Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology
Michael Brownstein and Jennifer Saul

Print publication date: 2016
Print ISBN-13: 9780198713241
Published to Oxford Scholarship Online: May 2016
DOI: 10.1093/acprof:oso/9780198713241.001.0001

Revisiting Current Causes of Women's Underrepresentation in Science

Carole J. Lee

DOI:10.1093/acprof:oso/9780198713241.003.0011

Abstract and Keywords

Given the body of experimental studies on gender bias in the evaluation of women in academia (e.g. Steinpreis, Anders, and Ritzke, 1999; Moss-Racusin et al., 2012), many expected implicit bias to be a major cause of women's underrepresentation in math-intensive sciences (STEM). However, large-scale correlational studies have discovered no gender disparities in real-life hiring and manuscript and grant outcomes (Ceci and Williams, 2011). Why might this be so? This chapter discusses methodological challenges that go beyond classic problems of external validity in extrapolating psychological effects and explanations to scientific communities. These problems include more complex external validity issues raised by the introduction of multi-process models of cognition (e.g. implicit versus explicit social cognition) as well as the reflexive role that folk and experimental theories of social psychology play in guiding the behavior of scientists at the individual and community level.

Keywords: implicit bias, gender, gender bias, STEM, science, external validity, manuscript review, grant review, hiring

What explains the continued underrepresentation of women in Science, Technology, Engineering, and Mathematics (STEM) disciplines? In a recent article in the *Proceedings of the National Academy of Sciences*, Stephen Ceci and Wendy Williams argue that claims of gender discrimination in journal review, grant funding, and hiring are "no longer valid" (Ceci and Williams, 2011). They conclude that

> the ongoing focus on sex discrimination in reviewing, interviewing, and hiring represents costly, misplaced effort: Society is engaged in the present in solving problems of the past rather than in addressing meaningful limitations deterring women's participation… today (Ceci and Williams, 2011).

The basis for this conclusion is what they deem a "more recent and robust empiricism" (Ceci and Williams, 2011) in particular, very large correlation studies of actual publication, grant award, and hiring outcomes that demonstrate no gender effect.

From the perspective of the robust social psychological literature on gender bias, this result may seem surprising. Experimental studies have added credence to the ecological validity and generalizability of implicit gender bias studies to STEM contexts by demonstrating that faculty in psychology (Steinpreis et al., 1999), biology, chemistry, and physics (Moss-Racusin et al., 2012) are susceptible to gender bias in hiring decisions. Furthermore, large meta-analyses measuring the predictive validity of Implicit Association Test measures on behavioral, judgment, and physiological test scores found small-to-moderate effect sizes, with correlation sizes ranging from 0.148 to 0.274 (Greenwald et al., 2009; (p.266) Oswald et al., 2013). Even when dealing with small effect sizes, we should expect to see differential outcomes (in this case, for women versus men) in the context of large samples/ populations (Greenwald et al., 2015).

How do we reconcile the substantial experimental evidence (which demonstrates implicit *gender bias*) with the large-scale correlational studies (which demonstrate no *gender effect*)?[1] How we answer this question has important social and institutional consequences. Rhetorically, Ceci and Williams's appeal to large empirical studies demonstrating no gender effect has the potential to be very damaging to the credibility of programs designed to recruit and retain women in STEM disciplines. After all, marshaling empirical evidence for gender bias played a central role in persuading faculty, administrators, and grant institutions about the need to undertake institutional changes.[2]

In this chapter I will try to reconcile these two literatures by motivating alternative (and not mutually exclusive) interpretations of the correlational studies that are consistent with the continued presence of implicit gender bias alongside women's reliance on strategies for counterbalancing such bias and the effectiveness of gender-equity programs to moderate its effects. In Section 1 I will argue that, despite their laudable efforts to control for quality indirectly through proxy, the correlational studies cited by Ceci and Williams cannot rule out the possibility that a quality confound is responsible for cancelling out the impact of ongoing implicit gender bias in journal review, grant funding, and hiring. Nor can the correlational studies rule out the possibility of a quality-related sample bias in the sample of women and men—a worry aggravated by the systematic underrepresentation of women among those submitting manuscripts (Lane and Linden, 2009), grant proposals (Grant et al., 1997; RAND, 2005; Marsh et al., 2008), and job applications (Committee on Gender Differences in the Careers of (p.267) Science et al., 2010). Studies in experimental psychology, by contrast, use stricter methods for controlling for quality and demonstrate clear gender bias.

In Section 2 I will identify how moderator variables present in STEM contexts may have contributed to the null results cited by Ceci and Williams by diminishing the influence of implicit gender bias in hiring. To motivate this possibility I will identify moderating factors that have been experimentally demonstrated to diminish the impact of implicit bias and whose presence in STEM contexts have been promoted by ADVANCE and other resource-intensive gender-equity efforts.

Please note that this chapter does not question Ceci and Williams's positive claims about additional causes of women's underrepresentation in STEM disciplines or about the relative strength of those other causes. Instead, my goal is to challenge their negative claim that gender discrimination is no longer a problem facing women scientists—that gender discrimination is a "historical rather than current" problem (Ceci and Williams, 2011)—by explaining how gender discrimination qua implicit gender bias can persist alongside counterbalancing and moderating factors.

1 Controlling for Quality in Correlational Studies

In the correlational studies, gender effects are discovered when the rate with which women garner successful outcomes is not proportionate to their representation in submission/application rates (Lee et al., 2013). Ceci and Williams rightly take the best studies to be those that control for quality-impacting factors such as author institution type (Xie and Shauman, 1998), experience (RAND, 2005), rank (Ley and Hamilton, 2008), discipline (RAND, 2005), and number of publications (Committee on Gender Differences in the Careers of Science et al., 2010). However, inferring lack of gender bias, as a causal claim, on the basis of these analyses requires adopting an important simplifying assumption: we must assume that, in the aggregate, women and men—with similar institutional resources, experience, rank, disciplines, and number of publications—submit manuscripts, grant applications, and job dossiers of comparable quality (Lee et al., 2013).

1.1 The sociality of peer review

However, the peer evaluation process is social in a very basic Weberian sense (Weber, 1947): actors make decisions (about, for example, the content or quality of work to submit) that are conditioned on beliefs about what others (e.g. reviewers and search committees) implicitly and explicitly believe and value (Lee et al., 2013). We can see this kind of social decision-making by manuscript (p.268) authors and grant proposal writers across the sciences. Among published studies across the physical, social, and human sciences, "positive" results favoring the experimental hypothesis are disproportionately valued and published (Fanelli, 2010). As a result, it should not be surprising that authors choose not to write up and submit studies reporting null results in preemptive anticipation of their perceived unimportance (Easterbrook et al., 1991) and likely rejection (Dickersin et al., 1992; Ioannidis, 1998). Authors also engage in research practices aimed at crafting positive results (Lee, 2013): a meta-analysis suggests that 33.7% of scientific researchers admit to using at least one questionable research practice in order to achieve a positive result (Fanelli, 2009), though estimates have surpassed 90% for specific disciplines (John et al., 2012). Analogous findings can be found in the literature on grant review. Applicants to the National Institutes of Health and National Science Foundation perceive the grant review process to be biased against highly innovative research (Gillespie et al., 1985; McCullough, 1989; National Research Council, 2007). Not surprisingly, then, grant applicants report downplaying the more transformative aspects of their research proposals (McCullough, 1989; Travis and Collins, 1991). It is "normal practice" for scientists to tune the content of their publication and grant submissions to forestall or defuse anticipated reviewer biases.

If women in STEM contexts anticipate gender bias in the evaluation of their manuscripts, grant proposals, and job applications, then we should expect that they try to offset this discrimination by submitting higher-quality work to counterbalance it (Gopnik, 2011; Budden et al., 2008b). Surveys support the idea that women believe that gender bias is a problem for them as individuals and as a group. A large survey of STEM researchers sponsored by the American Association for the Advancement of Science found that 62% of women scientists believed that gender biases were barriers faced by individuals working in the sciences (Cell Associates, 2010). 52% of women scientists reported having personally experienced gender bias (Cell Associates, 2010). A survey of female STEM faculty at the University of Michigan found that 42% of white women and 48% of women of color reported experiencing gender discrimination (UM ADVANCE Program, 2008). Surveys also suggest ways in which women strategize around anticipated gender bias. Even though single-anonymous journal review is the norm across the sciences (Ware, 2008), women prefer double-anonymous over single-anonymous journals (Budden et al., 2008a, Ware, 2008). A survey of women in physics revealed that women felt the need “to work twice as hard, do twice as much just to be considered half as qualified” (Ivie and Guo, 2006).

Even if women do not believe in the existence of gender bias (and do not strategically try to counterbalance it), they may have, through their education, (p.269) training, and professional experiences (Lee and Schunn, 2011), internalized stricter standards for what counts as quality work—a possibility indirectly evidenced by the relatively consistent and robust finding that female reviewers and editors are tougher than male ones in the evaluation of manuscripts (Wing et al., 2010; Lane and Linden, 2009; Gilbert et al., 1994) and grant proposals (Broder, 1993; Jayasinghe et al., 2003) authored by women or men.

1.2 Some of the fine print

The fine print of some of the large-scale studies cited by Ceci and Williams acknowledge the quality-related methodological limitations that qualify their results. The Faculty Committee on Women in Science, Engineering, and Medicine as well as NSF’s report on *Gender Differences in the Careers of Academic Scientists and Engineers* acknowledged that using “simple numerical counts” as measures for publishing success and productivity is problematic insofar as it accounts for “neither [the] quality nor the importance of scholarship” (Committee on Gender Differences in the Careers of Science et al., 2010; Bentley and Adamson, 2003). The large-scale study on grant review at NIH, NSF, and USDA made clear that “[i]f women are in fact less likely to apply for funding” as discovered across a range of large-scale studies on grant review (Grant et al., 1997; RAND, 2005; Marsh et al., 2008), then "female and male applicants for federal research grants likely differ in ways not observed in the data sets" (RAND, 2005). The Committee on Gender Differences in the Careers of Science notes that the slightly "higher rates of success" for female job candidates "may be explained by the possibility that only the strongest female candidates applied for Research I positions" (Committee on Gender Differences in the Careers of Science et al., 2010). So long as the correlational studies control for quality through proxy, this leaves open the possibility of a quality confound, where implicit gender bias co-exists alongside the submission of higher-quality work by women who anticipate and/or have been socialized to find strategies for overcoming implicit (or explicit) gender bias.

1.3 Controlling for quality in experimental studies

In light of the sociality of peer review, it is critical to control for the quality of submissions directly before drawing conclusions about the influence of gender on outcomes. The most convincing studies do this by submitting for evaluation work that is *identical with the exception of the gender of the author/applicant*. This is effectively what Steinpreis and her colleagues famously did: their study discovered that academic psychologists deemed a female job candidate as having less adequate teaching, research, and service than a male candidate with an (p.270) identical CV (Steinpreis et al., 1999). This result is reinforced by a more recent study showing that biology, chemistry, and physics faculty at research institutions deemed a female student less competent and hirable for a lab manager position than a male one with identical application materials, and offered the female applicant a 12% lower starting salary than the male applicant (Moss-Racusin et al., 2012).[3]

As Ceci and Williams note, ecologists sought to replicate this style of experimental research by having graduate students and postdocs review a paper that was identical save for the gender of the author (faculty were invited but failed to participate, making this study less generalizable to the STEM faculty population than the studies mentioned above). They discovered no gender effect (Borsuk et al., 2009). However, it is important to note that from a methodological point of view, this study used as their test manuscript a paper that had already been published (Borsuk et al., 2009) in a field that (at the time) enjoyed a rather generous 35–40% acceptance rate (Wardle, 2012). As I will explain, this choice of target manuscript runs the risk of being too clearly over the threshold for publishability to serve as a challenge to the experimental work on gender bias.

Within social psychology, the "convention established in classic experimental studies" is to use target materials for which it is ambiguous whether that target lies above or below the threshold for a particular attribute, such as publishability or hirability (Heilman et al., 2004; Biernat and Kobrynowicz, 1997). So, for example, Moss-Racusin and her colleagues elicited gender bias by crafting application materials for a student sufficiently "qualified to succeed in academic science" while not being "irrefutably excellent" (Moss-Racusin et al., 2012). Steinpreis and her colleagues' widely cited work elicited gender bias for an ordinary job applicant (Steinpreis et al., 1999).

We need not expect implicit gender bias in judgments about whether a target lies above some threshold in less ambiguous cases where the target clearly lies above the threshold, either because the target is exceptional or because the threshold is sufficiently low. So, for example, Steinpreis and her colleagues did not discover gender bias in judgments about the hirability of job candidates (p.271) whose credentials were exceptional and merited early tenure (Steinpreis et al., 1999).

1.4 Possible implications for interpreting the correlational studies

I propose that Borsuk and her colleagues did not find gender bias in the evaluation of the ecology manuscript because rejection rates at the time were sufficiently low for the already published manuscript to be well above the threshold for publishability. Contrast Borsuk's result with that of the classic Peters and Ceci study which found that an already published paper covertly resubmitted for review was more likely to be accepted for publication when the fictional author was affiliated with a prestigious institution than a non-prestigious one (Peters and Ceci, 1982). In this study, determining whether a strong manuscript exceeded the threshold for publishability was more difficult than in the Borsuk case, since the psychology journals under study had lower acceptance rates (about 20%, compared to Borsuk et al.'s 35–40%). Borsuk et al.'s result—what should be the most convincing result cited by Ceci and Williams with respect to journal reviewing because of its ability to control for quality directly rather than by proxy—is not inconsistent with the robust literature on gender bias in social psychology or with methodologically similar studies on stereotype-based evaluation bias.

A similar problem may afflict a correlational study that is frequently cited. Rebecca Blank randomly assigned manuscripts at *The American Economic Review* to single-anonymous and double-anonymous review conditions and discovered no statistically significant gender effects after controlling for author's institutional rank (Blank, 1991). However, Blank notes that double-anonymous review did not impact acceptance rates for authors at the highest and lowest ranking institutions and that the vast majority of submitting female authors were at lower-ranked universities. If we expect gender bias to play a larger role in ambiguous cases where a manuscript is not clearly above or below the threshold for publication, then the results of this study are not inconsistent with qualified psychological claims about seeing the disproportional effects of gender bias in near-threshold cases.

This observation animates yet another possible explanation for the null results cited by Ceci and Williams. It may be that in STEM contexts the distribution of quality among women's submissions is such that there are too few near-threshold cases for implicit gender bias to impact overall outcome measures; implicit gender bias persists, but impacts too few cases to skew the overall outcome.

Ideally, we would be able to compare all the above hypotheses (including Ceci and Williams's) by finding gender-independent ways of measuring the quality of submissions and individuals and identifying their distribution across the (p.272) population of women versus men in STEM disciplines.[4] In the absence of such techniques, we are left to control for quality through institutional/structural proxy (as the correlational studies do) or through stricter experimental means (as psychological studies do). Because of the inescapable sociality of peer review and the methodological advantage experimental studies have when controlling for quality, I give preferential weight to their results. However, I do not take the experimental results to be fully decisive, since the process of generalizing to STEM disciplines (and the distribution of quality found among women versus men there) requires its own set of inductive and interpretive inferences.

2 Moderators for Implicit Gender Bias

So far, my critiques have identified weaknesses in the correlational studies that may have systematically skewed the observed association between gender and outcomes towards the null result. This analysis motivated the possibility of an *unimpeded* psychological process between perceived gender (of author/applicant by evaluators) and outcomes, whose negative effects are not observed because they are cancelled out by counterbalancing gender differences in the quality of submitted work/applications. In contrast, this section will be dedicated to identifying moderating variables that reduce and even eliminate the influence of implicit gender bias by *impinging directly upon* the psychological process between (perceived) gender and outcomes.[5]

Psychologists have long recognized that moderating variables and competing processes present in real-life contexts may diminish or even eliminate a behavioral effect otherwise elicited in controlled laboratory conditions (Aronson et al., 1998). This appreciation has given rise to a crucial shift in questions about which properties from the laboratory are most important to project to the target context when evaluating the external validity of psychological studies: the primary object of generalization is no longer the stimulus–behavior pair (Campbell, 1957), but the *process* mediating the link between them and explaining their covariation in laboratory conditions (Mook, 1983). This means that failing to observe the expected behavior in the target context does not rule out the possibility that the cognitive process was triggered, carried out, and contributed causally toward (p.273) the behavior even in cases where moderating variables and competing processes tempered or even eliminated its effect. A null result in the target context does not imply the lack of a causal relationship in that context.

Since implicit biases are triggered automatically even in the absence of explicit attention being drawn to the stereotype-relevance of perceptual cues across a range of different perceptual modalities (Rudman and Lee, 2002, Dasgupta and Greenwald, 2001, Dasgupta and Asgari, 2004), I take the present question not to be whether the cognitive processes underwriting implicit gender bias play a causal role in STEM contexts, but what moderators may diminish their effect. In what follows, I will identify some of the moderating factors that have been found to diminish implicit bias's behavioral impact, and identify ways that resource-intensive gender-equity efforts and institutions have contributed to the presence of such moderators in the context of hiring.

2.1 Motivation

Motivation to avoid behaving with prejudice attenuates implicit bias (Maddux et al., 2005) and the influence of implicit bias on explicit social judgment tasks (Payne, 2005). Motivation to be similar to one's peers can decrease implicit bias and discriminatory behavior in contexts where one's peers are thought to be more egalitarian than one's self (Stewart et al., 2007b). And, for aversive racists (those with low explicit prejudice but high implicit bias), being reminded of one's previous discriminatory behavior can increase motivation to control prejudice and decrease prejudiced behavior (Son Hing et al., 2002).

How can institutions increase motivation and diminish implicit gender bias and prejudiced behavior in STEM contexts? Programs like ADVANCE can inspire and foster such motivation. If we infer that the institutionalization and endorsement of gender-equity programs such as ADVANCE change faculty perceptions about what their peers believe about women—especially by selecting highly credible senior male and female faculty to lead evaluation bias training programs (Stewart et al., 2007b)—then we have psychological reasons for thinking that this helps motivate faculty to control for implicit bias and decrease discriminatory behavior (Sechrist and Stangor, 2001). And, if ADVANCE-organized training about evaluation bias and reporting past hiring decisions reminds faculty with low explicit prejudice but high implicit bias about previous prejudicial behavior (such as gendered patterns of hiring in their department), we have psychological reason to think this will increase motivation to control prejudice and decrease discriminatory behavior (Son Hing et al., 2002).

Deans provide an additional source of motivation: the large hiring study cited by Ceci and Williams reported a "dean effect" in which women were more likely (p.274) to be offered a faculty position when the institution's dean reviewed and approved job offers as opposed to cases where the dean played no role (Committee on Gender Differences in the Careers of Science et al., 2010). The beneficial influence of upper administration has been noted at MIT (Hopkins, 2006), the University of Michigan (Stewart et al., 2007b), Georgia Tech (Fox et al., 2007), and Case Western Reserve University (Bilmoria et al., 2007).

As a locus of sustained change, diversity-dedicated programs such as ADVANCE may be more consistent over time than the dean effect. The dean effect is sensitive to the commitment of individual deans to gender equity and is, therefore, subject to variation across deans: for example, at MIT the number of women faculty hired in the School of Science increased sharply as a function of Dean Birgeneau's response to the 1996 Report on Women Faculty in Science but decreased when he left (Hopkins, 2006).[6] In contrast, we have reason to think that programs like ADVANCE and MIT's Gender Equity Committee (Hopkins, 2006) enjoy more stability over time by providing an ongoing organizational structure responsible for monitoring and improving the representation of female faculty in STEM disciplines.[7]

2.2 Cognitive capacity

When cognitive capacity has been hampered, due to the spontaneity of judgment (Hofmann et al., 2005), limited executive control (Payne, 2005), or low need for cognition (Florack et al., 2001), implicit bias proceeds along unhampered by the influence of more deliberate and careful processing. However, when attention and effort is increased, so is the moderating influence of explicit processing (Hofmann et al., 2005; Payne, 2005; Florack et al., 2001; Smith and DeCoster, 2000; Strack and Deutsch, 2004).

In an effort to increase attention and effort in hiring deliberations, ADVANCE programs have evaluation tools that focus on "specific and individuating evidence" and "emphasized clarity and completeness of candidates' contributions, mentoring and faculty development" (Fox et al., 2007). Twenty-one unique institutions have uploaded their best practices (including check lists) for search (p.275) committees on the ADVANCE web portal (ADVANCE). Since the institution of ADVANCE programs, the relative rate with which women have been hired has improved at the University of Michigan (Stewart et al., 2007b), Georgia Tech (Fox et al., 2007), New Mexico State University, and the University of California at Riverside, San Diego, and Irvine (ADVANCE, 2006). According to Beth Mitchneck, the program director for ADVANCE at the National Science Foundation, they hope to publish a more complete portfolio analysis in the future (Email, October 29, 2014). Note that reviving the institutional knowledge embodied by hiring evaluation tools with regular, systematic training/education requires sustained investment in institutional structures charged with implementing and enforcing these best practices.

2.3 Context effects

Experimental studies have demonstrated that contextual cues can attenuate implicit bias, as measured by decreased IAT scores (Barden et al., 2004; Wittenbrink et al., 2001): participants primed with violent, misogynist rap music showed more implicit bias against black men compared to those in the control group (Rudman and Lee, 2002); exposing participants to exemplars of famous women in leadership positions increased implicit associations between leadership qualities and women (Dasgupta and Asgari, 2004); and, exposure to admired black and disliked white individuals attenuates implicit bias against blacks while exposure to disliked black and admired white individuals increases it (Dasgupta and Greenwald, 2001).[8]

In light of this research, we would expect the presence of counterstereotypical female faculty on search committees to attenuate implicit gender bias. The large-scale hiring study cited by Ceci and Williams found that the odds that female job candidates get interviewed for a position improves with an increase in the percentage of female faculty on the search committee and when the search committee is chaired by a woman (Committee on Gender Differences in the Careers of Science et al., 2010). Note that putting women in leadership positions is a central goal of ADVANCE (National Science Foundation).

Ceci and Williams are right that gender-equity efforts focused on high-stakes gate-keeping moments are resource-intensive.[9] They are also right to worry (p.276) about *comparative* questions about the strongest factors responsible for the underrepresentation of women in STEM disciplines and to think about how equity-supporting resources should be allocated accordingly. However, if my analysis is correct, then taking away resources from current gender-equity efforts and institutions may erode the "fragility of progress" (Hopkins, 2006) achieved thus far.

3 Conclusion

Ceci and Williams argue that claims of gender discrimination in journal review, grant funding, and hiring are "no longer valid"—that worries about gender discrimination target "historical rather than current problems facing women scientists" (Ceci and Williams, 2011). I challenged this conclusion, arguing that null results in correlation studies may be consistent with the continued presence of implicit gender bias alongside women's reliance on strategies for counterbalancing such bias and the effectiveness of gender-equity programs to moderate its effects. I have not challenged

Ceci and Williams's positive claims about other causes of women's underrepresentation or the relative strength of those other causes. It is likely that there are multiple factors responsible for the underrepresentation of women in STEM disciplines, including the need to change disciplinary cultures and environments to increase women's feelings of belonging (Cheryan et al., 2009), the fact of women's underrepresentation itself which leads to women's lower self-identification with and motivation to pursue STEM careers (Murphy et al., 2007; Stout et al., 2011), as well as "gendered expectations, lifestyle choices, and career preferences" and "factors surrounding family formation and childrearing" (Ceci and Williams, 2011). However, if gender discrimination qua implicit gender bias persists, then withdrawing resources and institutions dedicated to monitoring and preventing gender discrimination in high-stakes gatekeeping moments may erode hard-earned progress.

Acknowledgments

I am grateful for the Career Enhancement Fellowship (funded by the Mellon Foundation and administered by the Woodrow Wilson Foundation) and Royalty Research Fund grant #A79071 (University of Washington) for the research time to complete this article. I would also like to thank Diana Bilimoria, Michael Brownstein, Sapna Cheryan, (p.277) Catherine Elgin, Jennifer Saul, Miriam Solomon, Abigail Stewart, Joyce Wen, and the anonymous reviewers for comments and/or references that helped to sharpen this chapter.

References

Bibliography references:

ADVANCE (online). "Search Committees." <http://www.portal.advance.vt.edu/index.php/tags/Search-Committees> (Accessed October 29, 2014).

ADVANCE (2006, online). "Using program evaluation to ensure the success of your advance program." <http://

www.advance.vt.edu/documents/other/advance_evaluation_toolkit.pdf> (Accessed October 29, 2014).

Aronson, E., Wilson, T. D., and Brewer, M. B. (1998). “Experimentation in social psychology.” In Gilbert, D., Fiske, S. and Lindzey, G. (eds.), *The Handbook of Social Psychology*. Boston, MA: McGraw-Hill.

Barden, J., Maddux, W. W., Petty, R. E., and Brewer, M. B. (2004). “Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes.” *Journal of Personality and Social Psychology* 87: 5–22.

Baron, R. M. and Kenny, D. A. (1986). “The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.” *Journal of Personality and Social Psychology* 51: 1173–82.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts.” *Memory and Cognition* 10: 82–93.

Baesalou, L. W. (1987). “The instability of graded structure: Implications for the nature of concepts.” In Neisser, U. (ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. New York, NY: Cambridge University Press.

Bentley, J. T. and Adamson, R. (2003). *Gender Differences in the Careers of Academic Scientists and Engineers: A Literature Review*. Arlington: National Science Foundation.

Biernat, M. and Kobrynowicz, D. (1997). “Gender- and race-based standards of competence: Lower minimum standards but higher ability standards for devalued groups.” *Journal of Personality and Social Psychology* 72: 544–57.

Bilmoria, D., Hopkins, M. M., O’Neil, D. A., and Perry, S. R. (2007). “Executive coaching: An effective strategy for faculty development.” In Stewart, A. J., Malley, J. E., and LaVaque-Manty, D. (eds.), *Transforming Science and Engineering:*

Advancing Academic Women. Ann Arbor, MI: University of Michigan Press.

Blank, R. M. (1991). "The effects of double-blind versus single-blind reviewing: Experimental evidence from *The American Economic Review*." *The American Economic Review* 81: 1041–68.

Bornmann, L., Mutz, R., and Daniel, H.-D. (2007). "Gender differences in grant peer review." *Journal of Informetrics* 1: 226–38.

Borsuk, R. M., Aarssen, L. W., Budden, A. E., Koricheva, J., Leimu, R., Tregenza, T., and Lortie, C. J. (2009). "To name or not to name: The effect of changing author gender on peer review." *BioScience* 59: 985–9.

Brewer, M. B. (2000). "Research design and issues of validity." In Reis, H. T. and Judd, C. M. (eds.), *Handbook of Research Methods in Social and Personality Psychology*. New York, NY: Cambridge University Press.

(p.278) Broder, I. E. (1993). "Review of NSF economics proposals: Gender and institutional patterns." *The American Economic Review* 83: 964–70.

Budden, A. E., Aarssen, L. W., Koricheva, J., Leimu, R., Lortie, C. J., and Tregenza, T. (2008a). Does double-blind review favor female authors? Reply." *Frontiers in Ecology and the Environment* 6: 356–7.

Budden, A. E., Aarssen, L. W., Koricheva, J., Leimu, R., Lortie, C. J., and Tregenza, T. (2008b). "Response to Whittaker: Challenges in testing for gender bias." *Trends in Ecology and Evolution* 23: 480–1.

Campbell, D. T. (1957). "Factors relevant to the validity of experiments in social settings." *Psychological Bulletin* 54: 297–312.

Ceci, S. J. and Willliams, W. M. (2011). "Understanding current causes of women's underrepresentation in science."

Proceedings of the National Academy of Sciences 108: 3157–62.

Cell Associates (2010). *Barriers for Women Scientists Survey Report*. American Association for the Advancement of Science.

Cheryan, S., Plaut, V. C., Davies, P. G., and Steele, C. M. (2009). “Ambient belonging: How stereotypical cues impact gender participation in computer science.” *Journal of Personality and Social Psychology* 97: 1045–60.

Committee on Gender Differences in the Careers of Science, Engineering, and Mathematics Faculty; Committee on Women in Science, Engineering, and Medicine; Committee on National Statistics; National Research Council (2010). *Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty*. Washington: National Academies Press.

Dasgupta, N. and Asgari, S. (2004). “Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping.” *Journal of Experimental Social Psychology* 40: 642–58.

Dasgupta, N. and Greenwals, A. G. (2001). “On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals.” *Journal of Personality and Social Psychology* 81: 800–14.

Dickersin, K., Min, Y.-I., and Meinert, C. L. (1992). “Factors influencing publication of research results: Follow-up of applications submitted to two institutional review boards.” *Journal of the American Medical Association* 267: 374–9.

Easterbrook, P. J., Berlin, J. A., Gopalan, R., and Matthews, D. R. (1991). “Publication bias in clinical research.” *The Lancet* 337: 867–72.

Fanelli, D. (2009). “How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data.” *PLoS One* 4: 1–11.

Fanelli, D. (2010). "'Positive' results increase down the hierarchy of the sciences." *PLoS One* 5: 1–10.

Florack, A., Scarabis, M., and Bless, H. (2001). "When do associations matter? The use of automatic associations toward ethnic groups in person judgments." *Journal of Experimental Social Psychology* 37: 518–24.

Fox, M. F., Colatrella, C., McDowell, D., and Realff, M. L. (2007). "Equity in tenure and promotion: An integrated institutional approach." In Stewart, A. J., Malley, J. E., and LaVaque-Manty, D. (eds.), *Transforming Science and Engineering: Advancing Academic Women*. Ann Arbor, MI: University of Michigan Press.

(p.279) Gawronski, B. and Bodenhausen, G. V. (2006). "Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change." *Psychological Bulletin* 132: 692–731.

Gilbert, J. R., Williams, E. S., and Lundbert, G. D. (1994). "Is there gender bias in JAMA's peer review process?" *Journal of the American Medical Association* 272: 139–42.

Gillespie, G. W., Chubin, D. E., and Kurzon, G. M. (1985). "Experience with NIH peer review: Researchers' cynicism and desire for change." *Science, Technology, and Human Values* 10: 44–54.

Ginther, D. K. (2003). "Is MIT an exception? Gender pay differences in academic science." *Bulletin of Science, Technology and Society* 23: 21–6.

Ginther, D. K. (2004). "Why women earn less: Economic explanations for the gender salary gap in science." *AWIS Magazine* 33: 1–5.

Ginther, D. K. and Kahn, S. (2009). "Does science promote women? Evidence from academia 1973–2001." In Freeman, R. B. and Gorgoff, D. L. (eds.), *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*. Chicago, IL: University of Chicago Press.

Gopnik, A. (2011, online). “What John Tierney gets wrong about women scientists: Understanding a new study about discrimination.” <http://www.slate.com/id/2285355/>.

Grant, J., Burden, S., and Breen, G. (1997). “No evidence of sexism in peer review.” *Nature* 390: 438.

Greenwald, A. G., Banaji, M. R., and Nosek, B. A. (2015). “Statistically small effects of the Implicit Association Test can have societally large effects.” *Journal of Personality and Social Psychology* 108: 553–61.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). “Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity.” *Journal of Personality and Social Psychology* 97(1): 17–41.

Heilman, M. E., Wallen, A. S., Fuchs, D., and Tamkins, M. M. (2004). “Penalties for success: Reactions to women who succeed at male gender-typed tasks.” *Journal of Applied Psychology* 89: 416–27.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., and Schmitt, M. (2005). “A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measure.” *Personality and Social Psychology Bulletin* 31: 1369–85.

Hopkins, N. (2006). “Diversification of a university faculty: Observations on hiring women faculty in the schools of science and engineering at MIT.” *MIT Faculty Newsletter* XVIII(1): 16–23.

Ioannidis, J. P. A. (1998). “Effect of the statistical signifcance of results on the time to completion and publication of randomized efficacy trials.” *Journal of the American Medical Association* 279: 281–6.

Ivie, R. and Guo, S. (2006). “Women physicists speak again.” American Institute of Physics, Pub. Number R-441.

Jayasinghe, U. W., Marsh, H. W., and Bond, N. (2003). “A multilevel cross-classified modelling approach to peer review of grant proposals: The effects of assessor and researcher attributes on assessor ratings." *Journal of the Royal Statistical Society (A)* 166: 279–300.

(p.280) John, L. K., Loewenstein, G., and Prelec, D. (2012). "Measuring the prevalence of questionable research practices with incentives for truth telling." *Perspectives on Psychological Science* 23: 524–32.

Kalev, A., Dobbin, F., and Kelly, E. (2006). "Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies." *American Sociological Review* 71: 589–617.

Lane, J. A. and Linden, D. J. (2009). "Is there gender bias in the peer review process at *Journal of Neurophysiology*?" *Journal of Neurophysiology* 101: 2195–6.

Larivière, V., Ni, C., Gingras, Y., Cronin, B., and Sugimoto, C. (2013). "Bibliometrics: Global gender disparities in science." *Nature* 504: 211–13.

LaVaque-Manty, D. (2007). "Transforming the Scientific Enterprise: An interview with Alice Hogan." In Steward, A. J., Malley, J. E., and LaVaque-Manty, D. (eds.), *Transforming Science and Engineering: Advancing Academic Women*. Ann Arbor, MI: University of Michigan Press.

Lee, C. J. (2013). "The limited effectiveness of prestige as an intervention on the health of medical journal publications." *Episteme* 10: 387–402.

Lee, C. J. and Schunn, C. D. (2011). "Social biases and solutions for procedural objectivity." *Hypatia* 26: 352–73.

Lee, C. J., Sugimoto, C. R., Zhang, G., and Ceonin, B. (2013). "Bias in peer review." *Journal of the American Society for Information Science* 64: 2–17.

Ley, T. J. and Hamilton, B. H. (2008). "The gender gap in NIH grant applications." *Science* 322: 1472–4.

Maddux, W. W., Jamie, B., Brewer, M. B., and Petty, R. E. (2005). "Saying no to negativity: The effects of context and motivation to control prejudice on automatic evaluative responses." *Journal of Experimental Social Psychology* 41: 19–35.

Marsh, H. W., Jayasinghe, U. W., and Bond, N. W. (2008). "Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability." *American Psychologist* 63: 160–8.

Massachusetts Institute of Technology (1999). "A Study on the status of women faculty in science at MIT." *The MIT Faculty Newsletter*. Cambridge, MA: Massachusetts Institute of Technology.

McCullough, J. (1989). "First comprehensive survey of NSF applicants focuses on their concerns about proposal review." *Science, Technology, and Human Values* 14: 78–88.

Mitchneck, B. (2014). Email, October 29, 2014.

Mook, D. G. (1983). "In defense of external validity." *American Psychologist* 38: 379–87.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., and Handelsman, J. (2012). "Science faculty's subtle gender biases favor male students." *Proceedings of the National Academy of Sciences* 109: 16474–9.

Murphy, M. C., Steele, C. M., and Gross, J. J. (2007). "Signaling threat: How situational cues affect women in math, science, and engineering settings." *Psychological Science* 18: 879–85.

National Research Council (2007). *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. Washington: National Academies Press.

(p.281) National Science Foundation (online). "ADVANCE at a glance." <http://www.nsf.gov/crssprgm/advance/> (Accessed August 29, 2013).

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., and Tetlock, P. E. (2013). "Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies." *Journal of Personality and Social Psychology* 105: 171–92.

Payne, B. K. (2005). "Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping." *Journal of Personality and Social Psychology* 89: 488–503.

Peters, D. P. and Ceci, S. J. (1982). "Peer-review practices of psychological journals: The fate of published articles, submitted again." *Behavioral and Brain Sciences* 5: 187–255.

RAND Corporation (2005). "Is there gender bias in federal grant programs?" Santa Monica: RAND Corporation.

Rudman, L. A. and Lee, M. R. (2002). "Implicit and explicit consequences of exposure to violent and misogynous rap music." *Group Processes and Intergroup Relations* 5: 133–50.

Sechrist, G. B. and Stangor, C. (2001). "Perceived consensus influences intergroup behavior and stereotype accessibility." *Journal of Personality and Social Psychology* 80: 645–54.

Smith, E. R. and Decoster, J. (2000). "Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review* 4: 108–31.

Son Hing, L. S., Li, W., and Zanna, M. P. (2002). "Inducing hypocrisy to reduce prejudicial response among aversive racists." *Journal of Experimental Social Psychology* 38: 71–7.

Steinpreis, R. E., Anders, K. A., and 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: 509–28.

Stewart, A. J., Malley, J. E., and LaVaque-Manty, D. (2007a). "Analyzing the problem of women in science and engineering: Why do we need institutional transformation?" In Stewart, A. J., Malley, J. E., and LaVaque-Manty, D. (eds.), *Transforming*

Science and Engineering: Advancing Academic Women. Ann Arbor, MI: University of Michigan Press.

Stewart, A. J., Malley, J. E., and LaVaque-Manty, D. (2007b). “Faculty recruitment: Mobilizing science and engineering faculty.” In Stewart, A. J., Malley, J. E., and LaVaque-Manty, D. (eds.), *Transforming Science and Engineering: Advancing Academic Women*. Ann Arbor, MI: University of Michigan Press.

Stout, J. G., Dasgupta, N., Hunsinger, M., and 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: 255–70.

Strack, F. and Deutsch, R. (2004). “Reflective and impulsive determinants of social behavior.” *Personality and Social Psychology Review* 8: 220–47.

Travis, G. D. L. and Collins, H. M. (1991). “New light on old boys: Cognitive and institutional particularism in the peer review system.” *Science, Technology, and Human Values* 16: 322–41.

(p.282) UM ADVANCE Program (2008). “Assessing the academic work environment for science and engineering and social science faculty at the University of Michigan in 2006: Gender, race, and discipline in department- and university-related climate factors.” <http://advance.umich.edu/resources/ADVANCE-2006-R2-FullReport.pdf> (accessed September 28, 2015).

Valian, V. (1998). *Why So Slow? The Advancement of Women*. Cambridge, MA: MIT Press.

Wardle, D. A. (2012). “On plummeting manuscript acceptance rates by the main ecological journals and the progress of ecology.” *Ideas in Ecology and Evolution* 5: 13–15.

Ware, M. (2008). “Peer review in scholarly journals.” *PRC Summary Papers* 4. London: Publishing Research Consortium.

Weber, M. (1947). *The Theory of Social and Economic Organization*, New York: Oxford University Press.

Wing, D. A., Benner, R. S., Petersen, R., Newcomb, R., and Scott, J. R. (2010). "Differences in editorial board reviewer behavior based on gender." *Journal of Women's Health* 19: 1919–23.

Wittenbrink, B., Judd, C. M., and Paek, B. (2001). "Spontaneous prejudice in context: Variability in automatically activated attitudes." *Journal of Personality and Social Psychology* 81: 815–27.

Xie, Y. and Shauman, K. A. (1998). "Sex differences in research productivity: New evidence about an old puzzle." *American Sociological Review* 63: 847–70.

Notes:

([1]) For the sake of clarity I will distinguish questions about gender bias from questions about gender effects. By convention, I will refer to "gender bias" as a psychological disposition/tendency, underwritten by cognitive representations and processes, that is studied experimentally. In contrast, I will refer to "gender effects" as differential outcomes for men versus women in "natural" settings, where possible confounds are controlled for by proxy rather than through direct experimental means (Bornmann et al., 2007). I will distinguish "effects" (i.e. the outcomes of causal processes) *simpliciter* from "gender effects" in particular by marking *gender* effects as such.

([2]) Virginia Valian's classic (1998) compellingly synthesized psychological work on cognitive schemas and evaluation bias and helped to catalyze the creation of the National Science Foundation's ADVANCE initiative (Stewart et al., 2007a; LaVaque-Manty, 2007). MIT's 1999 Report on Women in Science (Massachusetts Institute of Technology, 1999), which revealed unequivocal gender differences in resource allocation among faculty, also reverberated throughout academe, leading to national media attention as well as internal (Hopkins, 2006) and external (Ginther, 2003) examinations of hiring, tenure, and resource allocation at other institutions.

([3]) A note about STEM salaries: Ceci and Williams cite work by Ginther and Hayes to support their claim that salary disparities by gender are fully attributable to other factors. However, the cited author's work suggests a more complex picture in which there is no statistically significant gender gap except among full professors in science: men earned 12% more than female full professors, where a third of the salary gap is not explained by observable, non-gender factors (Ginther, 2004; Ginther and Kahn, 2009). The cited authors go on to suggest that gender factors—in particular, the accumulation of disadvantage by women versus men—may be to blame for this discrepancy.

([4]) Controlling for quality via number of publications or citations is not, for example, gender independent, since women publish fewer articles and are cited less frequently than men (Larivière et al., 2013).

([5]) Moderator variables diminish or enhance the degree of the final effect in the presence of the cause (Baron and Kenny, 1986). Unlike mediator variables, moderator variables are not thought to be necessary to complete the process connecting the cause to the effect (Brewer, 2000).

([6]) The increase in female hires in the sciences happened despite no change in the percentage of women completing PhDs at MIT, where the institution has the unusual practice of hiring its own graduates (Hopkins, 2006).

([7]) Along these lines, a large-scale correlational study found that when it comes to improving diversity among managers employed in the private sector, the presence of organizational structures responsible for increasing diversity (such as diversity officers, committees, departments, and task forces) is more effective than diversity training, evaluation, mentoring, or networking. The same study found that the presence of diversity-related organizational structures enhances the effectiveness of the other initiatives (Kalev et al., 2006).

([8]) Contexts are thought to moderate IAT scores through the temporary activation of select characteristics/properties associated with the concept in question (Barden et al., 2004, Barsalou, 1982, 1987, Wittenbrink et al., 2001, Gawronski and Bodenhausen, 2006).

([9]) The National Science Foundation's ADVANCE initiative has invested more than $130 million in more than a hundred institutions of higher education since 2001 to support the hiring, retention, and advancement of women in STEM disciplines (National Science Foundation, 2013); and individual institutions and investigators have invested additional resources dedicated to assessing gender differences in resource allocation and outcomes (Massachusetts Institute of Technology, 1999; Ginther, 2003).