

Threats to objectivity in peer review: the case of gender

Anna Kaatz¹, Belinda Gutierrez^{1,2}, and Molly Carnes^{1,2,3,4}

¹ Center for Women's Health Research University of Wisconsin–Madison, Madison, WI 53715, USA

² William S. Middleton Memorial Veterans Hospital, Geriatric Research Education and Clinical Center, Madison, WI, USA

³ Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI 53715, USA

⁴ Departments of Medicine and Psychiatry, University of Wisconsin–Madison, Madison, WI 53715, USA

Scientists strive to be objective in their peer review of grant applications and manuscript submissions. Nevertheless, all humans are susceptible to biases in decision-making. To illustrate how cognitive bias unrelated to the merit of the science could influence scientific peer review we describe the potential impact of applicant gender on the judgment of reviewers. [Table 1](#) describes some different types of cognitive biases and [Table 2](#) describes conditions that might facilitate the operation of cognitive biases in peer review.

Despite the large number of women who have entered science since the passage by the US government, in 1972, of Title IX legislation prohibiting gender discrimination in education, science remains a male gender-typed domain. A substantive body of theoretically sound, experimentally rigorous research demonstrates that pervasive cultural stereotypes depicting women as communal (e.g., dependent, relationship-oriented) and men as agentic (e.g., logical, goal-oriented) lead to the implicit assumption that women are less competent than men in male gender-typed domains, where competence is associated with agentic traits [1]. Although explicit gender bias still exists, implicit (unconscious) gender bias is particularly troubling because it can lead well-intentioned people to unwittingly perpetuate inequalities.

Research confirms the pervasive implicit association of science with male. In one study 70% of ~300 000 men and women, participants linked science words more quickly with male than female names in a timed dual-categorization task (the Implicit Association Test) that bypasses conscious decision-making [2]. Furthermore, gender stereotypes cause subtle adjustments of performance expectations in a science setting. In one experiment, male and female faculty in science departments, randomized to evaluate a male or female applicant for a laboratory manager position, preferred to hire the male applicant, pay him a higher salary, and provide him with more career mentorship than the identical female applicant – whom they nevertheless deemed more likeable [3]. Congruent with previous research, the evaluative decisions in this study were driven primarily by the implicit assumption that the male applicant was more competent in the male gender-typed domain of science.

Further experiments find that, before women are deemed equivalently hireable, competent, or worthy of promotion in male gender-typed professions, they must demonstrate a higher level of achievement than identically credentialed men [4]. Paradoxically, raters may enthusiastically praise a woman's performance at an agentic task, even though she may not reach the higher standard needed to confirm competence, because she is exceeding the lower performance expectations 'for a woman' [5]. Unfortunately, laboratory and field studies confirm that such praise may not translate into the same rewards or promotions for women as it does for men [1,4]. How might this research be relevant to scientific peer review?

When examined by age group, individuals of the same age as most scientific reviewers (i.e., 40–65 years) held the strongest implicit association of male and science on the Implicit Association Test [2]. Because science is a male gender-typed domain, evaluators of individual scientists filter information through the type of cognitive framework that is predicted to advantage male and disadvantage female applicants. Consequently, one would predict that reviewers require more proof of ability (e.g., more papers, more prestigious awards) to confirm the competence of a female than a male scientist. Leadership is also a male gender-typed domain [6], thus being evaluated for performance as a scientific leader would set the cognitive stage for a further burden of proof that a woman's competence is equivalent to a man's. This subtle adjustment of performance standards generally occurs inadvertently, unintentionally, and is usually in conflict with both data (e.g., men and women are equally effective leaders [7]) and explicit beliefs about the abilities of men and women [2]. Although randomized, controlled experiments confirm a negative bias against women in mock hiring settings for male-gender typed employment [8], no such studies have similarly tested the extent to which women are disadvantaged in scientific peer review. If such bias in the evaluation of women scientists and their work were occurring, we would expect to see that manuscripts written by women scientists are less likely to be accepted for publication than those written by men, and that women applicants for scientific grants are less successful than their male counterparts. Is this the case?

Goldberg published the first experimental study assessing the effect of manipulating author gender of scientific papers by randomly assigning students to read the same manuscript written by an author with either the first name

Corresponding author: Carnes, M. (mlcarnes@wisc.edu).

0165-6147/

© 2014 Published by Elsevier Ltd. <http://dx.doi.org/10.1016/j.tips.2014.06.005>

Table 1. Some types of cognitive bias and potential relevance to scientific peer review

Cognitive bias	Relevance to peer review
<i>Anchoring bias</i> : over-reliance on one piece of information (the anchor) or a ‘first impression’ when adjusting one’s opinion in response to subsequent information.	If the reviewer likes the lab where the principal investigator (PI) trained or, for example, does not like the particular animal model being used, this will affect the processing of additional information in the proposal.
<i>Backfire effect</i> : instead of adjusting an opinion in response to disconfirming information, one more strongly adheres to the original opinion.	A reviewer might hold to an initial opinion or score despite compelling information to the contrary.
<i>Blind-spot bias</i> : the ability to see cognitive bias in others but not in oneself.	A reviewer is sincere in their belief that they are objective and without bias.
<i>Confirmation bias</i> : information is selectively marshaled to find support for an existing impression or assumption.	If a reviewer rates a proposal highly, weaknesses might be overlooked.
<i>Ecological fallacy</i> : making an inference about an individual based on aggregate data or assumptions about a group; making assumptions about individuals based on cultural stereotypes is a type of ecological fallacy.	Because fewer women lead large center grants, making the assumption that an individual woman is less able to lead a large program; because overall Black student achievement is lower than for White students, making the assumption that an individual Black student is less prepared.
<i>Halo effect</i> : assuming, because someone is competent (or incompetent) in one area, that they are competent (or incompetent) in other areas. Individuals in high-status groups are assumed to be competent across a range of activities.	May result in more ‘benefit of the doubt’ for a PI at a prestigious institution and the need for a higher level of proof of competence for a PI at a lower-ranked school; could also lead to more influence of a statement by a high-versus lower-status individual in a study section.
<i>In group/out group bias</i> : it takes less information for positive judgments when the rater and target share even minimal affinity than when they do not, and less information is required for a negative judgment when they share no affinity.	The alignment of discipline, social networks, previous interactions, age, gender, or race between reviewer and R01 applicant or author could influence the review of the grant or manuscript positively or negatively.
<i>Shifting standards of reference</i> : cultural stereotypes set different mental frameworks for judging the work of individuals. For example, verbal skills were rated lower if raters thought an author was Black versus White. When the stereotype is of lower competence (e.g., women in mathematics, Blacks in academia), the performance needed for judgment of minimal competence is lower (e.g., good in mathematics – for a woman) but higher to confirm competence (i.e., there is empirical support for the adage that a member of a low-status group ‘needs to be twice as good’ to get ahead).	It is possible that grants or manuscripts will be perceived as being less well written if the reviewer knows the PI is Black. The greatest differences in funding outcomes at NIH for low- versus high-status groups is not at the T- or K-award level (minimal standard) but at the R01 level (confirmatory standard): this is what the shifting-standards type of implicit bias would predict. It is also possible that highly positive descriptors for a low-status PI do not translate into a fundable score and that more negative descriptors for a high-status PI would not prevent a fundable score.

‘John’ or ‘Joan’ [9]. Participants, all of whom were female students, consistently gave higher ratings to the manuscripts written by John than to the same paper written by Joan. Although field studies of gender bias in scientific publication have given mixed results, and no firm conclusions could be drawn (e.g., [10]), women are generally found to have fewer but overall higher-quality peer-reviewed scientific publications than men [11]. Broadly, this aligns with research showing that women face a disadvantage and can be held to a higher performance standard than men in review processes in male-typed fields.

Given the impact of gender stereotypes on judgment, one would also predict that grant reviewers may unknowingly and unintentionally adjust competence standards such that a female applicant would need to demonstrate more proof of ability to earn the same assessment of competence as a male applicant. Lending support to this is the classic study by Weneras and Wold in which female

applicants for a postdoctoral research fellowship needed more than twice as many publications to receive the same competence scores as comparable male applicants [12]. In the USA, awards from the National Institutes of Health (NIH) are the cornerstone of grant support for individual investigators in biomedical, behavioral, and clinical research. Overall, there is little difference in the success rates of male and female applicants for NIH grants [13–15]. If gender bias operates in grant peer review, one would expect that different evaluative standards for male and female investigators would be exaggerated when assumptions of performance most strongly align with male gender stereotypes; that is, where science and leadership conflate. This would occur in the review of highly prestigious scientific awards, large program proposals where the investigator is a leader of other scientific leaders, and R01 (NIH grant support for human health-related research) renewals from experienced investigators. Gender differences are

Table 2. Bias-enhancing conditions surrounding scientific peer review

Condition promoting cognitive bias	Potential relevance to R01 review
<i>Time pressure</i> : because cognitive biases are efficient, time pressure promotes their influence on decision-making.	Reviewers have multiple demands on their time, and are frequently under time-constraints to finish reviews.
<i>Belief in one’s personal objectivity</i> : in a constructed hiring setting, raters who were primed to believe they were objective gave more biased ratings than did non-primed controls.	Being an NIH scientific reviewer may prime belief in one’s objectivity as a scientist.
<i>Semantic gender priming</i> : exposure to words more strongly associated with male (e.g., aggressive, competitive) or female (e.g., supportive, nurturing) stereotypes affects subsequent evaluation of male or female targets.	Emphasis on funding scientists willing to engage in ‘risk-taking’ or achieve ‘technological breakthrough’ would be predicted to enhance evaluations of male applicants.

found in these situations [13]. Between 2003 and 2007, experienced female investigators had significantly lower R01 funding success rates than did equivalent male applicants [14]. Similarly, in 2008, experienced female investigators who submitted Type 2 (renewal) R01 applications had significantly lower funding success rates than did equivalent male applicants [15]. Data from the NIH website show persistently lower Type 2 R01 or R01-equivalent success rates for women versus men, and there was no change after the 2009 change in scoring and critique structure (range, 2–8% lower; average, 5%) [13]. We examined written critiques of R01 proposals and found significant differences in those for applications submitted by male versus female investigators, despite similar scores or funding outcomes [16]. Consistent with implicitly lower performance expectations for women in science, peer reviewers offered more praise and acclamation for proposals from female investigators, and greater reference to competence and ability for funded proposals from experienced female investigators (i.e., this is amazing work for a woman scientist; she must have exceptional ability). Critiques for male applicants contained significantly more negative words. This is also consistent with stereotype-based assumptions causing a subtle adjustment of evaluative standards because it may require more proof of a man's lack of competence in a male gender-typed domain for him to be deemed incompetent [1]. Although this is not an experimental study, the differences are those expected if gender stereotypes lead R01 reviewers to subtly adjust performance standards.

After reviewing published research on grant peer review, the Cochrane Collaboration concluded that 'Experimental studies assessing the effects of grant giving peer review on ... funded research are urgently needed' and that 'Practices aimed to control and evaluate the potentially negative effects of peer review should be implemented...' [17]. We agree. Extrapolating from experimental research on interventions that can reduce gender bias in other evaluative contexts (reviewed in [8,18]), we recommend the following in the setting of scientific review for manuscripts or grants:

- (i) Instead of feeling confident in your objectivity, before engaging in peer review reflect on the susceptibility of all humans to bias in judgment.
- (ii) Allow sufficient time and try to avoid 'multi-tasking' when reviewing a scientific work.
- (iii) Before engaging in peer review, imagine in detail a female scientific leader.
- (iv) As far as possible, undertake to review the assessment criteria before evaluating manuscripts or applications.
- (v) Challenge yourself with thought experiments: would your evaluation change if the investigator were of a different gender (or race, or from a different institution)?
- (vi) Journals and funding agencies should make sure that review criteria do not implicitly prime the selection of

male over female scientists (e.g., avoid the use of male-typed terms such as 'high-risk' or 'aggressive').

Women have lower rates of publication and lower success rates for high-status research awards than do men. Theoretically supported experimental and field research make a strong case that gender stereotypes operate in review processes in male-typed domains such as science. Future experimental studies will be necessary to examine the extent to which gender bias may operate in peer review despite the best intentions of reviewers, their personal commitment to egalitarian principles, and their goal of advancing the best and most innovative science.

Acknowledgments

This work was supported by NIH grant number R01 GM111002.

References

- 1 Biernat, M. (2012) Stereotypes and shifting standards: forming, communicating and translating person impressions. In *Advances in Experimental Social Psychology* (Devine, P., ed.), pp. 1–50, Academic Press
- 2 Nosek, B.A. et al. (2007) Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* 18, 36–88
- 3 Moss-Racusin, C.A. et al. (2012) Science faculty's subtle gender biases favor male students. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16474–16479
- 4 Heilman, M.E. and Haynes, M.C. (2008) Subjectivity in the appraisal process: a facilitator of gender bias in work settings. In *Beyond Common Sense: Psychological Science in the Courtroom* (Borgida, E. and Fiske, S.T., eds), pp. 127–155, Blackwell Publishing
- 5 Biernat, M. and Vescio, T.K. (2002) She swings, she hits, she's great, she's benched: implications of gender-based shifting standards for judgment and behavior. *Pers. Soc. Psychol. Bull.* 28, 66–77
- 6 Eagly, A.H. and Karau, S.J. (2002) Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109, 573–598
- 7 Eagly, A.H. et al. (2003) Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. *Psychol. Bull.* 129, 569–591
- 8 Isaac, C. et al. (2009) Interventions that affect gender bias in hiring: a systematic review. *Acad. Med.* 84, 1440–1446
- 9 Goldberg, P. (1968) Are women prejudiced against women? *Transaction* 5, 316–322
- 10 Nakhaie, M.R. (2002) Gender differences in publication among university professors in Canada. *Can. Rev. Sociol. Anthropol.* 39, 151–179
- 11 Symonds, M.R.E. et al. (2006) Gender differences in publication output: towards an unbiased metric of research performance. *PLoS ONE* 1, e127
- 12 Wenneras, C. and Wold, A. (1997) Nepotism and sexism in peer-review. *Nature* 387, 341
- 13 NIH Research Portfolio Online Reporting Tools (RePORT); NIH Data Book
- 14 Ley, T.J. and Hamilton, B.H. (2008) The gender gap in NIH grant applications. *Science* 322, 1472–1474
- 15 Pohlhaus, J.R. et al. (2011) Sex differences in application, success, and funding rates for NIH extramural programs. *Acad. Med.* 86, 759–767
- 16 Kaatz, A. et al. (2014) A quantitative linguistic analysis of NIH R01 application critiques from investigators at one institution. *Acad. Med.* (in press)
- 17 Demicheli, V. and Di Pietrantonj, C. (2007) Peer review for improving the quality of grant applications. *Cochrane Database of Syst. Rev.* 2007, MR000003
- 18 Carnes, M. et al. (2012) Promoting institutional change through bias literacy. *J. Divers. High. Educ.* 5, 63–77