Literature review: Bias in peer review and grant applications

Feb 2024

In the dynamic landscape of academia, the allocation of research funding serves as a cornerstone for advancing knowledge. Traditionally, peer review has been the gatekeeper, ensuring a fair and unbiased distribution of resources. However, emerging evidence suggests that this system is not immune to bias, presenting significant challenges to the integrity of the academic funding process (Schmaling & Gallo, 2023). This literature review embarks on an exploration of biases inherent in peer review and grant applications, with a specific focus on understanding the multifaceted dynamics that permeate this critical process.

Efforts by major funding bodies, such as the Engineering and Physical Sciences Research Council (EPSRC), and their approaches to mitigating bias are key aspects of our inquiry. The UK Research and Innovation (UKRI) has proactively released diversity data, revealing disparities in award rates categorised by age, ethnicity, gender, and disability annually (UK Research and Innovation, 2022). However, there are limitations to these analyses, such as overlooking non-additive interactions and not tracking multiple applications across years by the same applicant, which may potentially distort the estimation of bias pervasiveness. EPSRC has also separately published diversity data of the reviewer population (EPSRC, 2020b) and gender disparities (EPSRC, 2020a).

With regard to the gender gap in funding allocation, diverse sources of discrimination have been proposed, ranging from individual biases of reviewers and systemic bias in the evaluation criteria, to broader social barriers beyond the peer review process (Sato et al., 2021; Witteman et al., 2019). The same mechanisms might be at work regarding other protected characteristics too. To address these complexities, we propose a participatory approach, ensuring the active involvement of the community, not merely as subjects but as co-creators of knowledge. This participatory model proves particularly pertinent for understanding biases in peer review, where lived experiences and perceptions play a pivotal role in comprehending this complex phenomenon.

Motivated by the increasing importance of equity, diversity, and inclusion in academic research, this study seeks to explore the issue of bias in peer review and grant applications. Our aims are 1) to understand the funding portfolio of EPSRC in terms of protected group characteristics, 2) to detect and understand potential biases in reviewers' scoring and comments. The learning will feed into practical recommendations that can be implemented to enhance the fairness of the funding process.

1. Introduction

Reviewer bias in peer review refers to the presence of systematic prejudice or favouritism based on factors unrelated to the quality or relevance of the research. These biases can be based on the author's gender, nationality, institutional affiliation, or other non-scientific factors. Although there is a large body of literature on biases in peer review in scientific publication, in the context of this project we mainly consider peer review in grant allocation. Such biases can undermine the objectivity and fairness of the

peer review process, leading to the acceptance or rejection of grant applications for reasons other than their scientific merit. Biases can have profound implications, influencing not only the trajectory of individual researchers' careers but also the direction and priorities of entire fields of study.

The number of empirical studies looking into biases in the peer review process has steadily increased since the 1990s. Below we first give examples of review articles and empirical studies, then summarise the findings along different dimensions of discrimination and the methods used in the studies.

Examples of review papers and meta-analysis:

Ceci and Williams (Ceci & Williams, 2011) address the prevalent issue of women's underrepresentation in math-intensive fields of science. The traditional explanations often attribute this underrepresentation to gender discrimination in various aspects, including grant and manuscript reviewing, interviewing, and hiring. However, the authors argue that recent and more robust empirical evidence fails to support these claims of discrimination. After reviewing data from the past 20 years, the authors suggest that some of the previously asserted claims may no longer be valid. They emphasize that focusing on historical discrimination can divert attention from contemporary determinants of women's underrepresentation. The paper concludes that differential gender outcomes result from choices, whether free or constrained, and advocates for directing efforts toward education and policy changes to address the current challenges faced by women in science, technology, engineering, and mathematics careers. The authors propose potential interventions aligned with present findings to enhance gender fairness in these fields.

Schmaling and Gallo (Schmaling & Gallo, 2023) delve into the potential factors contributing to women's underrepresentation in the sciences, specifically examining gender differences in grant application outcomes. In contrast to narrative reviews, they conducted a meta-analysis following PRISMA 2020 standards to investigate grant award acceptance rates, reapplication award acceptance rates, and other grant-related outcomes. The findings indicate that, while award acceptance rates were similar for women and men, women tended to receive smaller awards and fewer awards after reapplying. This suggests potential disparities in the support provided to women in scientific research, with implications for their continued scientific productivity. The paper highlights the importance of greater transparency to monitor and verify these data globally, contributing valuable insights into the gender dynamics within peer-reviewed grant outcomes. Connecting to the previous work by Ceci and Williams, this study adds empirical evidence to the ongoing discussion about gender disparities in science and emphasizes the need for nuanced approaches in addressing the challenges faced by women in research and academia.

Cruz-Castro et al. gave a thorough discussion of the contradictory findings about discrimination in grant allocation in one review from the gender perspective (Cruz-Castro & Sanz Menéndez, 2020) and another covering other protected characteristics too (Cruz-Castro et al., 2022). In the former, they also unpacked the complexities around detecting bias or discrimination (e.g. unobserved heterogeneity, confusion between outcome and process, and dynamic processes), which contributed to differences in research designs and subsequent conclusions. They recommended incorporating the context of both the employing organisations and research funding organisations into the analysis. Similarly, they raised that the literature represents largely informative independent studies that are difficult to interpret and harmonise due to variations in their context (Cruz-Castro et al., 2022).

While there is limited research on effective interventions to mitigate these biases, a report commissioned by UKRI presents the findings of a study reviewing 38 interventions aimed at optimising and innovating the peer review processes used in Research and Innovation (R&I) award funding (Kolarz et al., 2023). The interventions range from small process adjustments to fundamental changes such as partial randomisation and complete bypass of peer review. The study identifies seven common aims of these interventions, including saving time, optimizing relevance, identifying high-risk projects, reducing burden, managing application volume, addressing bias and promoting inclusion, and improving the overall quality of reviews. The report emphasizes the need for tailored funding processes aligned with specific aims and contexts. While some interventions have ample evidence supporting their effectiveness, others require further research. The recommendation underscores the importance of integrating process design into scheme design for each funding opportunity, with a call for funders to share findings and consider widespread adoption of certain interventions across portfolios.

Examples of empirical studies:

In their 2022 study, McKenzie et al undertook a comprehensive mixed-methods investigation into the peer review process employed by the American Society of Clinical Oncology (ASCO) (McKenzie et al., 2022). The study utilised various data sources to examine implicit bias regarding sex, race, and institution within ASCO's peer review system. Firstly, a literature search was conducted to define implicit bias in peer review, identify evidence of its presence, and explore potential mitigation strategies. Additionally, the researchers created and analysed an ASCO database, examining sex, race, and institutional affiliation concerning peer review success. To gain in-depth insights, qualitative interviews were conducted with key stakeholders within ASCO, providing valuable perspectives on their experiences with implicit bias.

Exploring the realm of language features in the context of gender bias, van der Lee and Ellemers conducted a manual annotation study in 2015 (van der Lee & Ellemers, 2015). Their research delved into the gendered language used in the written communication originating from funding bodies. By scrutinizing linguistic nuances, this study aimed to uncover any biases embedded in the language employed in communication related to funding. This investigation contributes to a broader understanding of how language choices may influence perceptions and outcomes in the realm of grant applications, shedding light on yet another facet of the multifaceted issue of gender bias in research funding. In a similar vein, van den Besselaar and Mom employed NLP (Natural Language Processing) and algorithmic techniques . Their study specifically focused on identifying gender stereotyping through the analysis of negation words and agentic words in review reports. Utilizing the LIWC (Linguistic Inquiry and Word Count) tool, this research aimed to uncover subtle biases embedded in the language used during the peer review process. By employing computational methods, the study delved into the intersection of language and gender dynamics, contributing valuable insights to the ongoing discourse on gender bias in peer review.

2. Axes of bias/discrimination

2.1 Gender

Gender is the most often examined aspect regarding biases in grant making. While women's under-representation in the research workforce – especially in the STEM fields, and in higher positions within the research profession – is not disputed, studies disagree over whether the mechanisms behind involves bias or discrimination. A ground-breaking 1997 study by Wennerås and Wold showing female applicants for a postdoctoral fellowship were scored lower after controlling for scientific productivity shocked academia (Wennerås & Wold, 1997). However, subsequent replication studies conducted in various countries yielded mixed results. Many recent studies continue to detect gender bias while showing variations between disciplines, countries, and analysis methods (Cruz-Castro & Sanz Menéndez, 2020).

Several studies also found that women received smaller award amounts even when the overall success rates were similar (Blake & La Valle, 2001; Waisbren et al. 2008; Bedi et al. 2012). In some cases this is due to women requesting smaller amounts, possibly from differences in types and costs of research, sense of entitlement, or salaries associated with academic rank. Where funders have the option to allocate less than the requested amount, gender bias in the review and decision process could also lead to reduced funding for women (Schmaling & Gallo, 2023).

Since the early 2010s there has been studies supporting that the current underrepresentation of women in science is not mainly due to gender biases in the review process during grant funding or hiring, but pre-existing social barriers (e.g. lifestyle choices, free or unfree) and systemic biases (e.g. less likely to have work published/cited (Ceci & Williams, 2011). Common indicators for assessing excellence may trap women in a vicious cycle: having less publications and citations due to limited resources and lower positions, which in turn limits their access to resources in the future (Schiffbaenker et al., 2022).

Gender differences could also exist in the responses to feedback: women showed reduced motivation to re-apply for funding following review highlighting inadequacy (Biernat et al., 2020). This aligns with findings from Hosek et al. (Hosek et al., 2005), indicating that women tend to reapply less frequently than men to the same funding agency, irrespective of the outcome. Relatedly, women have been found to switch out of male-dominated STEM majors following poor performance more often than men (Kugler et al., 2017).

In most recent studies and review papers, a mixed picture of contradictory findings seems to have emerged as the consensus, highlighting the relevancy of context not just around the review process which is not homogenous between funding agencies, but also the organisational practices of institutions to understand the study outcomes.

Some studies also explored possible effect of the gender of the reviewers. Most studies did not detect any significant effect of reviewers' gender on gender bias in the outcome (Bornmann et al., 2007; Marsh et al., 2009; van der Lee & Ellemers, 2015). An study using a decade's data from the Austrian Science Fund did not detect gender bias in the final decision, but found that the approval probability decreased when there is parity or a majority of women among the reviewers (Mutz et al., 2012). A similar trend was detected in peer review reports submitted to the Swiss National Science Foundation, where male reviewers gave higher scores than female reviewers (Severin et al., 2020). Separately, regarding peer

review in scientific publication, homophily between gatekeepers (editors and reviewers) and authors was detected among submissions to *eLife* between 2012 and 2017 (Murray et al., 2018). In their result the gender inequality in submission outcome was the most pronounced when all reviewers are male.

2.2 Race/ethnicity and nationality

The literature on differences in research funding regarding race or ethnicity is much scarcer, as the data is not often collected outside the United States. Exploring disparities in National Institutes of Health (NIH) R01 grant funding, Ginther et al. (Ginther et al., 2011) uncovered racial disparities, with later research (Ginther et al., 2018) attributing a substantial portion of the funding gap to differences in field-adjusted bibliometric measures. Factors such as topic choice, discussion at a study session, and impact score emerged as key contributors to the gap (Hoppe et al., 2019).

Related to ethnicity, potential biases related to the nationality of applicants and reviewers have also been explored. The homophily in scientific publishing reported by Murray et al. also pertains to nationalities or countries of affiliation, although the details varied between countries (Murray et al., 2018). Additionally, foreign reviewers were found to provide more generous reviews in a study in Australia (Jayasinghe et al., 2001) and another in Switzerland (Severin et al., 2020).

2.3 Disciplines/topics

In the realm of peer review, the influence of reviewers' beliefs and past work on their evaluations introduces a potential confirmation bias, wherein research aligned with their perspectives may receive more favorable treatment. This cognitive bias raises concerns about the objectivity of the evaluation process. Additionally, a bias against novelty, interdisciplinary and a tendency toward risk aversion may lead to heightened scrutiny or skepticism for innovative or unconventional research proposals.

A pertinent example of the impact of topic choice on funding outcomes is evident in NIH grants, where Black scientists face disparities. Research indicates that Black scientists tend to propose projects on topics with lower award rates, often focusing on community and population-level studies rather than more fundamental and mechanistic investigations (Hoppe et al., 2019). This discrepancy highlights the need for addressing biases in topic evaluation and promoting equitable consideration of diverse research themes.

2.4 Age

The influence of age on research grant applications has been a subject of investigation, with early studies, such as Jayasinghe et al. (Jayasinghe et al., 2003) and those cited within, suggesting that age may not have a significant effect on the outcomes of grant applications. However, a more nuanced picture emerges from recent research. Notably, a study found that "[s]ignificantly lower application scores were also associated with applicants who were older, evaluated by female reviewers only" (Tamblyn et al., 2018). This indicates a potential age-related bias in the peer review process, raising questions about the equitable evaluation of proposals based on age.

It might be challenging to separate the effect of age from academic ranks and productivity metrics, especially considering these have been reported to interact with gender. For example, some research suggests that women in senior academic positions are as successful as men in obtaining funding but there remains a notable underrepresentation of women in senior positions (Holliday et al., 2014;

Waisbren et al., 2008). This disparity in the occupation of senior roles may contribute to the broader gender-related challenges observed in the research funding landscape. Understanding the multifaceted factors influencing grant outcomes, including age and gender dynamics, is essential for fostering a fair and inclusive research environment.

2.5 Connection to evaluators

Preferential treatment to candidates affiliated with a committee member was demonstrated to have a significant impact comparable to gender bias in the landmark study by Wennerås and Wold (Wennerås & Wold, 1997). Although reviewers with conflicts of interests were not allowed to score applicants they are associated with, the rule did not stop other committee members raising the scores for applicants affiliated with their peers. Such nepotism persisted in a replicate study conducted a decade later (Sandström & Hällsten, 2008). Notably, the authors of the latter study issued a correction in 2021 (Sandström & Hällsten, 2021), acknowledging that the effect of conflicts of interest should be smaller but still significant.

In the context of grant applications, the role of affiliations and connections is also evident. Research using NIH data indicated that, holding quality constant, an applicant's chances of being funded increase with each additional permanent member they have personal ties with (Li, 2017).ERC starting grant applicants sharing the same affiliation with a panelist have a higher probability of getting funded (Mom & Besselaar, 2022). However, this effect varies between countries and disciplines, with a more pronounced impact among men than women.

Furthermore, the choice of reviewers in the peer review process plays a crucial role. Applicant-nominated reviewers tend to give higher scores compared to reviewers invited by the panel, as observed in studies in Australia and Switzerland (Jayasinghe et al., 2001; Marsh et al., 2007; Severin et al., 2020). These findings underscore the intricate interplay of personal connections, affiliations, and the selection of reviewers in shaping the outcomes of grant applications, highlighting the need for ongoing efforts to enhance fairness and transparency in the peer review process.

2.6 Intersectionality

In a rare experimental study, Forscher et al. modified the PI names of 48 NIH R01 grant proposals to create white male, white female, black male, and black female versions of each proposal, which were then sent to reviewers (Forscher et al., 2019). They did not detect any pragmatically important pro-white, pro-male, or pro-white male biases. Their simulation-based power analysis suggested very high power to detect gap at least one-quarter the size between high-quality and moderate-quality proposals, although the authors did note their set-up differed from the true NIH process as the reviewers had lower workload and knew they were participating in a study.

3. Language in reviewer comments

Having access to reviewers' comments provides a valuable opportunity to explore more nuanced dimensions of implicit bias that may not be reflected in the scoring, but nevertheless influence the decisions of the panelists. Therefore, we also surveyed the literature for the language features that have been examined to detect potential gender biases.

3.1 Overall positivity/negativity

Quantifying the language used in reviewer comments provides valuable insights into implicit biases within the grant review process, with studies exploring the overall tone and uncovering patterns of positivity and negativity. Reviews might focus on the presence of required qualifications in men and the absence of them in women (Uhlmann & Cohen, 2005; Vinkenburg et al., 2014); a selection criteria (e.g. independence) may be more frequently mentioned in evaluations for men when they meet it, but for women when they do not (van den Besselaar & Mom, 2022).

3.2 Focus/themes

In a study looking at the critiques received by NIH K award applications, researchers reported subtle variations in themes for male and female investigators: low productivity led to doubt about future independence in women, but not in men; only women's career development plan was described as over-ambitious; concerns about research plan were directed at problems with the proposal for male applicants and followed by advice, but directed at perceived deficiencies in the ability for female applicants (Kaatz, Dattalo, et al., 2016). The last point echoes another study detecting reduced gender gap in application outcome when the review focus was explicitly placed on the proposed science compared to the calibre of the principal investigator (Witteman et al., 2019).

3.3 Agentic traits in men

Gender stereotypes depict women as more communal and men as more agentic, which led to the implicit bias that women are less competent in the male gender-typed domain of academic research (Kaatz et al., 2014). In the reviewers' critiques to NIH R01 grant applicants, men are more often described as "leaders" or "pioneers" in their fields, while women described as having "expertise" and working in "excellent" environment (Magua et al., 2017).

3.4 More praise for women

Perhaps surprisingly, one study found implicit bias masked behind positive comments for female applicants. Despite similar application scores and funding outcomes, female investigators received more praise in the reviewers' critiques while male investigators received more negative evaluation words (Kaatz et al., 2015). The authors believe gender stereotypes lead evaluators to give women greater praise than men for achieving the same performance. Others have raised different objective standards that the subjective descriptions map onto due to stereotypes, namely women were compared against an objectively lower standard of "good" than men (Biernat et al., 2012, 2020).

3.5 Negation bias

Gender stereotyping can also lead to more use of negation words for the "outgroup" even when describing positive qualities (van den Besselaar & Mom, 2022). Such "negation bias" reflects an underlying stereotypical expectation and has communicative consequences: the recipient would infer the description as an exception to the rule and more attributable to situational circumstances than inherent qualities of the person (Beukeboom et al., 2010).

4. Methods used

In the exploration of gender bias across different domains, researchers employ diverse methodologies, providing nuanced insights into the multifaceted nature of this phenomenon.

4.1 Quantitative observational studies of application outcomes

There have been many quantitative studies using datasets about the grant-making results, mostly focused on gender bias. They are mostly observational studies using multivariate regression models. There is a distinction between focusing on application level and person level (where multiple applications from the same applicant are treated as one record). Some studies only compared the reviewer scores or success rates segregated by gender (Bautista-Puig et al., 2019; Friesen, 1998; van der Lee & Ellemers, 2015), sometimes across career stages (Ley & Hamilton, 2008; Pohlhaus et al., 2011), but the majority attempted to control for the merit, impact or productivity of applicants using available proxies, including the first study that put gender bias in the selective process on the agenda (Wennerås & Wold, 1997). Commonly used proxies for productivity include bibliometrics (the number of publications, the impact factor of journals that the applicants published in, the number of citations received), and previous award history (usually only available for the same award). Institutions, year of application, age, job title, field of research, and academic degree are often included as covariates in the context of researching gender bias.

Some studies also include the characteristics of the decision makers as independent variables on their own and/or through interaction terms with the applicants' characteristics. These include, for example, the gender, nationality, ethnicity, and experience of the reviewers, and whether they were nominated by the applicant or the panel. At the aggregate level, the (dis-)agreement between reviewers for the same application has been explored too (Tamblyn et al., 2018). With regard to the panel, studies have looked into their gender composition, nationality composition, and workload (van den Besselaar & Mom, 2022). Beyond ratings and success rate, some studies also examined the total value of funding (Blake & La Valle, 2001; Schmaling & Gallo, 2023) and the propensity to apply or reapply (Biernat et al., 2020; Grant et al., 1997; van der Lee & Ellemers, 2015).

4.2 Experimental or quasi-experimental studies

As a rare example, Forscher et al. assigned the same grant research proposals different applicant names from which the gender and race can be inferred before sending out to recruited reviewers (Forscher et al., 2019). Witteman et al. Took advantage of a natural experiment where the Canadian Institutes of Health channelled applications into two new grant programmes, one with and the other without an explicit focus on the calibre of the principal investigator (Witteman et al., 2019). The statistical models used are similar to those used for analysing observational data. Another experimental study tested the effect of redacting grant applicant identity and institutional affiliation on peer review outcomes, although the redaction was found not to eliminate reviewers' ability to correctly infer features of identity (Nakamura et al., 2021).

4.3 Surveys and interviews

Although focusing on the context of scientific publishing, not grant competition, McKenzie et al. included interviews with key stakeholders within the ASCO board, publication and grants committee, as well as surveys among committee volunteers to evaluate the perception of implicit bias and possible

measures to mitigate (McKenzie et al., 2022). Van den Besselaar et al. interviewed panelists about the decision-making process and their interpretation and operationalization of the assessment criteria to provide context for their linguistic and bibliometric analysis (van den Besselaar et al., 2018). Schiffbaenker et al. (Schiffbaenker et al., 2022) interviewed reviewers of the ERC starting grant for how they understand and interpret the independence criterion, through which gender biases were revealed.

4.4 Linguistic analysis

Research on the unstructured texts ranges from qualitative to quantitative, from manual to automated and various combinations in between. Van der Lee and Ellemers examined the use of gendered language not in the reviewers' comments, but in written communications (grant calls, online application information, instructions for reviewers and interview committee) from the grant-making body through manual coding (van der Lee & Ellemers, 2015). Kaatz et al. conducted qualitative thematic analysis to identify common themes in each section of the critique, and compared how they differ between male and female applicants and between funded and unfunded applications (Kaatz, Dattalo, et al., 2016).

A more recent development is the application of algorithmic tools, including natural language processing (NLP), to linguistic analysis of reviewers' comments. Aided by tools like LIWC, the percentage of words under predefined linguistic categories (e.g. "ability", "agentic", "positive evaluation") has been analysed using linear models (Forscher et al., 2019; Kaatz et al., 2015; Kaatz, Lee, et al., 2016; van den Besselaar et al., 2018). Magua et al. applied topic modeling and identified one of the topics to correspond most closely to reviewers' sentiment (Magua et al., 2017). They then searched for other words that co-occur frequently with each word in the sentiment topic and extract the context around them, and compared the difference between those addressing male and female PIs combined with qualitative thematic analysis. Although van den Besselaar et al. (van den Besselaar et al., 2018) did not in particular examine reviewers' bias, they extended Kaatz et al. (Kaatz et al., 2015)'s LIWC categories and used linear regression to predict reviewers' scores from these categories.

Beyond the context of peer review in grant applications, Orgeira-Crespo et al. trained a convolutional neural network to detect gender non-inclusive language from a vector representation of the sentences in a corpus of doctoral theses from Spanish universities (Orgeira-Crespo et al., 2021). This approach, however, requires a manually labelled dataset for training, as do other work applying deep learning models in sentiment analysis and the assessment of review quality (Kang et al., 2018; Wang & Wan, 2018). The most common techniques to identify gender bias in NLP are based on word embeddings in the vector space: by analogy queries ("man is to A as woman is to X"), by demonstrating predictive classification of the gender a word refers to, or by showing male and female words form distinct clusters (Costa-jussà, 2019). Word embedding association test has become a benchmark for testing gender bias. Stanczak and Augenstein (Stanczak & Augenstein, 2021) gave a detailed review of other methods as well as the lexica and datasets available. However, the focus in this field is more on detecting (and mitigating) gender bias in NLP algorithms instead of implicit bias of the authors of the texts.

4.5 Previous Meta-Analyses: Building on Insights

Building on the foundation laid by earlier meta-analyses, such as the work by Bornmann et al. in 2007, subsequent research by Marsh et al. in 2009 aimed to refine the models used for gender effect analysis. The follow-up study suggested that compared to the fixed- and random-effects model, more sophisticated multilevel analyses that controlled for discipline, country, and year mitigate many gender-

related effects, highlighting the evolving nature of research methodologies and the continuous quest for improved models (Bornmann et al., 2007; Marsh et al., 2009). Nevertheless, a more recent meta-analysis returned to the more traditional methodology (Schmaling & Gallo, 2023).

5. Conclusion

This literature review highlights findings indicating the existence of biases in academic funding processes, including reviewer bias, gender and ethnicity disparities in grant outcomes, and the influence of personal connections on funding success. The outcomes underscore the importance of tailored interventions to address these biases effectively. The study recommends future research to explore practical measures for promoting fairness, diversity, and inclusion in peer review processes. Moreover, it advocates for participatory methods to gain nuanced insights into the experiences of underrepresented groups and enhance the overall integrity of academic funding processes.

References

Bautista-Puig, N., García-Zorita, C., & Mauleón, E. (2019). European Research Council: Excellence and leadership over time from a gender perspective. *Research Evaluation*, *28*(4), 370–382. https://doi.org/10.1093/reseval/rvz023

Bedi, G., Van Dam, N. T., & Munafo, M. (2012). Gender inequality in awarded research grants. The Lancet, 380(9840), 474. https://doi.org/10.1016/S0140-6736(12)61292-6

Beukeboom, C. J., Finkenauer, C., & Wigboldus, D. H. J. (2010). The negation bias: When negations signal stereotypic expectancies. *Journal of Personality and Social Psychology*, *99*(6), 978–992. https://doi.org/10.1037/a0020861

Biernat, M., Carnes, M., Filut, A., & Kaatz, A. (2020). Gender, Race, and Grant Reviews: Translating and Responding to Research Feedback. *Personality and Social Psychology Bulletin*, *46*(1), 140–154. https://doi.org/10.1177/0146167219845921

Biernat, M., Tocci, M. J., & Williams, J. C. (2012). The Language of Performance Evaluations: Gender-Based Shifts in Content and Consistency of Judgment. *Social Psychological and Personality Science*, *3*(2), 186–192. https://doi.org/10.1177/1948550611415693

Blake, M., & La Valle, I. (2001). *Who applies for research funding?: Key factors shaping funding application behaviour among women and men in British higher education institutions*. Wellcome Trust. https://wellcome.org/sites/default/files/wtd003209_0.pdf

Bornmann, L., Mutz, R., & Daniel, H. (2007). Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, *1*(3), 226–238. https://doi.org/10.1016/j.joi.2007.03.001

Buljan, I., Garcia-Costa, D., Grimaldo, F., Squazzoni, F., & Marušić, A. (2020). Large-scale language analysis of peer review reports. *eLife*, *9*, e53249. https://doi.org/10.7554/eLife.53249

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, *108*(8), 3157–3162. https://doi.org/10.1073/pnas.1014871108

Chakraborty, S., Goyal, P., & Mukherjee, A. (2020). Aspect-based Sentiment Analysis of Scientific Reviews. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 207–216. https://doi.org/10.1145/3383583.3398541

Costa-jussà, M. R. (2019). An analysis of gender bias studies in natural language processing. *Nature Machine Intelligence*, *1*(11), 495–496. https://doi.org/10.1038/s42256-019-0105-5

Cruz-Castro, L., Ginther, D. K., & Sanz-Menendez, L. (2022). *Gender and Underrepresented Minority Differences in Research Funding* (Working Paper 30107). National Bureau of Economic Research. https://doi.org/10.3386/w30107

Cruz-Castro, L., & Sanz Menéndez, L. (2020). *Grant Allocation Disparities from a Gender Perspective: Literature Review. Synthesis Report*. https://doi.org/10.20350/DIGITALCSIC/10548

EPSRC. (2020a, September 21). *Understanding our portfolio: A gender perspective*. https://www.ukri.org/publications/epsrc-understanding-our-portfolio-a-gender-perspective/

EPSRC. (2020b, December 10). *EPSRC peer review diversity data 2014-15 to 2019-20*. https://www.ukri.org/publications/epsrc-peer-review-diversity-data-2014-15-to-2019-20/

Forscher, P. S., Cox, W. T. L., Brauer, M., & Devine, P. G. (2019). Little race or gender bias in an experiment of initial review of NIH R01 grant proposals. *Nature Human Behaviour*, *3*(3), 257–264. https://doi.org/10.1038/s41562-018-0517-y

Friesen, H. G. (1998). Equal opportunities in Canada. *Nature*, *391*(6665), 326–326. https://doi.org/10.1038/34768

Ginther, D. K., Basner, J., Jensen, U., Schnell, J., Kington, R., & Schaffer, W. T. (2018). Publications as predictors of racial and ethnic differences in NIH research awards. *PLOS ONE*, *13*(11), e0205929. https://doi.org/10.1371/journal.pone.0205929

Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., & Kington, R. (2011). Race, Ethnicity, and NIH Research Awards. *Science*, *333*(6045), 1015–1019. https://doi.org/10.1126/science.1196783

Grant, J., Burden, S., & Breen, G. (1997). No evidence of sexism in peer review. *Nature*, *390*(6659), 438–438. https://doi.org/10.1038/37213

Holliday, E. B., Jagsi, R., Wilson, L. D., Choi, M., Thomas, C. R., & Fuller, C. D. (2014). Gender Differences in Publication Productivity, Academic Position, Career Duration, and Funding Among U.S. Academic Radiation Oncology Faculty: *Academic Medicine*, *89*(5), 767–773. https://doi.org/10.1097/ACM.0000000000000229

Hoppe, T. A., Litovitz, A., Willis, K. A., Meseroll, R. A., Perkins, M. J., Hutchins, B. I., Davis, A. F., Lauer, M. S., Valantine, H. A., Anderson, J. M., & Santangelo, G. M. (2019). Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Science Advances*, *5*(10), eaaw7238. https://doi.org/10.1126/sciadv.aaw7238

Hosek, S. D., Cox, A. G., Ghosh-Dastidar, B., Kofner, A., Ramphal, N. R., Scott, J., & Berry, S. H. (2005). *Gender Differences in Major Federal External Grant Programs*. RAND Corporation. https://www.rand.org/pubs/technical_reports/TR307.html

Jayasinghe, U. W., Marsh, H. W., & Bond, N. (2001). Peer Review in the Funding of Research in Higher Education: The Australian Experience. *Educational Evaluation and Policy Analysis*, *23*(4), 343–364.

Jayasinghe, U. W., Marsh, H. W., & 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 Series A: Statistics in Society*, *166*(3), 279–300. https://doi.org/10.1111/1467-985X.00278

Kaatz, A., Dattalo, M., Regner, C., Filut, A., & Carnes, M. (2016). Patterns of Feedback on the Bridge to Independence: A Qualitative Thematic Analysis of NIH Mentored Career Development Award Application Critiques. *Journal of Women's Health*, *25*(1), 78–90. https://doi.org/10.1089/jwh.2015.5254

Kaatz, A., Gutierrez, B., & Carnes, M. (2014). Threats to objectivity in peer review: The case of gender. *Trends in Pharmacological Sciences*, *35*(8), 371–373. https://doi.org/10.1016/j.tips.2014.06.005

Kaatz, A., Lee, Y.-G., Potvien, A., Magua, W., Filut, A., Bhattacharya, A., Leatherberry, R., Zhu, X., & Carnes, M. (2016). Analysis of National Institutes of Health R01 Application Critiques, Impact, and Criteria Scores: Does the Sex of the Principal Investigator Make a Difference? *Academic Medicine*, *91*(8), 1080–1088. https://doi.org/10.1097/ACM.0000000000001272

Kaatz, A., Magua, W., Zimmerman, D. R., & Carnes, M. (2015). A Quantitative Linguistic Analysis of National Institutes of Health R01 Application Critiques From Investigators at One Institution: *Academic Medicine*, *90*(1), 69–75. https://doi.org/10.1097/ACM.0000000000000442

Kang, D., Ammar, W., Dalvi, B., Zuylen, M., Kohlmeier, S., Hovy, E., & Schwartz, R. (2018). A Dataset of Peer Reviews (PeerRead): Collection, Insights and NLP Applications. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 1647–1661. https://doi.org/10.18653/v1/N18-1149

Kolarz, P., Vingre, A., Vinnik, A., Neto, A., Vergara, C., Rodriguez, C. O., Nielsen, K., & Sutinen, L. (2023). *Review of peer review*. https://www.ukri.org/publications/review-of-peer-review/

Kugler, A., Tinsley, C., & Ukhaneva, O. (2017). *Choice of Majors: Are Women Really Different from Men?* (w23735; p. w23735). National Bureau of Economic Research. https://doi.org/10.3386/w23735

Ley, T. J., & Hamilton, B. H. (2008). The Gender Gap in NIH Grant Applications. *Science*, *322*(5907), 1472–1474. https://doi.org/10.1126/science.1165878

Li, D. (2017). Expertise versus Bias in Evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics*, *9*(2), 60–92. https://doi.org/10.1257/app.20150421

Luo, J., Feliciani, T., Reinhart, M., Hartstein, J., Das, V., Alabi, O., & Shankar, K. (2021). Analyzing sentiments in peer review reports: Evidence from two science funding agencies. *Quantitative Science Studies*, *2*(4), 1271–1295. https://doi.org/10.1162/qss_a_00156

Magua, W., Zhu, X., Bhattacharya, A., Filut, A., Potvien, A., Leatherberry, R., Lee, Y.-G., Jens, M., Malikireddy, D., Carnes, M., & Kaatz, A. (2017). Are Female Applicants Disadvantaged in National Institutes of Health Peer Review? Combining Algorithmic Text Mining and Qualitative Methods to Detect Evaluative Differences in R01 Reviewers' Critiques. *Journal of Women's Health*, *26*(5), 560–570. https://doi.org/10.1089/jwh.2016.6021

Marsh, H. W., Bond, N. W., & Jayasinghe, U. W. (2007). Peer review process: Assessments by applicant-nominated referees are biased, inflated, unreliable and invalid. *Australian Psychologist*, *42*(1), 33–38. https://doi.org/10.1080/00050060600823275

Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., & O'Mara, A. (2009). Gender Effects in the Peer Reviews of Grant Proposals: A Comprehensive Meta-Analysis Comparing Traditional and Multilevel Approaches. *Review of Educational Research*, *79*(3), 1290–1326. https://doi.org/10.3102/0034654309334143

McKenzie, N. D., Liu, R., Chiu, A. V., Chavez-MacGregor, M., Frohlich, D., Ahmad, S., & Hendricks, C. B. (2022). Exploring Bias in Scientific Peer Review: An ASCO Initiative. *JCO Oncology Practice*, *18*(12), 791–799. https://doi.org/10.1200/OP.22.00275

Miner, A. S., Stewart, S. A., Halley, M. C., Nelson, L. K., & Linos, E. (2023). Formally comparing topic models and human-generated qualitative coding of physician mothers' experiences of workplace discrimination. *Big Data & Society*, *10*(1), 205395172211491. https://doi.org/10.1177/20539517221149106

Mom, C., & Besselaar, P. van den. (2022). *Do interests affect grant application success? The role of organizational proximity*. https://doi.org/10.48550/ARXIV.2206.03255

Murray, D., Siler, K., Larivière, V., Chan, W. M., Collings, A. M., Raymond, J., & Sugimoto, C. R. (2018). *Author-Reviewer Homophily in Peer Review* [Preprint]. Scientific Communication and Education. https://doi.org/10.1101/400515

Mutz, R., Bornmann, L., & Daniel, H.-D. (2012). Does Gender Matter in Grant Peer Review?: An Empirical Investigation Using the Example of the Austrian Science Fund. *Zeitschrift Für Psychologie*, *220*(2), 121–129. https://doi.org/10.1027/2151-2604/a000103

Nakamura, R. K., Mann, L. S., Lindner, M. D., Braithwaite, J., Chen, M.-C., Vancea, A., Byrnes, N., Durrant, V., & Reed, B. (2021). An experimental test of the effects of redacting grant applicant identifiers on peer review outcomes. *eLife*, *10*, e71368. https://doi.org/10.7554/eLife.71368

Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2021). The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods. *Sociological Methods & Research*, *50*(1), 202–237. https://doi.org/10.1177/0049124118769114

Orgeira-Crespo, P., Míguez-Álvarez, C., Cuevas-Alonso, M., & Rivo-López, E. (2021). An analysis of unconscious gender bias in academic texts by means of a decision algorithm. *PLOS ONE*, *16*(9), e0257903. https://doi.org/10.1371/journal.pone.0257903

Pennebaker, J., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. (2007). *The development and psychometric properties of LIWC2007*. https://www.semanticscholar.org/paper/The-development-and-psychometric-properties-of-Pennebaker-Chung/8fb7d85cefa1a0dbc5cec05f2c052c58ced71390

Pohlhaus, J. R., Jiang, H., Wagner, R. M., Schaffer, W. T., & Pinn, V. W. (2011). Sex Differences in Application, Success, and Funding Rates for NIH Extramural Programs: *Academic Medicine*, *86*(6), 759–767. https://doi.org/10.1097/ACM.0b013e31821836ff

Sandström, U., & Hällsten, M. (2008). Persistent nepotism in peer-review. *Scientometrics*, *74*(2), 175–189. https://doi.org/10.1007/s11192-008-0211-3

Sandström, U., & Hällsten, M. (2021). Correction to: Persistent nepotism in peer-review. *Scientometrics*, *126*(2), 1863–1865. https://doi.org/10.1007/s11192-020-03739-4

Sato, S., Gygax, P. M., Randall, J., & Schmid Mast, M. (2021). The leaky pipeline in research grant peer review and funding decisions: Challenges and future directions. *Higher Education*, *82*(1), 145–162. https://doi.org/10.1007/s10734-020-00626-y

Schiffbaenker, H., Haas, M., & Holzinger, F. (2022). The gendered nature of independence in the context of research funding and excellence. *SN Social Sciences*, *2*(12), 275. https://doi.org/10.1007/s43545-022-00563-w

Schmaling, K. B., & Gallo, S. A. (2023). Gender differences in peer reviewed grant applications, awards, and amounts: A systematic review and meta-analysis. *Research Integrity and Peer Review*, *8*(1), 2. https://doi.org/10.1186/s41073-023-00127-3

Severin, A., Martins, J., Heyard, R., Delavy, F., Jorstad, A., & Egger, M. (2020). Gender and other potential biases in peer review: Cross-sectional analysis of 38 250 external peer review reports. *BMJ Open*, *10*(8), e035058. https://doi.org/10.1136/bmjopen-2019-035058

Stanczak, K., & Augenstein, I. (2021). *A Survey on Gender Bias in Natural Language Processing*. https://doi.org/10.48550/ARXIV.2112.14168

Tamblyn, R., Girard, N., Qian, C. J., & Hanley, J. (2018). Assessment of potential bias in research grant peer review in Canada. *Canadian Medical Association Journal*, *190*(16), E489–E499. https://doi.org/10.1503/cmaj.170901

Uhlmann, E. L., & Cohen, G. L. (2005). Constructed Criteria: Redefining Merit to Justify Discrimination. *Psychological Science*, *16*(6), 474–480. https://doi.org/10.1111/j.0956-7976.2005.01559.x

UK Research and Innovation. (2022, December 1). *Diversity data*. https://www.ukri.org/what-we-do/supporting-healthy-research-and-innovation-culture/equality-diversity-and-inclusion/diversity-data/

van den Besselaar, P., & Mom, C. (2022). *Gender differences in research grant allocation—A mixed picture*. https://doi.org/10.48550/ARXIV.2205.13641

van den Besselaar, P., Sandström, U., & Schiffbaenker, H. (2018). Studying grant decision-making: A linguistic analysis of review reports. *Scientometrics*, *117*(1), 313–329. https://doi.org/10.1007/s11192-018-2848-x

van der Lee, R., & Ellemers, N. (2015). Gender contributes to personal research funding success in The Netherlands. *Proceedings of the National Academy of Sciences*, *112*(40), 12349–12353. https://doi.org/10.1073/pnas.1510159112

Vinkenburg, C. J., Jansen, P. G. W., Dries, N., & Pepermans, R. (2014). Arena: A Critical Conceptual Framework of Top Management Selection. *Group & Organization Management*, *39*(1), 33–68. https://doi.org/10.1177/1059601113492846

Waisbren, S. E., Bowles, H., Hasan, T., Zou, K. H., Emans, S. J., Goldberg, C., Gould, S., Levine, D., Lieberman, E., Loeken, M., Longtine, J., Nadelson, C., Patenaude, A. F., Quinn, D., Randolph, A. G., Solet, J. M., Ullrich, N., Walensky, R., Weitzman, P., & Christou, H. (2008). Gender Differences in Research Grant Applications and Funding Outcomes for Medical School Faculty. *Journal of Women's Health*, *17*(2), 207–214. https://doi.org/10.1089/jwh.2007.0412

Wang, K., & Wan, X. (2018). Sentiment Analysis of Peer Review Texts for Scholarly Papers. The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 175–184. https://doi.org/10.1145/3209978.3210056

Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, *387*(6631), 341–343. https://doi.org/10.1038/387341a0

Witteman, H. O., Hendricks, M., Straus, S., & Tannenbaum, C. (2019). Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *The Lancet*, *393*(10171), 531–540. https://doi.org/10.1016/S0140-6736(18)32611-4