) contains supplementary material, which is available to authorized users.

✉ Corinne A. Moss-Racusin
cmossrac@skidmore.edu

[1] Department of Psychology, Skidmore College, 815 North Broadway, Saratoga Springs, NY 12866, USA

Published online: 12 February 2018

report regularly encountering gender bias (Robnett 2016; Steele et al. 2002; Williams et al. 2016), papers by women are less likely to be cited than those by men (Lariviere et al. 2013), female junior biomedical researchers receive significantly less start-up support than do comparable men (Sege et al. 2015), women are less likely than men to be invited to present colloquium talks at top universities (across six academic fields; Nittrouer et al. 2017), and when men introduce invited medical grand rounds speakers, they are less likely to refer to female (relative to male) colleagues by their professional title (Files et al. 2017). Experimental evidence echoes these correlational patterns, revealing that relative to the identical men, women are less likely to be hired for a psychology faculty position (Steinpreis et al. 1999), a job in mathematics (Reuben et al. 2014), and a lab manager position (Moss-Racusin et al. 2012). They are also judged to be less competent, less deserving of mentoring, are paid less (Moss-Racusin et al. 2012), and are less likely to receive valuable pre-doctoral mentoring (Milkman et al. 2012, 2015). Finally, a scientific abstract is judged as poorer in quality when attributed to a female (vs. male) author (Knobloch-Westerwick et al. 2013). Participant gender differences did not emerge across these experiments, suggesting that both women and men are likely to exhibit STEM gender biases.

Although a large body of evidence reveals gender biases favoring men, it is important to note that one recent paper found the opposite pattern. Results from a series of experiments revealed a preference for women tenure-track STEM candidates on the part of STEM faculty compared to similarly qualified men (Williams and Ceci 2015). These results may signal that there are some situations in which women benefit from gender bias. However, certain methodological aspects of these experiments may constrain their generalizability (e.g., Williams and Smith 2015). For example, target applicants were irrefutably excellent, a condition well-known to suppress typical gender bias effects that emerge in more realistic, nuanced, and common situations (Heilman et al. 2004). Additionally, this research focused only on faculty hiring, whereas existing evidence suggests that women encounter substantial bias at other career junctures. Finally, given extensive media coverage of prior experimental evidence of gender bias in STEM (e.g., Pollack 2013) and discussion of these results in the STEM community, Williams and Ceci's (2015) faculty participants may have been aware of the true purpose of their experiment, experienced social desirability pressures, and "bent over backwards" to avoid favoring male candidates.

This concern is further bolstered by the fact that virtually no prior research has revealed a preference for female candidates in male-typed domains. Exceptions to this rule are very few; to our knowledge, they include only a preference for female over male leaders under severe system threats inspiring a need for change (Brown et al. 2011) and when a company was said to be on the brink of failure (i.e., "the glass cliff"; Bruckmuller and Branscombe 2010). In other words, the existing research suggests that women are preferentially selected only when situations are particularly dire and when women are viewed as offering the potential for change, suggesting that evidence of gender biases favoring women is particularly limited. Thus, although future research is needed to further explore the nature and limits of STEM gender bias, the majority of existing evidence suggests that gender biases favoring men remain an unfortunate problem.

However, evidence of bias does not necessarily indicate that it *causes* women's underrepresentation. Indeed, although prior research has sought to determine whether STEM gender bias exists, whether the existence of gender bias contributes to the STEM gender gap is its own empirical question. That is, skeptics may argue that women simply persist in the face of gender biases in evaluation, hiring, and pay, such that these experimentally-demonstrated biases ultimately may not be causally responsible for women's underrepresentation. This possibility remains because researchers have yet to experimentally manipulate the alleged existence of STEM gender bias and examine its causal impact on women's and men's STEM engagement. In a review of the potential causes of women's underrepresentation in math-intensive fields, Ceci and colleagues (2014, p. 26) express this idea clearly:

> One might imagine that, given the plethora of allegations, there would be compelling evidence that biased interviewing and hiring is a cause of women's underrepresentation in STEM fields and/or that discriminatory remuneration and promotion practices are responsible for the gender gap in pay and rank. However, the evidence in support of biased hiring *as a cause of underrepresentation* is not well-supported…We do not claim that there have not been many excellent demonstrations of implicit bias or stereotyping and explicit bias; rather, our claim is that the literature has failed to demonstrate a causal link between such demonstrations and the underrepresentation of female faculty. (italics in the original)

Simply put, does the existence of gender bias contribute to women's underrepresentation in STEM? Alternately, it is possible that women are unaffected by STEM gender bias, perhaps because they manage to ignore or dismiss it, or even utilize it as motivation to work harder. Although we argue in the following that the former possibility is more likely, answering this question is of the utmost importance for shrinking the STEM gender gap. Overstating the impact of gender bias if it does not contribute to the gender gap (i.e., committing a Type I error) could lead to interventions that are by definition ineffective because they target the wrong culprit (Ceci and Williams 2011; Ceci et al. 2014). Conversely, dismissing the role of gender bias if it does contribute to the gender gap (i.e., committing a Type II error) could distract from needed policy changes, trivialize women's well-documented subjective experiences of bias (Robnett 2016; Steele et al. 2002; Williams et al. 2016), and perhaps leave room for more biased behaviors, which could widen the current gender gap. Of importance, if evidence emerges to support the causal impact of gender bias on STEM gender gaps, we do not claim that is the sole contributing factor. As we stated, numerous other contributors have been identified by other researchers; a systematic review of this literature is beyond the scope of the current research and has been provided elsewhere (e.g., Cheryan et al. 2017; Ceci et al. 2014). However, a clear experimental test is needed to determine whether gender bias should be added to the list of factors responsible for the STEM gender gap.

Related research suggests that gender bias likely does contribute to women's underrepresentation in STEM. Notably, work on stereotype threat highlights the detrimental effects of knowing that one's own behavior could confirm a negative stereotype about one's group (Steele and Aronson 1995). For example, when a math test was described as producing gender differences favoring men, women scored lower than similarly-qualified men; these differences vanished when the test was described as not producing gender differences (Spencer et al. 1999; Walsh et al. 1999). Moreover, social identity threats can undermine women's interest in math and science (Murphy et al. 2007; Smith et al. 2007). Broadly, the stereotype threat literature highlights the role of negative group stereotypes in undermining outcomes such as stigmatized group members' sense of belonging, motivation, and performance (Inzlicht and Schmader 2012). By extension, this raises the possibility that knowledge of the existence of gender bias may be similarly detrimental. However, although the stereotype threat literature is suggestive, the potential impact of the existence of gender bias itself (rather than the threat of fulfilling negative group stereotypes) on gender gaps in STEM engagement remains largely unexplored.

More closely related to our central research question, several studies have demonstrated that interacting with one sexist individual can undermine women's performance. For example, male interviewers' implicit gender stereotypes negatively predicted female applicants' interview performance as assessed by third-party raters (Latu et al. 2015), and male engineering students' levels of subtle sexism negatively predicted the performance of their female engineering student partners' performance on an engineering test (Logel et al. 2009). Further, Adams et al. (2006) explored the impact of the mere suggestion that one person in a position of authority could be sexist. When told that an instructor in a laboratory experiment might be sexist, women (but not men) reported a less positive instructional experience, demonstrated worse performance on a logic test, and rated the instructor himself as less competent relative to women who were not told that the instructor might be sexist. Similarly, Thoman and Sanson (2016) investigated the impact of receiving gender-biased feedback from an experimenter on students' interest in science activities completed during a laboratory experiment. Women who were given no reason for the fact that they were not chosen for a leadership role in the experiment reported being more interested in a science-related task they had just completed than did women who were told that they were not chosen because "men usually do better than women at science."

These studies suggest that one person's gender bias can have negative consequences for women, but they did not explore the more systemic impact of knowledge that an entire field (or department) typically exhibits bias. Additionally, the existing studies focused on outcomes acutely relevant to the particular instructional situation (e.g., task performance, individual instructor competence, task interest) whereas the current research focuses on consequences for general STEM engagement (i.e., positive attitudes, sense of belonging, aspirations, trust and comfort). More broadly, existing results suggest (but do not demonstrate) that field- or department-level gender bias may indeed be linked to gender gaps in engagement and that women do not simply ignore the existence of gender bias when making decisions about what field to enter (or exit, as the case may be). However, a direct test of this causal relationship—as well as an exploration of factors that may ameliorate negative effects of the existence of bias on women's STEM engagement—remained forthcoming.

## Downstream Impact of Diversity Training

If the existence of gender bias does indeed produce gender gaps in STEM engagement, then it is crucial to identify strategies that successfully mitigate bias. Diversity training reflects one intervention approach that has become an increasingly common organizational practice (Paluck 2006), in part due to concerns about the professional consequences associated with bias (Green and Kalev 2008). Although a comprehensive review of the diversity training literature is beyond the scope of the current research, we note that although a growing number of programs have been shown to boost women's representation in the workforce (e.g., Devine et al. 2017; Smith et al. 2015), very few theoretically-grounded interventions capable of demonstrably reducing STEM gender bias have historically been available (Moss-Racusin et al. 2014). Of importance, successful efforts to increase women's STEM participation cannot stop at the hiring stage. Although it is encouraging that available interventions can boost hiring rates of women in STEM, retention rates may remain low without more comprehensive programs capable of successfully targeting pernicious STEM gender bias itself. Hearteningly, recent evidence reveals that some newer, evidence-based programs do successfully reduce gender bias in STEM (e.g., Carnes et al. 2015; Moss-Racusin et al. 2018; Pietri et al. 2017), suggesting that rigorous diversity trainings may provide a fruitful avenue for reducing STEM gender bias and improving gender parity in the future.

However, the bulk of existing research on diversity trainings has focused on their ability to improve bias-related outcomes for attendees. That is, success has been conceptualized largely in terms of whether *individual attendees* demonstrate more positive outcomes (e.g., less bias, more awareness of diversity issues, more hiring of stigmatized group members, etc.) after the training than before it, and sometimes relative to a control group (Paluck 2006; Paluck and Green 2009). In contrast, very little work has considered the impact of diversity trainings from a broader perspective by assessing their downstream consequences for *non-attendees* (i.e., other members of stigmatized groups within the organization). We thus explored whether knowing that a successful institutional diversity training has taken place might ameliorate the negative effects of the existence of gender bias for women in STEM. We reasoned that if women's underrepresentation indeed results from gender bias (among other factors) as predicted, then knowledge of organizational diversity trainings designed to ameliorate gender bias could equalize men's and women's STEM engagement. If so, this would highlight dual potential benefits of diversity trainings—first for attendees themselves and then for stigmatized group members within the broader community.

Some related work on corporate diversity messages suggests that under the right circumstances, knowledge of diversity and inclusion efforts might positively impact stigmatized group members' experiences in organizations. For example, when African American professionals read about a company with a low proportion of Employees of Color, they reported more trust and comfort when the company espoused a progressive value-diversity philosophy relative to a colorblind diversity philosophy (Purdie-Vaughns et al. 2008). Additionally, Women of Color expected more organizational diversity, had greater performance expectations, and actually performed better on a math test after exposure to a multicultural (relative to a colorblind) corporate diversity message (Wilton et al. 2015). Moreover, the presence of organizational diversity structures targeting one stigmatized group (e.g., People of Color) resulted in identity safety for members of other stigmatized groups as well (e.g., White women; Chaney et al. 2016). Further, the mere presence of diversity trainings led women to view organizations as procedurally fair (Brady et al. 2015).

However, other work suggests that the existence of diversity trainings may not be a magic bullet. Indeed, the mere presence of diversity structures may create an "illusion of fairness" that paradoxically undermines organizational equity (Kaiser et al. 2013). For example, although Brady et al. (2015) determined that the existence of diversity trainings boosts women's perceptions of fairness, this ironically also led them to view sexist hiring outcomes as reasonable and express less support for litigation alleging sexism. Additionally, people may be aware that diversity trainings are often ineffective or counterproductive (Anand and Winters 2008; Dobbin et al. 2015; Legault et al. 2011). Thus, women in STEM may not believe that diversity trainings will effectively eradicate bias, and thus the gender gap may still persist. Given the paucity and contradictory nature of existing evidence on this topic, it was unclear whether knowledge of a diversity training might undo the negative effects of gender bias and close the STEM gender gap. Rather than make an a-priori prediction, we provided the first known direct test of this possibility.

## The Current Research

The current work sought to help resolve debates about the role of gender bias in contributing to women's underrepresentation in STEM. To do so, we explored the extent to which knowledge that entire academic fields (Experiment 1) or specific departments (Experiment 2) exhibit gender bias favoring men can produce gender disparities in engagement with those fields/departments. To do so, we manipulated the alleged existence of gender bias, and then we assessed various manifestations of men's and women's engagement. This design allowed us to isolate the causal impact of the existence of gender bias on the gender gap in STEM.

Although past research has often focused on predictors of STEM performance (Logel et al. 2009; Spencer et al. 1999; Steele and Aronson 1995; Walsh et al. 1999), in the current work, we were interested in assessing women's and men's STEM engagement more broadly. Although performance is certainly relevant to understanding women's underrepresentation, a host of studies have suggested that women's STEM underrepresentation is not due to performance detriments (Ceci et al. 2014; Halpern et al. 2007; Hyde and Linn 2006; Spelke 2005). Thus, we focused on engagement (relying on variables drawn from the existing literature) rather than on performance. Specifically, in Experiment 1, we assessed the extent to which participants projected positive attitudes toward STEM, sense of belonging in STEM, and aspirations to participate in STEM. Given the specific organizational context of Experiment 2, we modified measures to assess positivity and sense of belonging in relation to the particular chemistry department participants read about. We also replaced the broader STEM aspirations scale with a more acute measure of participants' trust and comfort in the specific chemistry department. Finally, because Experiment 2 included a manipulation of the existence of diversity training, we added a direct assessment of the extent to which participants' anticipated experiencing discrimination in the department.

Across both experiments, we predicted that men would show better outcomes than women on all dependent variables only when participants were exposed to the reality of gender bias. In contrast, we expected that gender gaps would vanish when participants were told that STEM as a whole (Experiment 1) or a specific chemistry department (Experiment 2) exhibited gender equality. Further, in Experiment 2, we explored whether knowledge that the chemistry department had successfully completed a gender diversity training might be enough to counteract the negative effects of gender bias and close the STEM engagement gender gap.

## Experiment 1

The primary goal of Experiment 1 was to explore the causal impact of the existence of gender bias on women's and men's STEM engagement. To do so, we exposed adult participants to a news article reporting published experimental evidence of STEM gender bias (Moss-Racusin et al. 2012), the identical article modified to report that the experiment revealed no evidence of gender bias, or no article in a control condition. We then measured participants' sense of belonging in STEM, positive attitudes toward STEM, and aspirations to work in STEM.

We predicted that women would show lower outcomes than men would in the gender bias condition in which they were explicitly provided evidence of the reality of gender bias. We expected to observe similar gender differences in the control condition because the absence of explicit information about gender bias should allow participants to respond as they naturally would. Given the large body of evidence revealing the existence of gender bias in STEM (Knobloch-Westerwick et al. 2013; Milkman et al. 2012, 2015; Moss-Racusin et al. 2012; Reuben et al. 2014; Steinpreis et al. 1999), we predicted that the control condition would thus reflect a gender gap stemming from prevalent gender bias. However, in the gender equality condition, we expected that men's and women's outcomes would be equal. These results would reflect the idea that women are deterred from STEM fields due (at least in part) to the gender biases that exist across them.

### Method

#### Participants

Participants ($n = 322$, 180 women or 56% female, all percentages rounded to nearest whole number) were fluent English-speaking adults living in the United States who were at least 18-years-old ($M_{age} = 35$, $SD = 12.25$, range = 18–80). Of all participants, 242 (75%) were White, 24 (8%) were Black, 18 (6%) were Hispanic, 20 (6%) were East Asian, 4 (1%) were South Asian, 5 (1%) were Southeast Asian, 4 (1%) were Native American/ Pacific Islander, and 5 (2%) were Multiracial. Neither participants' age, $t(299) = .54$, $p = .587$, nor race, $\chi^2(7, 322) = 8.52$, $p = .289$, varied as a function of participants' gender—in other words, demographics were similar for men and women in the sample. Additionally, neither participant age (all $B$s < .21, all $p$s > .18) nor racial background (all $F$s < 1.40, all $p$s > .21) significantly impacted our results.

Of all participants, 24 (7%) had completed high school or obtained a GED, 76 (24%) had completed some college, 36 (11%) had completed a 2-year college degree, 145 (45%) had completed a 4-year college degree, 29 (9%) had received a Master's-level degree, 8 (3%) had received a Doctorate, and 4 (1%) had received a Professional degree (e.g., JD, MD). Of all participants, 79 (25%) reported that they intended to be, currently were, or had been a STEM major, and 75 (23%) reported having worked in STEM. Although there were some unsurprising main effects associated with participants' STEM background (such that those who reported more extensive STEM experience were also generally more engaged in STEM), in no case did these variables significantly interact with either the bias-existence condition or participants' gender (all $F$s < 1.53, all $p$s > .22). In other words, the results reported here appear to be largely unaffected by participants' prior levels of STEM experience.

#### Design, Procedure and Materials

The experiment utilized a 3 (existence of gender bias: gender bias, gender equality, control) × 2 (participant gender; male, female) between-subjects design. Participants were recruited through Amazon's Mechanical Turk (MTurk), an online marketplace for data collection providing multiple Human Intelligence Tasks (HITs) to eligible workers (Buhrmester et al. 2011). MTurk quickly generates large, diverse samples that are as representative as undergraduate participant pools (Paolacci et al. 2010), with superior test-retest reliability (Buhrmester et al. 2011). Participants could elect to participate in our experiment (entitled "Academic Fields Study") from among a list of HITs. Participation was restricted to those residing with the United States. All experimental materials were presented using the survey program Qualtrics. Participants read that we were interested in learning about their "views of different academic fields." After providing informed consent, they were randomly assigned to read the gender bias article, the gender equality article, or no article. All participants then completed the three dependent variable scales, which were presented in a random order (and items within each scale were randomized). Finally, participants reported their demographic information, completed the manipulation check, and were fully debriefed and compensated $.75.

**Existence of Gender Bias Manipulation** We manipulated the existence of gender bias in STEM using news articles reporting the results of scientific research. (All materials used in both experiments are available in an online supplement.) To do so, we modified an actual news article (Midura 2013) covering an experiment on STEM gender bias (Moss-Racusin et al. 2012). In the gender bias condition, the article accurately conveyed the existence of STEM gender bias as revealed by experimental data. Specifically, the article described the methodology and major results of Moss-Racusin et al. (2012) where in science faculty members rated a male lab manager applicant as more competent, hireable, and deserving of mentoring than the identical female applicant, and they also offered him a higher starting salary. These results were presented in a figure, in which we increased the magnitude of effect sizes obtained in the actual published research in order to ensure that the critical pattern of results was clear to non-specialists. Throughout the article, the actual evidence of gender bias was explicitly communicated. For example, the article was titled, "Research Finds Evidence of Gender Bias in Science Fields," and it stated that: "It appears that female scientists are not evaluated fairly based on their skills and abilities, but instead, are judged as inferior simply because of their gender."

We utilized the same article in the gender equality condition, but modified the language and figure in order to convey that research had not revealed evidence of gender bias. Throughout the article, the alleged evidence of gender equality was explicitly communicated. For example, the article was titled, "Research Finds Evidence of Gender Equality in Science Fields," and stated that, "It appears that female scientists are evaluated fairly based on their skills and abilities, and are not judged as inferior simply because of their gender." In the control condition, participants did not read an article.

**Sense of Belonging in STEM** To assess this construct, we utilized eight items modified from previous research (Cheryan et al. 2011; Nosek et al. 2002) on a Likert scale ranging from 1 (*not at all*) to 7 (*very much*). Items included, "How much do you think you belong in STEM?" and "To what extent do you 'feel at home' in STEM?" Items were averaged to form the sense of belonging index ($\alpha = .97$), with higher scores reflecting greater sense of belonging.

**Positive Attitudes toward STEM** We measured this construct using four items modified from previous research (Cheryan et al. 2009; Diekman et al. 2011) on a Likert scale ranging from 1 (*not at all*) to 7 (*very much*). Items included: "How positive is your impression of a career in STEM?" and "How positively do you feel toward STEM?" Items were averaged to form the positive attitudes scale ($\alpha = .91$), with higher scores indicating more positivity.

**STEM Aspirations** We measured participants' STEM aspirations utilizing seven items modified from previous research (Cheryan et al. 2011; Cheryan and Plaut 2010) as well as novel items developed for the purposes of the current research. Reponses were provided on a Likert scale ranging from 1 (*not at all*) to 7 (*very much*). Items included: "How interested are you in working in a STEM field?" and "If you were offered a job in a STEM field, how likely would you be to take it?" Items were averaged to create the aspirations scale ($\alpha = .96$), with higher scores reflecting greater aspirations.

**Manipulation Check** Participants responded to a manipulation check question, "What did the research described in the article you read about find?" Response options were: "Women encounter gender bias in science," "Women do not encounter gender bias in science," or "I was not asked to read an article." All participants answered the manipulation check correctly.

## Results

To test our predictions, we assessed participant gender differences in each condition (means, standard deviations, gender difference effect sizes, and correlations among study variables appear in Table 1; ANOVA results appear in Table 1s in the online supplement). For sense of belonging, as expected, men reported a higher sense of belonging than did women in the bias condition, $t(101) = 4.07$, $p < .001$, $d = .81$, and in the control condition, $t(117) = 2.14$, $p = .035$, $d = .40$. However, this difference vanished in the equality condition, $t(98) = .77$, $p = .441$, $d = .16$. This suggests that the existence of gender bias lead women to experience less sense of belonging in STEM than did men.

For positive attitudes toward STEM, contrary to expectations, women did not project significantly less positive attitudes toward STEM than did men in the control condition, $t(117) = 1.51$, $p = .134$, $d = .29$, although this difference was in the predicted direction. More importantly (and consistent with expectations), women did project significantly less positivity toward STEM than did men in the bias condition, $t(101) = 3.06$, $p = .003$, $d = .60$. This difference vanished in the equality condition, $t(98) = -.66$, $p = .511$, $d = .13$. These results suggest that women project less positivity toward STEM than men do only in the face of explicitly-identified gender bias.

For STEM aspirations, consistent with expectations, men reported greater aspirations to participate in STEM than did women in the bias condition, $t(101) = 3.25$, $p = .002$, $d = .64$, and in the control condition, $t(117) = 2.06$, $p = .042$, $d = .39$. However, in the gender equality condition, men and women aspired to participate in STEM at equal levels, $t(97) = .26$, $p = .797$, $d = .06$. This suggests that gender differences in STEM aspirations were only observed in the presence of gender bias.

**Table 1** Means and standard deviations of dependent variables by bias existence and participants' gender, Experiment 1

| Variables | Equality | | | | | Bias | | | | | *Control* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Men | | Women | | | Men | | Women | | | Men | | Women | | | Correlations | |
| | *M* | *SD* | *M* | *SD* | *d* | *M* | *SD* | *M* | *SD* | *d* | *M* | *SD* | *M* | *SD* | *d* | 1 | 2 |
| 1.Belonging | 4.48 | 1.35 | 4.24 | 1.65 | .16 | 4.51 | 1.93 | 3.11 | 1.50 | .81* | 4.61 | 1.51 | 3.94 | 1.81 | .40* | – | |
| 2. Positivity | 5.43 | 1.36 | 5.60 | 1.31 | .13 | 5.43 | 1.52 | 4.50 | 1.60 | .60* | 5.63 | 1.22 | 5.23 | 1.52 | .29 | .68*** | – |
| 3..Aspirations | 4.37 | 1.73 | 4.27 | 1.89 | .06 | 4.47 | 2.05 | 3.24 | 1.79 | .64* | 4.59 | 1.70 | 3.89 | 1.87 | .39* | .86*** | .71*** |

Positive scores reflect more sense of belonging, positivity, and aspirations. By convention, .20 reflects a small effect, .50 a medium effect, .80 a large effect (Cohen 1988)

* pairwise comparison significant at $p < .05$

***$p < .001$

### Moderated Mediation Analyses

Reflecting a conservative approach, we initially treated sense of belonging, positive attitudes, and aspirations as three outcome measures of interest because there was not sufficient pre-existing evidence to warrant hypothesizing that one of these variables might function as a mediator in this context. However, some prior research does suggest that sense of belonging may help to explain important gender differences in STEM. For example, Cheryan et al. (2011) found that in virtual classrooms with designs that conveyed computer science stereotypes, women reported lower interest and anticipated less success in computer science than did men *because* they felt less sense of belonging. Thus, we conducted similar moderated mediation analyses to explore the potential role of sense of belonging in explaining the effect of gender on both positive attitudes toward STEM and aspirations in STEM separately by condition (i.e., control, bias, and equality).

Specifically, we anticipated that there would be a significant indirect effect of gender on positive attitudes and aspirations via belonging in the control and bias conditions, but *not* in the equality condition. In other words, we expected that in the presence of gender bias, women would project less positivity and aspirations for STEM than men because they experienced less sense of belonging. To test this possibility, we ran a moderated mediation analysis using Hayes' PROCESS macro model 8 and 10,000 bootstrap resamples. We included gender as the predictor, sense of belonging in STEM as the mediator, condition as the moderator, and positive attitudes toward STEM and aspirations in STEM as the outcome variables (across two separate models, see Figs. 1 and 2). We expected that condition would moderate the effects of participant gender on both the mediator and outcomes (but not the effect of the mediator on the outcomes).

First, with positive attitudes toward STEM as the outcome variable (see Fig. 1), there were significant indirect effects (i.e., the confidence interval did not cross 0) of gender on attitudes via sense of belonging in the control condition and in the bias condition. As expected, the indirect effect was not significant in the equality condition. Second, focusing on STEM aspirations as the outcome variable (see Fig. 2), there were significant indirect effects of gender on aspirations through sense of belonging in the control condition and in the bias condition. As predicted, the indirect effect again was not significant in the equality condition.

Thus, in the presence of gender bias in the control and bias conditions, women had a lower sense of belonging in STEM than men did, and this depressed sense of belonging was associated with less positive attitudes toward STEM aspirations to participate in STEM. In the equality condition, however, gender did not predict sense of belonging (i.e., men and women felt similar levels of sense of belonging) and, consequently, the indirect effects in that condition were not significant. This suggests that sense of belonging might help to explain why women report less STEM positivity and aspirations than do men in the presence of gender bias.

## Discussion

Results from Experiment 1 provide support for our prediction that the existence of gender bias produces STEM gender gaps. When explicitly exposed to the reality of bias in the gender bias condition, women expected to experience less sense of belonging, positive attitudes, and aspirations to participate in STEM than did men. Of concern, these results provide novel evidence of the causal consequences of STEM gender bias, suggesting that it directly results in gender gaps in STEM engagement. However, these gender differences were fully eliminated when participants were exposed to the idea of gender equality. This suggests that women may project being just as enthusiastically engaged in STEM as men in the absence of gender bias.

When tacitly exposed to gender bias in the control condition (i.e., because gender bias exists in daily life), women projected significantly less sense of belonging and aspirations for STEM than did men. However, women and

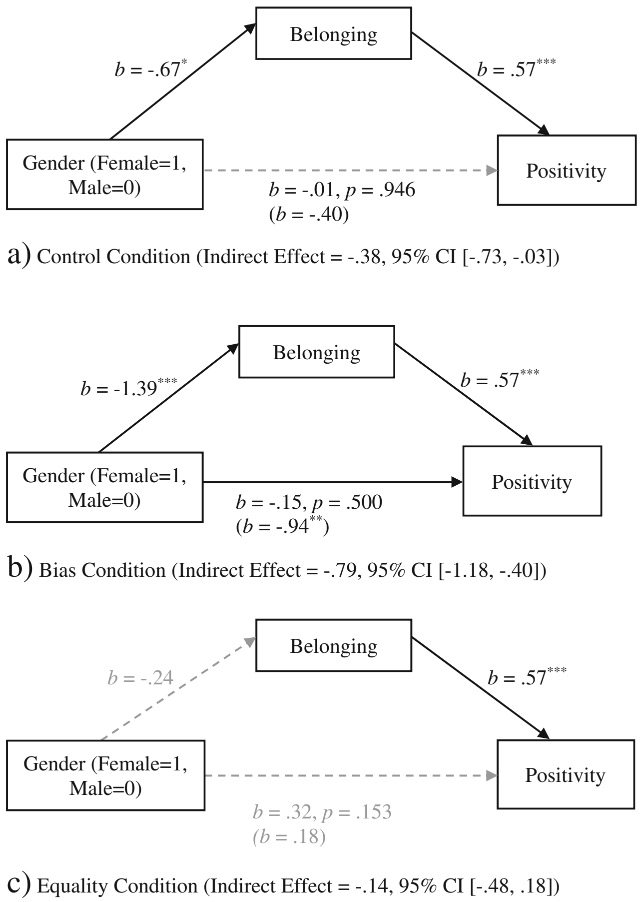


**Fig. 1** Moderated mediation results for positive attitudes toward STEM, Experiment 1. Moderated mediation model testing the indirect effect of gender on positive attitudes toward STEM via sense of belonging in STEM in the control, bias, and equality conditions. Figure 1a shows the mediational model in the control condition, Fig. 1b presents the mediational model in the bias condition, and Fig. 1c shows the mediational model in the equality condition. The total effects are shown in parenthesis, and the direct effects (i.e., controlling for sense of belonging) are shown without parenthesis. *b* indicates the unstandardized regression coefficient. Solid lines indicate significant effects at $p < .05$, and dotted lines represent nonsignificant effects. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$

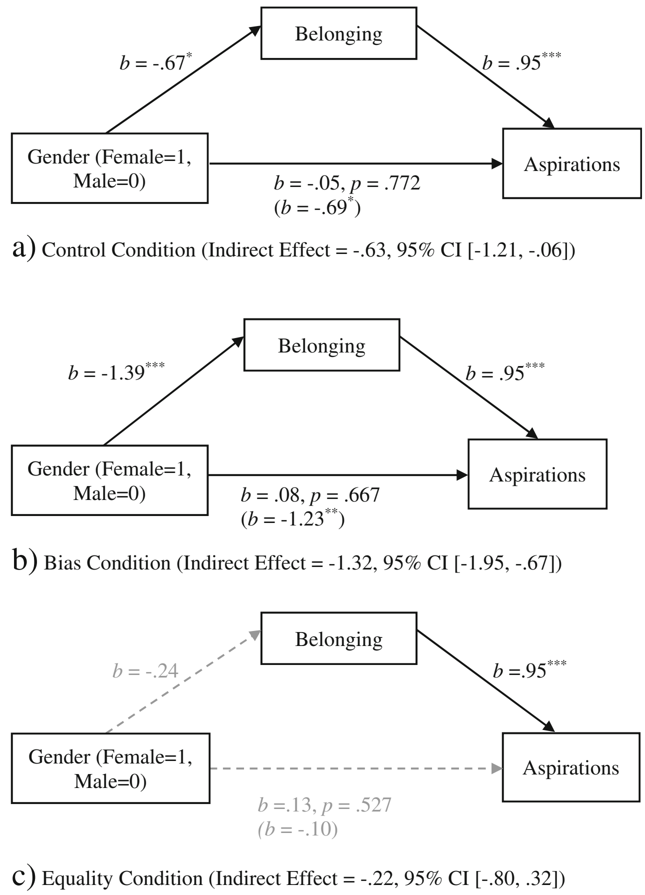


**Fig. 2** Moderated mediation results for STEM aspirations, Experiment 1. Moderated mediation model testing the indirect effect of gender on aspirations in STEM via sense of belonging in STEM in the control, bias, and equality conditions. Figure 2a shows the mediational model in the control condition, Fig. 2b presents the mediational model in the bias condition, and Fig. 2c shows the mediation model in the equality condition. The total effects are shown in parenthesis, and the direct effects (i.e., controlling for sense of belonging) are shown without parenthesis. *b* indicates the unstandardized regression coefficient. Solid lines indicate significant effects at $p < .05$, and dotted lines represent nonsignificant effects. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$

men did project equal levels of positivity toward STEM. These mixed findings may be attributable to the fact that the items in our positivity scale assessed judgments of STEM fields themselves, whereas the sense of belonging and aspirations scales measured participants' judgments of themselves (and their own potential experiences) in those fields. That is, the fact that positivity did not demonstrate a gender difference in the control condition suggests that women are not inherently less positive toward STEM itself than are men, and indeed, it is only in the presence of explicitly-identified gender bias that a positivity gender gap emerges. However, even with the control condition's absence of concrete information about the existence of STEM gender bias, women expected depressed personal outcomes (i.e., sense of belonging and aspirations) relative to men. The fact that results in the control condition mirrored the bias condition much more closely than the equality condition suggests that women may presume the existence of gender bias and respond by projecting their withdrawal from STEM. More broadly, only the equality condition's explicit, evidence-based claims of gender equality were sufficient to consistently close the gender gaps in projected STEM engagement.

Finally, consistent with some prior work (Cheryan et al. 2011), moderated mediation analyses suggested that the gender gaps in STEM positivity and aspirations that emerge in the presence of gender bias may be driven by sense of belonging. In other words, women may project less positivity toward and aspirations for STEM relative to men in part *because* they expect to feel less sense of belonging in STEM than do men when bias is present. Of importance, this relationship was not observed in the gender equality condition because women anticipated feeling just as much sense of belonging as men did when STEM was described as egalitarian.

## Experiment 2

Experiment 2 had two primary goals. First, we sought to replicate and extend the major pattern of results observed in Experiment 1 (i.e., that gender differences in STEM engagement emerge only in the face of gender bias) in a more acute context. To do so, we exposed participants to a summary of a chemistry department's re-accreditation review in which they were either found to have exhibited gender bias or not. Because Experiment 1 established that results for the control condition closely mirrored those in the gender bias condition, we eliminated the control condition in Experiment 2 in order to conserve power. Thus, the design of Experiment 2 allowed us to again explore the impact of the existence of gender bias on men's and women's STEM outcomes, but this time in a more proximal context (i.e., reactions to a specific department, rather than STEM as a whole). Accordingly, we modified existing scales and selected additional outcome variables from the literature that directly assessed participants' reactions to the specific department they read about rather than STEM more broadly.

We predicted that women would report more anticipated discrimination and less positive attitudes, sense of belonging, and trust and comfort in the chemistry department than men in the gender bias condition, but not in the gender equality condition. These results would further underscore the causal impact of gender bias on deterring women—not only from pursuing STEM globally, but also from joining particular STEM-learning environments and workplaces.

Second, we explored whether knowledge of organizational efforts to combat bias might close the perceived STEM engagement gender gap observed in Experiment 1. Specifically, participants read that the chemistry department had either successfully completed a gender bias training or had yet to complete the training. Although research has explored the impact of diversity training on attendees (e.g., Paluck 2006), relatively little work has explored the potential downstream impact on other members of the organization. Thus, we sought to determine whether diversity training might affect change not only by improving the attitudes and behaviors of attendees, but also by increasing expected engagement on the part of women who witness their organization taking action to address bias. Some related past work suggested that this might be likely. For example, Chaney et al. (2016, Experiment 1) found that White women exposed to a company that held either a "Fostering Women's Success" or a "Fostering Racial Minorities' Success" program anticipated greater workplace inclusion than those in a control condition (who merely read about a neutral "Fostering Employee Success" program). Because there were no differences in White women's perceived inclusion across the two experimental conditions (and because Men of Color showed similar effects when exposed to companies who won awards for either gender or racial diversity; Chaney et al. 2016, Experiment 3), these results suggest that the presence of diversity structures designed to benefit one stigmatized group may also positively impact members of other targeted groups.

However, the current work built upon this existing research in at least four critical ways. First, Chaney et al. (2016) explored reactions to general organizational diversity structures (e.g., a course designed to improve employees' intergroup communication in Experiment 1 or awards recognizing generally positive workplace climates for women or People of Color in Experiments 2–4), rather than trainings specifically targeting gender bias itself. Second, Chaney et al. (2016, Experiment 1) manipulated the presence of diversity trainings, but did not indicate whether the trainings were actually successful. Third, participant samples in this research were demographically homogeneous (i.e., White women in Experiments 1 and 2, Men of Color in Experiment 3, and White men in Experiment 4), such that it was not possible to determine whether diversity structures would produce differences in engagement across participants from stigmatized and non-stigmatized groups (a key question in the current research). Finally, the prior research focused on a corporate (rather than a STEM) context. Thus, we expanded upon existing work by exploring whether the existence of gender bias, the successful completion of a gender bias training, and participants' own gender would shape both male and female participants' projected STEM outcomes. Due to the paucity of existing research, we did not make an a priori prediction regarding the extent to which the completion of a successful diversity training might equalize men's and women's STEM engagement (even in the face of gender bias), but rather, we sought to provide the first known experimental test of this possibility.

### Method

#### Participants

Participants ($n = 429$, 224 women or 52% female) were fluent English-speaking adults living in the United States who were at least 18-years-old ($M_{age} = 37$, $SD = 11.87$, range = 18–75). Of all participants, 343 (80%) were White, 27 (6%) were Black, 18 (4%) were Hispanic, 12 (3%) were East Asian, 9 (2%) were South Asian, 5 (1%) were Southeast Asian, 2 (2%) were Native American/ Pacific Islander, and 10 (2%) were Multiracial. Consistent with Experiment 1, neither participants' age, $t(427) = -.69$, $p = .492$, nor race, $\chi^2(9, n = 429) = 11.03$, $p = .274$, varied as a function of participants' gender, nor did participants' age (all $B$s $< .11$, all $p$s $> .23$) or racial background (all $Fs < 1.40$, all $ps > .19$) significantly impact results.

Of participants, 87 (20%) had completed high school or obtained a GED, 92 (22%) had completed a 2-year college degree, 180 (42%) had completed a 4-year college degree, 53 (12%) had received a Master's-level degree, 8 (2%) had received a Doctorate, and 9 (2%) had received a Professional degree (e.g., JD, MD). Of participants, 138 (32%) reported that they intended to be, currently were, or had been a STEM major, and 111 (26%) reported having worked in STEM. Consistent with Experiment 1, there were no significant interactions between these variables and the bias existence condition or participants' gender (all $Fs < 1.70$, all $ps > .19$). Thus, once again, results appear to be largely unaffected by participants' prior levels of STEM experience.

### Design, Procedure, and Materials

The experiment utilized a 2 (existence of gender bias: gender bias, gender equality) × 2 (diversity training status: trained, not trained) × 2 (participant gender) between-subjects design. Participants were again recruited via MTurk, where they could elect to participate in our experiment (entitled "You be the Judge") from among a list of HITs. Workers who had completed Experiment 1 were not permitted to participate in Experiment 2. Participation was again restricted to those residing with the U.S. All experimental materials were presented using the survey program Qualtrics. Participants read that we were interested in learning about their "perceptions of different academic departments." After providing informed consent, they were randomly assigned to read one of four department re-accreditation review summaries. They next completed the three dependent variable scales, which were presented in a random order (and items within each scale were randomized). Finally, participants reported their demographic information, completed the manipulation checks, and were fully debriefed and compensated $1.00.

**Re-Accreditation Reviews** To manipulate both the existence of bias and the completion of diversity training, we created summaries of a re-accreditation review that had allegedly taken place for a chemistry department at a major "R1" American University in the last year. Participants viewed a summary form which contained information in four sections: "General Department Information" (e.g., size, institution type); "Department Performance" (e.g., number of competitive research grants awarded); "Department Climate" (e.g., number of discrimination complaints and academic integrity violations); and "Post-Review Information." The first two sections contained neutral filler data that was standardized across all conditions. Information was presented using tabular data from the current review (2016), the previous review (2013), and current averages for similar institutions. Categories with improvement since the last review were highlighted in grey, whereas those with decreases were highlighted in orange (allowing participants to easily visualize problem areas). The filler data presented first reflected a department that performed well over time and relative to peers.

Our manipulations were contained in the table on "Department Climate." To manipulate the existence of gender bias, we displayed the number of "reported gender bias/discrimination complaints." (Note that to bolster the cover story and allay suspicions, participants read that "information regarding racial bias is still being processed and is not yet available for this department.") In the gender bias condition, the current number of complaints was 31, higher than those reported in the last review (22) and the average for similar institutions (18). Further, the reviewer expressed strong concerns about gender bias in their comments. For example, the reviewer stated: "The current data revealed a disturbing pattern of gender bias." In the gender equality condition, the current number of complaints was 11, and the reviewer noted that the climate was no longer problematic and stated explicitly that they were not concerned about gender bias. For example, they wrote: "The current data revealed an encouraging pattern of gender equality."

The manipulation of diversity training status was contained in a final section, entitled "Post-Review Information." In both the gender bias and equality conditions, participants read that the reviewer had recommended completion of a "validated gender diversity training course," but in the gender bias condition, this was "out of necessity," and in the equality condition, this was merely "as a standard procedure." In the trained condition, participants read that the department had successfully completed the recommended course, the date of completion, the completion grade of "Highly Successful," along with the definition that this grade "indicates receptive faculty, active engagement with course material, demonstrated greater understanding of gender diversity issues, commitment to behavioral improvements and equitable treatment of male and female department community members." Finally, the reviewer noted that "The faculty was very successful. They are committed to gender equity." In the untrained condition, participants read that the department had not yet completed the recommended gender diversity training course. Under "Completion Date" and "Completion Grade," the words "NOT YET COMPLETED" appeared in red.

**Anticipated Discrimination in the Chemistry Department** We sought to directly assess the extent to which participants expected to experience gender-based discrimination in this particular chemistry department. To do so, we modified an existing scale of anticipated discrimination (Wilton et al. 2015). Participants responded to six items using a Likert scale ranging from 1 (*not at all*) to 7 (*very much*). Items included: "I would expect to personally experience gender bias in this chemistry department" and "I would expect to be excluded from some important professional opportunities in this chemistry department." Items were averaged to form the anticipated discrimination scale, with higher scores reflecting higher levels of anticipated discrimination ($\alpha = .89$).

**Belonging, Positive Attitudes, and Trust and Comfort** We utilized the same sense of belonging scale as in Experiment 1, but modified it to pertain to this particular chemistry department rather than to STEM as whole ($\alpha = .95$). We utilized the positive attitudes scale from Experiment 1, but modified it to pertain to this particular chemistry department rather than to STEM as whole ($\alpha = .96$). We sought to utilize a measure that was conceptually similar to the aspirations scale utilized in Experiment 1, but that would be appropriate for a specific departmental context (rather than STEM fields as a whole). To do so, we drew upon the existing organizational literature and utilized a modified version of the trust and comfort scale, which assess the extent to which individuals feel motivated, satisfied, and determined to work hard and succeed in an organization (Purdie-Vaughns et al. 2008). Participants responded to five items indicated on a Likert scale ranging from 1 (*not at all*) to 7 (*very much*). Items included: "I would like to be a member of this chemistry department" and "This chemistry department would inspire me to do the very best job that I can." Items were averaged to form the trust and comfort scale, with larger numbers reflecting greater levels of trust and comfort in the chemistry department ($\alpha = .96$).

**Manipulation Checks** To assess the efficacy of our manipulations, participants responded to two manipulation check questions. The first was: "What was the reviewer's opinion of the climate in this department?" Response options were "The reviewer had concerns about gender bias," and "The reviewer did not have concerns about gender bias." The second was "Have the department faculty completed the required gender diversity training?" Response options were: "Yes, they have completed the training" and "No, they have not yet completed the training." As in Experiment 1, all participants answered the manipulation checks correctly.

## Results

### Testing Predicted Gender Differences

To test our predictions, we again assessed participant gender differences in each condition (means, standard deviations, gender difference effect sizes, and correlations among study variables appear in Table 2; ANOVA results appear in Table 2s in the online supplement). For anticipated discrimination, as predicted, in the gender bias condition, women anticipated that they would encounter significantly more discrimination if they joined this chemistry department than did men, $t(219) = 5.20$, $p < .001$, $d = .70$. However, this difference vanished in the gender equality condition, $t(205) = 1.59$, $p = .114$, $d = .22$. This suggests that women believed they would encounter more discrimination than did men only in the presence of gender bias.

Results also supported predictions for the remaining variables. For sense of belonging, women predicted experiencing significantly less sense of belonging than men in the gender bias condition $t(219) = 3.74$, $p < .001$, $d = .51$. However, this difference was not observed in the gender equality condition, $t(205) = 1.88$, $p = .062$, $d = .26$. For positive attitudes toward this chemistry department, women projected less positive attitudes toward the chemistry department than did men in the gender bias condition, $t(218) = 3.09$, $p = .002$, $d = .42$. However, this difference disappeared in the gender equality condition, $t(205) = .37$, $p = .713$, $d = .05$. For trust and comfort, women perceived significantly less trust and comfort in the chemistry department than did men in the gender bias condition, $t(219) = 2.60$, $p = .010$, $d = .35$. However, this gender difference vanished in the gender equality condition,

**Table 2** Means and standard deviations of dependent variables by bias existence and participants' gender, Experiment 2

| Variables | Equality | | | | | Bias | | | | | Correlations | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Men | | Women | | | Men | | Women | | | | | |
| | *M* | *SD* | *M* | *SD* | *d* | *M* | *SD* | *M* | *SD* | *d* | 1 | 2 | 3 |
| 1. Anticipated discrimination | 2.55 | 1.10 | 2.81 | 1.25 | .22 | 3.57 | 1.17 | 4.50 | 1.48 | .70* | – | | |
| 2. Sense of belonging | 5.01 | 1.24 | 4.65 | 1.50 | .26 | 4.07 | 1.52 | 3.33 | 1.39 | .51* | −.58*** | – | |
| 3. Positivity | 5.47 | 1.20 | 5.41 | 1.80 | .05 | 4.05 | 1.47 | 3.39 | 1.70 | .42* | −.70*** | .78*** | – |
| 4. Trust and comfort | 5.26 | 1.35 | 5.31 | 1.25 | .04 | 3.81 | 1.52 | 3.25 | 1.69 | .35* | −.68*** | .78*** | .91*** |

Positive scores reflect more anticipated discrimination, sense of belonging, positivity, and aspirations. By convention, .20 reflects a small effect, .50 a medium effect, .80 a large effect (Cohen 1988)

* pairwise comparison significant at $p < .05$

***$p < .001$

$t(205) = .29$, $p = .772$, $d = .04$. Taken together, these results provide consistent support for predictions, such that women projected less STEM engagement relative to men only in the presence of gender bias.

### Assessing the Exploratory Research Question

Because we did not have specific a priori predictions regarding whether the completion of a successful diversity training might impact men's and women's STEM outcomes, we assessed this exploratory research question by submitting each of the four dependent variables to a three-way between-subjects ANOVA and examining main and interaction effects associated with the training status variable. For interested readers, Table 3s (available in the online supplement) contains all means, standard deviations, and gender difference effect sizes as a function of bias existence, diversity training status, and participant gender. There was a main effect of training status for each dependent variable (all $F$s > 8.02, all $p$s < .005), such that whether or not the department had exhibited bias, both male and female participants' expected more positive outcomes when departments had been successfully trained relative to when they had not been trained.

However, no two- or three-way interactions involving training status were significant for any variable (all $F$s < 2.89, all $p$s > .090), with one exception: for anticipated discrimination, there was a significant interaction between training status and participant gender, $F(1, 420) = 5.80$, $p = .017$, $\eta p^2 = .014$. Simple effects tests revealed that when the department had not been trained, women ($M = 4.16$, $SD = 1.58$) anticipated greater discrimination than did men ($M = 3.14$, $SD = 1.27$), $t(210) = 5.17$, $p < .001$, $d = .71$. However, when the department completed the diversity training, women ($M = 3.28$, $SD = 1.53$) anticipated similar levels of discrimination as did men ($M = 3.01$, $SD = 1.21$), $t(214) = 1.44$, $p = .150$, $d = .20$. These results indicate that knowledge of a successful diversity training equalized women's and men's levels of anticipated discrimination (and because the three-way interaction was not significant, this did not vary as a function of whether the department had been biased). However, we reiterate that because no additional two or three-way interactions involving training status were significant for other outcomes, it appeared that training status largely did not impact the predicted effects of gender bias and participant gender on STEM engagement in the current research.

### Moderated Mediation Analyses

As in Experiment 1, there was not sufficient pre-existing theoretical rationale to make a priori predictions regarding mediation. However, we again explored whether sense of belonging would emerge as a significant mediator in the gender bias (but not gender equality) condition. Further, because Experiment 2 included a direct measure of participants' anticipated discrimination in the department, we were able to test this as a mediator as well. Consistent with Experiment 1, we anticipated that there would be a significant indirect effect of gender on positive attitudes and trust and comfort via belonging and anticipated discrimination in the bias condition, but *not* in the equality condition. To test this possibility, we ran moderated parallel mediation analyses using Hayes' PROCESS macro model 8 and 10,000 bootstrap resamples. We included gender as the predictor, belonging and anticipated discrimination as two parallel mediators, condition as the moderator, and attitudes toward STEM and trust and comfort as the outcome variables (across two separate models, see Figs. 3 and 4). As in Experiment 1, we again expected that condition would moderate the effects of participant gender on both the mediators and outcomes (but not the effects of the mediators on the outcomes).

Results were consistent with and expanded upon the moderated mediation analyses conducted in Experiment 1. First, with positive attitudes toward STEM as the outcome variable (see Fig. 3), in the bias condition there were significant indirect effects (i.e., the confidence interval did not cross 0) of gender on positive attitudes via both sense of belonging and anticipated discrimination. As expected, in the equality condition, neither the indirect effect of belonging nor anticipated discrimination was significant. Second, focusing on trust and comfort as the outcome variable (see Fig. 4), in the bias condition, there were significant indirect effects of gender on trust and comfort via both sense of belonging and anticipated discrimination. As expected, in the equality condition, neither the indirect effect of belonging nor anticipated discrimination was significant.

Thus, in the bias condition, women reported lower sense of belonging and higher anticipated discrimination than did men, and feeling greater sense of belonging predicted more positive attitudes and trust and comfort. Simultaneously, anticipating discrimination predicted less positive attitudes and trust and comfort. However, in the equality condition, gender was not significantly related to belonging (i.e., men and women felt a similar sense of belonging in STEM) and did not predict anticipated discrimination. As a result, the indirect effects in the equality condition were not significant.

## Discussion

Using novel stimulus materials, additional dependent variables, and assessing outcomes at the level of one particular chemistry department rather than STEM as a whole, results from Experiment 2 provide additional support for our prediction that gender bias leads to gender gaps in STEM. When facing departmental gender bias, women anticipated more discrimination and projected

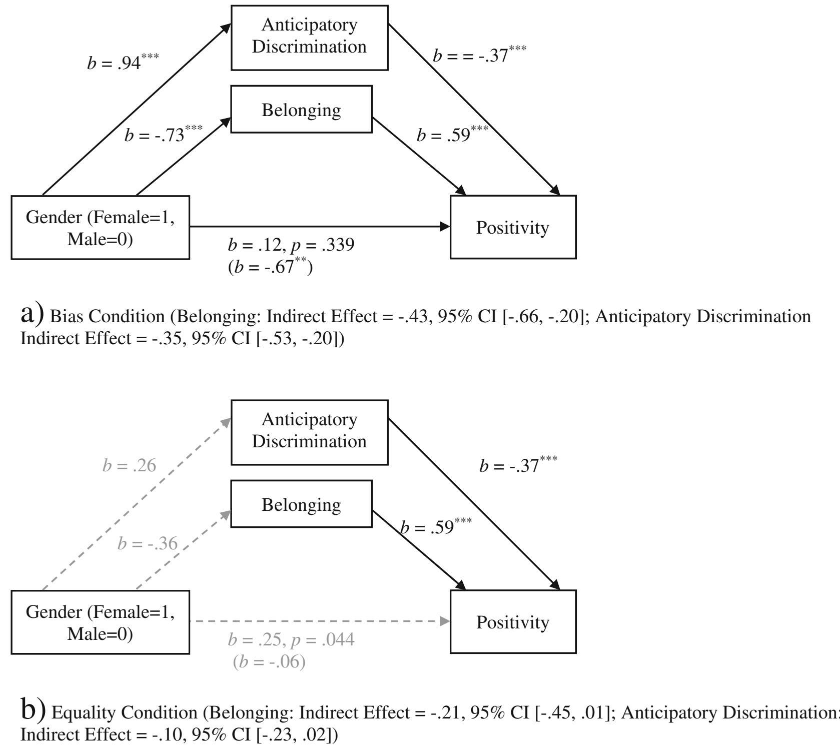


**Fig. 3** Moderated mediation results for positive attitudes toward STEM, Experiment 2. Moderated parallel mediation model testing the indirect effects of gender on positive attitudes via sense of belonging and anticipated discrimination in the bias and equality conditions. Figure 3a shows the mediational model in the bias condition, and Fig. 3 presents the mediational model in the equality condition. The total effects are shown with parenthesis, and the direct effects (i.e., controlling for belonging and anticipated discrimination) are shown without parenthesis. *b* indicates the unstandardized regression coefficient. Solid lines indicate significant effects at $p < .05$, and dotted lines represent nonsignificant effects. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

less sense of belonging, positivity, and trust and comfort toward a particular chemistry department than did men. However, when considering an unbiased department, women and men responded with equal levels of STEM engagement as assessed by these variables.

Results suggested that successfully completing a diversity training designed to eradicate gender bias was not a particularly potent tool in closing the gender gap that emerged in the presence of gender bias. Although main effects of training status emerged such that all participants anticipated less

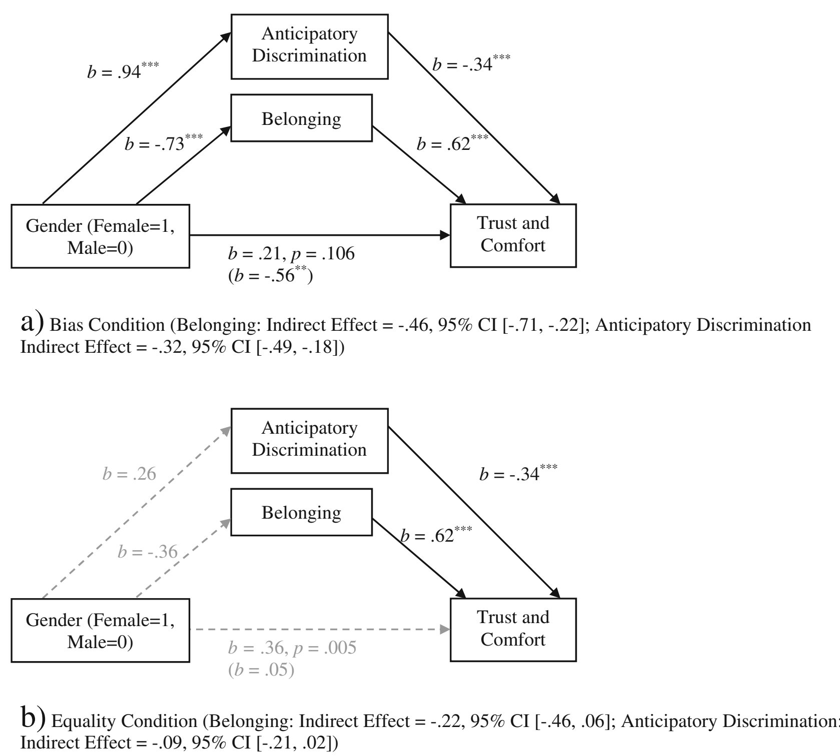


**Fig. 4** Moderated mediation results for trust and comfort, Experiment 2. Moderated parallel mediation model testing the indirect effects of gender on trust and comfort via sense of belonging and anticipated discrimination in the bias and equality conditions. Figure 4a shows the mediational model in the bias condition, and Fig. 4b presents the mediational model in the equality condition. The total effects are shown with parenthesis, and the direct effects (i.e., controlling for belonging and anticipated discrimination) are shown without parenthesis. *b* indicates the unstandardized regression coefficient. Solid lines indicate significant effects at $p < .05$, and dotted lines represent nonsignificant effects. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

discrimination and predicted experiencing more sense of belonging, positive attitudes, and trust and comfort toward the department when it had been trained relative to when it was untrained, these effects largely did not interact with participants' gender. In other words, although training was associated with general increases in engagement (which could benefit women as well as men), it did not consistently shrink the bias-linked gender gap. The exception to this trend was anticipated discrimination, which demonstrated an interaction between training status and gender such that women anticipated more discrimination than did men only when the department was untrained.

Consistent with Experiment 1, we explored the mediating role of sense of belonging in explaining the relationship between gender and positivity, as well as gender and trust and comfort. New to Experiment 2, we also assessed the mediating role of anticipated discrimination. Results of moderated mediation analyses revealed that the gender gaps in engagement that emerged in the bias condition were driven by both mediators, such that in the presence of bias, women projected less sense of belonging and anticipated more discrimination, and these in turn were associated with less positivity and trust and comfort toward the chemistry department (relative to men). In the gender equality condition, women and men projected similar levels of sense of belonging and anticipated discrimination, and thus they reported similar levels of positivity and trust and comfort.

## General Discussion

Taken together, results from two experiments provide support for the idea that the existence of gender bias is causally related to gender gaps in STEM engagement. When explicitly exposed to the reality of gender bias at the level of STEM as a whole (Experiment 1), women projected less sense of belonging, positive attitudes, and aspiration for STEM than did men. Further, gender gaps emerged for most variables when participants were tacitly exposed to gender bias in the control condition. Expanding upon these results, when women were exposed to the reality of gender bias at the level of one particular chemistry department (Experiment 2), they anticipated experiencing more discrimination and projected less sense of belonging, positive attitudes, and trust and comfort in the department than did men. Building upon prior research demonstrating the consequences of one instructor's sexism on women's performance (Adams et al. 2006) and interest in science (Thoman and Sanson 2016), the current research suggest that gender bias deters women from STEM.

These results are inconsistent with the idea that women are underrepresented in STEM simply because they choose not to enter it. Indeed, we show that gender gaps emerge only in the presence of gender biases that are detrimental to women's equitable treatment, well-being, and professional progress. As such, the current results stand in contrast to prior claims that gender bias does not contribute to STEM gender gaps. For example, Ceci and Williams (2011, p. 3157) asserted that, "Women's current underrepresentation in math-intensive fields is not caused by discrimination in these domains, but rather to sex [sic] differences in resources, abilities, and choices (whether free or constrained)." Although we agree that many factors jointly result in women's underrepresentation (Cheryan et al. 2009; Diekman et al. 2017; Murphy et al. 2007; Nosek et al. 2009; Stout et al. 2011; Williams and Ceci 2012), we strongly contend that gender bias should now be added to the list.

Although there was insufficient existing theory and evidence on which to base strong a priori process predictions, we conducted moderated mediation analyses to explore the potential mechanisms underscoring effects. Consistent with some prior work (Cheryan et al. 2011), evidence emerged in Experiment 1 suggesting that in the presence of gender bias (i.e., in the bias and control conditions), women projected less positivity and aspirations for STEM than did men because they experienced less sense of belonging. In contrast, because women and men experienced equivalent levels of sense of belonging in the gender equality condition, they in turn projected equal levels of positivity and aspirations. Expanding upon these results in Experiment 2, in the gender bias condition, women projected less positivity and trust and comfort toward a specific chemistry department than did men because they experienced less sense of belonging and also because they anticipated encountering more discrimination. However, in the gender equality condition, women and men anticipated equal levels of discrimination and experienced an equivalent sense of belonging, and thus they reported equal levels of positivity and trust and comfort. Although these analyses were exploratory, they further underscore the serious consequences of gender bias for women's STEM engagement, and they suggest that sense of belonging and anticipated discrimination may play key roles in explaining these effects.

### Limitations and Future Research Directions

The current research is not without limitations, which reveal interesting avenues for future research. Although our key prediction focused on exploring gender gaps in the gender bias versus gender equality conditions, we also generally expected to observe gender gaps in the control condition in Experiment 1 (due to the fact that participants in this condition were tacitly exposed to real-world gender bias in STEM; see the following). However, unlike the other variables, positive attitudes toward STEM did not exhibit a gender gap in the control condition. We speculated that this may be due to the fact that items in our positivity scale assessed judgments of STEM fields themselves (e.g., "How positively do you feel toward STEM?"). In contrast, the sense of belonging and aspirations scales more directly assessed participants' predictions of their own experiences in those fields (e.g., "How much do you think you belong in STEM?"; "How interested are you in working in a STEM field?"), and thus may have been more sensitive to the impact of tacit bias present in the control condition. However, future research should examine this possibility by utilizing different measures of positivity toward STEM.

More broadly, it is important to consider the nature of Experiment 1's control condition itself. We utilized a no-content control condition in order to help clarify the causal impact of gender bias on women's and men's STEM engagement. However, unlike many control conditions, ours cannot truly reflect a neutral level of the independent variable. Because participants are regularly exposed to the broader culture (and its accompanying stereotypes and biases), it was not possible to create a control condition that truly had *no* level of existence of gender bias. Results indicated that in the absence of explicit information about the existence of gender bias or equality, in most cases, participants responded as though they had been exposed to evidence of gender bias. In other words, likely because bias exists in the real world, it appears that it is not necessary to explicitly mention bias in order for it to produce a gender gap.

However, the lack of a truly neutral control condition does raise the possibility that rather than gender bias creating a gender gap, gender equality may remove a gender gap that already exists for a myriad of other reasons. Indeed, the current results could be interpreted as highlighting the potential positive impact of gender equality for women's STEM engagement. Future research should explore this possibility by seeking to develop an alternate or more informative control condition. For example, a control that minimizes the salience of gender bias, distracts participants from the reality of gender bias, or otherwise produces a more neutral condition could help to further isolate the conditions driving the causal action of the effect. In any case, the fact that men's and women's outcomes were indistinguishable in the gender equality condition (across two experiments and seven outcomes) is encouraging.

The current research did not reveal particularly strong effects associated with diversity training. Although knowledge that a chemistry department had successfully completed a gender diversity training did close the gender gap in anticipated discrimination, it did not significantly impact the gender gap for the remaining variables. However, because we orthogonally manipulated the existence of bias and the completion of a diversity training, our experimental design reflected a relatively conservative test of the impact of diversity trainings. Indeed, the "unbiased" department may have been viewed as more biased than intended, for three reasons. First, the design necessitated that all departments were noted to have struggled a bit with bias at the time of their prior review. Second, all departments were recommended to take the training, whether they had exhibited bias at the current review or not (although in the gender equality condition, the reviewer noted that the recommendation was merely a formality). Third, even the unbiased departments still had 11 current gender bias/discrimination complaints (although this number was lower than both the previous review and similar departments and the reviewer noted that they were not concerned about gender bias). Thus, future less conservative tests may reveal more powerful effects of diversity training for additional variables beyond anticipated discrimination. Conversely, future research could also include a condition in which departments took but *failed* the training. This should indicate even higher or more persistent levels of bias, resulting in exacerbations of gender gaps.

The current research cannot conclusively determine *why* the diversity training manipulation was not more successful. Future work should explore this idea by assessing the perceived success and long-term impact of diversity trainings in STEM. Perhaps it is the case that trainings are largely viewed as ineffective (consistent with some of the existing assessment literature; Anand and Winters 2008; Dobbin et al. 2015; Legault et al. 2011), or participants are skeptical that trainees will be willing and/or able to implement positive changes in their broader organizational culture over time. If so, additional future work could seek to develop a new manipulation that assuages these concerns and, thus, might be more effective in closing STEM gender gaps. Regardless, we do note that the effects for anticipated discrimination are heartening; mere knowledge of a successful training equalized the extent to which men and women anticipated experiencing discrimination in a male gender-typed field that has historically demonstrated gender bias. Because anticipated discrimination emerged as a mediator of the effect of gender on positive attitudes and trust and comfort, closing the gender gap for this variable could have additional positive ripple effects. Indeed, our results hint at the possibility of a second-level benefit to diversity trainings (i.e., for both attendees themselves and stigmatized group members throughout their organization).

Additional work is needed to determine the extent to which results generalize beyond the current context. For example, as is the case with many experiments, our sample was not randomly-selected nor representative of the broader underlying U.S. population. Although MTurk samples may be more diverse and conscientious than traditional undergraduate samples are (Buhrmester et al. 2011; Paolacci et al. 2010), results may be idiosyncratic to those with the time, means, and disposition to complete our HIT. Thus, future research should attempt to replicate the current results with a randomly-selected, representative sample. Additionally, the current research failed to incorporate attention checks to ensure that participants were carefully attending to the materials. Although we were heartened to see that all participants passed the manipulation checks (suggesting that they were paying adequate attention), future research should directly assess participants' engagement by including attention checks.

Of importance, the current research concerns the ways in which the existence of bias may impact STEM engagement. However, the current MTurk samples were composed of participants with a range of ages, educational backgrounds, and levels of STEM experience rather than solely those who were in the early stages of determining whether or not to participate in STEM. In other words, our samples included both those who were already committed to working in STEM (or not), as well as those still navigating these decisions. Future research should determine whether the current findings replicate with a sample of participants who have yet to establish their level of STEM involvement (e.g., elementary, high school, or early college students) and may thus be particularly likely to be influenced by factors such as the existence of gender bias. However, it is important to reiterate that our results did not appear to be moderated by these demographic variables, such that the existence of gender bias appeared to deter even those women who were already committed to (or participating in) STEM careers. Additionally, because many in our sample had already settled upon a particular career, it is possible that their levels of STEM engagement may be relatively fixed and thus less likely to be influenced by various situational factors (such as existence of gender bias) compared to younger, professionally-undecided participants. Thus, our samples may actually have afforded relatively conservative tests of our predictions. If so, the fact that results emerged even among these samples may speak to the powerful role of gender bias in shaping STEM engagement.

Further, the extent to which the same types of effects would be observed at other STEM career junctures (such as applying for postdoctoral and faculty positions, as well as promotion) remains unclear. It would also be interesting to determine whether similar results emerge in other organizational contexts in which women remain underrepresented, such as medicine, corporate leadership, and politics. Additionally, future work could determine whether the existence of gender biases targeting men produce engagement gender gaps favoring women in fields in which men are underrepresented, such as nursing and early elementary education. Also, recent work suggests that evidence-based confrontations of sexism (i.e., those in which participants receive concrete evidence of their sexist behaviors) are more effective than confrontations without concrete evidence (Parker et al. 2018). The current manipulation contained evidence-based claims of gender bias or equality (i.e., results of experimental research in Experiment 1 and a formal accreditation review in Experiment 2). It would thus be interesting to determine whether other types of claims (e.g., anecdotal or interpersonal) are sufficient to produce the same results or whether only evidence-based claims of gender equality can close STEM gender gaps.

Finally, the current work investigated the extent to which gender bias produced gender gaps in STEM engagement, but it did not take an intersectional perspective in exploring this question. Researchers have convincingly argued that because individuals are members of numerous stigmatized (and non-stigmatized) groups, it is essential to consider these multiple group identities simultaneously in order to avoid "intersectional invisibility" (Purdie-Vaughns and Eibach 2008, p. 377). In the context of the current research questions, future work should explore whether different forms of marginalization may compound the gender effects observed here. For example, although racial background did not significantly impact our results, 75% of participants in Experiment 1 and 80% in Experiment 2 identified as White. Future research should make an effort to recruit samples of racially diverse participants in order to guarantee sufficient statistical power to test for racial differences. Moreover, future work could fruitfully examine whether gender bias-linked gender gaps in STEM engagement are exacerbated by membership in additional stigmatized groups. For example, women from the LBTQ community, those with lower socio-economic backgrounds, or those who are differentially-abled may be particularly deterred by STEM gender bias. Relatedly, future work could meaningfully examine whether knowledge of other forms of bias (e.g., racial bias, homophobia) produce effects similar to those observed here with gender bias and, in turn, whether these effects vary as a function of individuals' multiple social identities.

## Practice Implications

The current results have important implications for interventions and policies aimed at increasing women's STEM participation. For example, proponents of the view that gender bias does not contribute to STEM gender gaps have argued that current policies aimed at reducing gender bias are unnecessary and should thus be eliminated (e.g., Ceci et al. 2014; Ceci and Williams 2011). Given the current evidence that gender bias does indeed produce gender gaps, we assert that evidence-based interventions targeting gender bias are needed to boost women's representation. Unfortunately, very few tested gender bias interventions have been developed (Moss-Racusin et al. 2014; Paluck 2006), perhaps in part because of persistent arguments that they are not needed. Of importance, recent work has identified at least two evidence-based gender bias interventions that appear to be effective with STEM attendees (Carnes et al. 2015; Devine et al. 2017; Moss-Racusin et al. 2018; Pietri et al. 2017). Because the current results highlight the detrimental role of STEM gender bias, we argue that new and existing interventions should be thoroughly tested and implemented on a wide scale. Given our moderated mediation results, interventions designed to increase women's sense of belonging and reduce their levels of anticipated discrimination may be particularly effective. Additionally, although current funding mechanisms (such as the National Science Foundation's ADVANCE program) have generously supported research aimed at broadening women's participation in STEM, we suggest that additional funding explicitly designated to support work on understanding and mitigating the impact of gender bias could help to further close the existing gender gaps.

However, one potential interpretation of the current results is that interventions and other messages referencing evidence of gender bias in STEM could deter talented women who would otherwise enter the STEM workforce (Williams and Ceci 2015). That is, if gender gaps emerge only in the presence of gender bias, perhaps it is not fruitful to pursue interventions that emphasize the existence of gender bias. And yet, we contend that ignoring the reality of bias does not constitute a realistic preventative approach. Even if young women are initially unaware of gender bias, the evidence suggests that they are likely to encounter it once they enter the STEM pipeline (Robnett 2016; Steele et al. 2002; Williams et al. 2016). At that point, women who are not familiar with established trends of bias may attribute biased incidents to their own shortcomings, with various negative consequences for their future performance and engagement (Crocker and Major 1989). Indeed, attributing negative feedback to their own abilities rather than to discrimination was negatively associated with women's self-esteem (Major et al. 2003). Relatedly, interventions that openly acknowledge the difficulty, pervasiveness, and malleability of challenging situations have ameliorated racial and ethnic gaps in achievement, engagement, and even health (Walton and Cohen 2011; Yeager et al. 2016).

Thus, although failing to discuss the reality of gender bias could potentially increase the number of women initially entering STEM, failing to meaningfully *address* gender bias would not be likely to halt women's attrition (nor enable a positive and productive working environment for those who do choose to stay). More broadly, remaining silent about bias has historically failed to protect stigmatized group members from its harmful effects. Thus, instead of deterring individuals from engaging with STEM due to its lingering gender biases, we hope that the current research will inspire additional work on the best ways to rid STEM of these pernicious biases and their consequences. Rather than frightening women away from STEM, frank conversations about the nature of the problem and steps to eradicate it may help to boost women's enthusiasm for a more equitable and meritocratic future STEM community.

## Conclusions

These results identify gender bias as one cause of women's underrepresentation in STEM. It does not appear to be the case that women are inherently disinterested in or prefer to avoid STEM. Instead, they "choose" to engage at lower rates than men only in the presence of systematic biases against them. The consequences of STEM gender biases are thus two-fold in that they can interrupt the progress of individual women in STEM who encounter them, as well as other women who may be (understandably) deterred from entering STEM in favor of more egalitarian educational and professional environments. Of importance, gender gaps in STEM engagement were fully eliminated when participants were told that bias was absent. This is heartening in that it reveals a road to gender parity; if future STEM communities exhibit gender equality instead of gender bias, then we have every reason to anticipate women's full participation.

**Acknowledgements** The authors thank Dr. Evava Pietri for statistical consultation. This research was funded in part by a research grant from the Smithsonian Institute and a faculty development grant from Skidmore College, both to the first author.

## Compliance with Ethical Standards

All data were collected in accordance with the highest established (APA) ethical standards (including obtaining informed consent from human participants). This work is not currently under review elsewhere, nor has it been previously published in whole or in part. This research was funded in part by a research grant from the Smithsonian Institute and a faculty development grant from Skidmore College, both to the first author.

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

Adams, G., Garcia, D. M., Purdie-Vaughns, V., & Steele, C. M. (2006). The detrimental effects of a suggestion of sexism in an instruction situation. *Journal of Experimental Social Psychology, 42*, 602–615. https://doi.org/10.1016/j.jesp.2005.10.004.

Anand, R., & Winters, M. F. (2008). A retrospective view of corporate diversity training from 1964 to the present. *The Academy of Management Learning and Education, 7*, 356–372. https://doi.org/10.5465/AMLE.2008.34251673.

Brady, L. M., Kaiser, C. R., Major, B., & Kirby, T. A. (2015). It's fair for us: Diversity structures cause women to legitimize discrimination. *Journal of Experimental Social Psychology, 57*, 100–110. https://doi.org/10.1016/j.jesp.2014.11.010.

Brown, E. R., Diekman, A. B., & Schneider, M. C. (2011). A change will do us good: Threats diminish typical preference for male leaders. *Personality and Social Psychology Bulletin, 37*, 930–941. https://doi.org/10.1177/0146167211403322.

Bruckmuller, S., & Branscombe, N. R. (2010). The glass cliff: When and why women are selected as leaders in crisis contexts. *British Journal of Social Psychology, 49*, 433–451. https://doi.org/10.1348/014466609X466S94.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science, 6*, 3–5. https://doi.org/10.1037/14805-009.

Carnes, M., Devine, P. G., Baier Manwell, L., Byars-Winston, A., Fine, E., Ford, C. E., … Sheridan, J. (2015). The effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic Medicine, 90,* 221–230. https://doi.org/10.1097/ACM.0000000000000552.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 3157–3162. Retrieved from http://www.jstor.org/stable/41060898.

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest, 15*, 75–141. https://doi.org/10.1177/1529100614541236.

Chaney, K. E., Sanchez, D. T., & Remedios, J. D. (2016). Organizational identity safety cue transfers. *Personality and Social Psychology Bulletin, 42*, 1564–1576. https://doi.org/10.1177/0146167216665096.

Cheryan, S., & Plaut, V. C. (2010). Explaining underrepresentation: A theory of precluded interest. *Sex Roles, 63*, 475–488. https://doi.org/10.1007/s11199-010-9835-x.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology, 97*, 1045–1060. https://doi.org/10.1037/a0016239.

Cheryan, S., Meltzoff, A. N., & Kim, S. (2011). Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. *Computers & Education, 57*, 1825–1835. https://doi.org/10.1016/j.compedu.2011.02.004.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143*, 1–35. https://doi.org/10.1037/bul0000052.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillside: Lawrence Erlbaum Associates.

Crocker, J., & Major, B. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological Review, 96,* 608–630. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.336.9967&rep=rep1&type=pdf.

Devine, P. G., Forscher, P. S., Cox, W. T. L., Kaatz, A., Sheridan, J., & Carnes, M. (2017). A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments. *Journal of Experimental Social Psychology, 73*, 211–215. https://doi.org/10.1016/j.jesp.2017.07.002.

Diekman, A. B., Clark, E. K., Johnston, A. M., Brown, E. R., & Steinberg, M. (2011). Malleability in communal goals and beliefs influence attraction to STEM careers: Evidence for a goal congruity perspective. *Journal of Personality and Social Psychology, 101*, 902–918. https://doi.org/10.1037/a0025199.

Diekman, A. B., Steinberg, M., Brown, E. R., Belanger, A. L., & Clark, E. K. (2017). A goal congruity model of role entry, engagement, and exit: Understanding communal goal processes in STEM gender gaps. *Personality and Social Psychology Review, 21*, 142–175. https://doi.org/10.1177/1088868316642141.

Dobbin, F., Schrage, D., & Kalev, A. (2015). Rage against the iron cage: The varied effects of bureaucratic personnel reforms on diversity. *American Sociological Review, 80*, 1014–1044. https://doi.org/10.1177/0003122415596416.

Files, J. A., Mayer, A. P., Ko, M. G., Friedrich, P., Jenkins, M., Bryan, M. J., … Hayes, S. N. (2017). Speaker introductions at internal medicine grand rounds: Forms of address reveal gender bias. *Journal of Women's Health, 26,* 413–419. https://doi.org/10.1089/jwh.2016.6044.

Green, T., & Kalev, A. (2008). Discrimination-reducing measures at the relational level. *Hastings Law Journal, 59*, 1435–1461. Retrieved from https://ssrn.com/abstract=1322504.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*, 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology, 89*, 416–427. https://doi.org/10.1037/0021-9010.89.3.416.

Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science, 314,* 599–600. Retrieved from http://www.jstor.org/stable/20031619.

Inzlicht, M., & Schmader, T. (Eds.). (2012). *Stereotype threat: Theory, process, and applications*. New York: Oxford University Press.

Kaiser, C. R., Major, B., Jurcevic, I., Dover, T. L., Brady, L. M., & Shapiro, J. R. (2013). Presumed fair: Ironic effects of organizational diversity structures. *Journal of Personality and Social Psychology, 104*, 504–519. https://doi.org/10.1037/a0030838.

Knobloch-Westerwick, S., Glynn, C. J., & Huge, M. (2013). The matilda effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication, 35*, 603–625. https://doi.org/10.1177/1075547012472684.

Lariviere, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature, 504*, 211–213. https://doi.org/10.1038/504211a.

Latu, I. M., Mast, M. S., & Stewart, T. L. (2015). Gender biases in (inter) action: The role of interviewers' and applicants' implicit and explicit stereotypes in predicting women's job interview outcomes. *Psychology of Women Quarterly, 39*, 539–522. https://doi.org/10.1177/0361684315577383.

Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages: How motivational interventions can reduce (but also increase) prejudice. *Psychological Science, 22*, 1472–1477. https://doi.org/10.1177/0956797611427918.

Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., von Hippel, W., & Bell, A. E. (2009). Interacting with sexist men triggers social identity threat among female engineers. *Journal of Personality and Social Psychology, 96*, 1089–1103. https://doi.org/10.1037/a0015703.

Major, B., Quinton, W. J., & Schmader, T. (2003). Attributions to discrimination and self-esteem: Impact of group identification and situational ambiguity. *Journal of Experimental Social Psychology, 39*, 220–231. https://doi.org/10.1016/S0022-1031(02)00547-4.

Midura, M. (2013). John vs. Jennifer: A battle of the sexes. *Yale Scientific.* Retrieved from http://www.yalescientific.org/2013/02/john-vs-jennifer-a-battle-of-the-sexes/.

Milkman, K. L., Akinola, M., & Chugh, D. (2012). Temporal distance and discrimination: An audit study in academia. *Psychological Science, 23*, 710–717. https://doi.org/10.1177/0956797611434539.

Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology, 100*, 1678–1712. https://doi.org/10.1037/apl0000022.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109,* 16474–16479. Retrieved from http://www.jstor.org/stable/41763373.

Moss-Racusin, C. A., van der Toorn, J., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2014). Scientific diversity interventions. *Science, 343*, 615–616. https://doi.org/10.1126/science.1245936.

Moss-Racusin, C. A., Pietri, E. V., Hennes, E. P., Dovidio, J. F., Brescoll, V. L., Roussos, G., … Handelsman, J. (2018). Reducing gender bias with VIDS (Video interventions for diversity in STEM). *Journal of Experimental Psychology: Applied*. Advance online publication. https://doi.org/10.1037/xap0000144.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18*, 879–885. https://doi.org/10.1111/j.1467-9280.2007.01995.x.

National Science Foundation, National Center for Science and Engineering Statistics (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017.* Special Report NSF 17–310. Arlington, VA. Retrieved from www.nsf.gov/statistics/wmpd/.

Nittrouer, C. L., Hebl, M. R., Ashburn-Nardo, L., Trump-Steele, R. C. E., Lane, D. M., & Valian, V. (2017). Gender disparities in colloquium speakers at top universities. *Proceedings of the National Academy of Sciences of the United States of America, 115*, 104–108. https://doi.org/10.1073/pnas.1708414115.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83,* 44–59. https://doi.org/10.1037//0022-3514.83.1.44.

Nosek, B. A., Smyth, F. L., Sriram, N., Linder, N. M., Devos, T., Ayala, A., … Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 10593-10597. https://doi.org/10.1073/pnas.0809921106.

Paluck, E. L. (2006). Diversity training and intergroup contact: A call to action research. *Journal of Social Issues, 62*, 577–595. https://doi.org/10.1111/j.1540-4560.2006.00474.x.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology, 60*, 339–367. https://doi.org/10.1146/annurev.psych.60.110707.163607.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical Turk. *Judgment and Decision Making, 5*, 411–419. Retrieved from https://ssrn.com/abstract=1626226.

Parker, L. R., Monteith, M. J., Moss-Racusin, C. A., & Van Camp, A. R. (2018). Promoting concern about gender bias with evidence-based confrontation. *Journal of Experimental Social Psychology, 74*, 8–23. https://doi.org/10.1016/j.jesp.2017.07.009.

Pietri, E. S., Moss-Racusin, C. A., Dovidio, J. F., Guha, D., Roussos, G., Brescoll, V. L., … Handelsman, J. (2017). Using video to increase gender bias literacy toward women in science. *Psychology of Women Quarterly, 41,* 175–196. https://doi.org/10.1177/0361684316674721.

Pollack, E. (2013, October 3). Why are there still so few women in science? *New York Times Magazine*. Retrieved from http://www.nytimes.com/2013/10/06/magazine/why-are-there-still-so-few-women-in-science.html.

Purdie-Vaughns, V., & Eibach, R. P. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles, 59*, 377–391. https://doi.org/10.1007/s11199-008-9424-4.

Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions. *Journal of Personality and Social Psychology, 94*, 615–630. https://doi.org/10.1037/0022-3514.94.4.615.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences, 111*, 4403–4408. https://doi.org/10.1073/pnas.1314788111.

Robnett, R. D. (2016). Gender bias in STEM fields: Variation in prevalence and links to STEM self-concept. *Psychology of Women Quarterly, 40*, 65–79. https://doi.org/10.1177/0361684315596162.

Sege, R., Nykiel-Bub, L., & Selk, S. (2015). Sex differences in institutional support for junior biomedical researchers. *JAMA, 314*, 1175–1177. https://doi.org/10.1001/jama.2015.8517.

Smith, J. L., Sansone, C., & White, P. H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology, 99*, 99–114. https://doi.org/10.1037/0022-0663.99.1.99.

Smith, J. L., Handley, I. M., Zale, A. V., Rushing, S., & Potvin, M. A. (2015). Now hiring! Empirically testing a three-step intervention to increase faculty gender diversity in STEM. *Bioscience, 65*, 1084–1087. https://doi.org/10.1093/biosci/biv138.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60*, 950–958. https://doi.org/10.1037/0003-066X.60.9.950.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28. https://doi.org/10.1006/jesp.1998.1373.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology, 69*, 797–811. https://doi.org/10.1037/0022-3514.69.5.797.

Steele, J., James, J. B., & Barnett, R. C. (2002). Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. *Psychology of Women Quarterly, 26*, 46–50. https://doi.org/10.1111/1471-6402.00042.

Steinpreis, R. E., Anders, K. A., & 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. https://doi.org/10.1023/A:1018839203698.

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology, 100*, 255–270. https://doi.org/10.1037/a0021385.

Thoman, D. B., & Sanson, C. (2016). Gender bias triggers diverging science interests between women and men: The role of activity interest appraisals. *Motivation and Emotion, 40*, 464–477. https://doi.org/10.1007/s11031-016-9550-1.

Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41*, 219–240. https://doi.org/10.1023/A:1018854212358.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science, 331*, 1447–1451. https://doi.org/10.1126/science.1198364.

Williams, W. M., & Ceci, S. J. (2012). When scientists choose motherhood. *American Scientist, 100*, 138–145. https://doi.org/10.1511/2012.95.138.

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences, 112*, 5360–5365. https://doi.org/10.1073/pnas.1418878112.

Williams, J. C., & Smith, J. L. (2015). The myth that academic science isn't biased against women. *The Chronicle of Higher Education.* Retrieved from http://www.chronicle.com/article/The-Myth-That-Academic-Science/231413/.

Williams, J. C., Phillips, K. W., & Hall, E. V. (2016). Tools for change: Boosting the retention of women in the STEM pipeline. *Journal of Research in Gender Studies, 1*, 11–75. https://doi.org/10.22381/JRGS6120161.

Wilton, L. S., Good, J. J., Moss-Racusin, C. A., & Sanchez, D. T. (2015). Communicating more than diversity: The effect of institutional diversity statements on expectations and performance as a function of race and gender. *Cultural Diversity and Ethnic Minority Psychology, 21*, 315–325. https://doi.org/10.1037/a0037883.

Yeager, D. S., Walton, G. M., Brady, S. T., Akcinar, E. N., Paunesku, D., Keane, L., … Dweck, C. S. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 3341–3348. https://doi.org/10.1073/pnas.1524360113.