Peer-reviewing, the multi-edged sword:

Distributions of power and burden among marginalized researchers

Julia Reiter*[1], Helena Hartmann*[2], Simon T. van Baal*[3], the Peer Review Project Consortium, & Alessandro Nai[4]

[1] *Department of Occupational, Economic and Social Psychology, University of Vienna, Austria.*

[2] *Clinical Neurosciences, Department of Neurology and Center for Translational Neuro- and Behavioral Sciences (C-TNBS), University Hospital Essen, Germany.*

[3] *Analytics, Technology and Operations Department, Leeds University Business School, University of Leeds, LS2 9JT Leeds, UK*

[4] *Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Amsterdam 1001 NG, The Netherlands.*

*Shared first authorship

Consortium authors (sorted alphabetically):

- William L. Allen, University of Oxford
- Isabelle Borucki, Philipps-University Marburg
- Gabriele Bossi, University of Amsterdam
- Mark Boukes, University of Amsterdam
- Elena Escalante-Block, University of Antwerp
- Max Grömping, Griffith University
- Alice Hamilton, University of Amsterdam
- Helena Hartmann, University Hospital Essen
- Alessandro Nai, University of Amsterdam

- Julia Reiter, University of Vienna
- Stuart J. Turnbull-Dugarte, University of Southampton
- Simon T. van Baal, University of Leeds
- Fabio Votta, University of Amsterdam

Author note

Correspondence regarding this article should be addressed to Dr. Julia Reiter, Department of Occupational, Economic and Social Psychology, University of Vienna, Universitätsstraße 7, 1010 Vienna, Austria. julia.reiter@univie.ac.at

Julia Reiter: https://orcid.org/0000-0002-7451-9441

Helena Hartmann: https://orcid.org/0000-0002-1331-6683

Simon T. van Baal: https://orcid.org/0000-0001-5351-4361

Alessandro Nai: https://orcid.org/0000-0001-7303-2693

Acknowledgements

This work was done as part of the Peer Review Consortium (https://www.peerreviewstudy.com). We thank all consortium members and study participants.

Declaration of Interest

The authors have no competing interests to declare.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used u:ai (the University of Vienna's instance of ChatGPT) in order to fix/improve R code. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Funding

HH was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 422744262 - TRR 289 (Deutsche Forschungsgemeinschaft, DFG).

Author contributions:

- Conceptualization: All authors.
- Data curation: All authors.
- Formal analysis: JR, HH, & SvB.
- Investigation: All authors.
- Methodology: All authors.
- Project administration: JR & HH.
- Software: AN, JR, HH & SvB
- Validation: JR & HH.
- Visualization: JR, HH, & SvB.
- Writing – original draft: JR, HH, & SvB.
- Writing – review and editing: All authors.

Abstract

Peer review is a pillar of the scientific system, serving as a pre-publication quality assurance mechanism. Active peer reviewers hold power over which ideas and approaches enter the scientific discourse. They are simultaneously burdened with unpaid labour, as peer reviewing is generally only incentivized by the notion of contributing to the public scientific good. The present large-scale mixed-methods global study (n = 3677 social scientists from 49 countries) investigates the power, burden, and perceptions of peer reviewing in terms of academic prestige, geography, gender identity, academic legacy, and structural marginalisation. We find a clear bias favouring reviewers in and from academically prestigious countries. Men reviewed slightly more than others, and marginalized researchers felt they were investing more time and effort in their reviews than others. Exploratory analyses of reasons for rejecting review requests and spontaneous associations with peer review add to a nuanced picture of how differences in academic prestige shape the peer review landscape. We discuss underlying mechanisms and potential implications for improving peer review processes.

Keywords

Peer review, WEIRD bias, gender bias, first-generation researchers, marginalization, discrimination

Highlights

- We present a large-scale mixed-methods global study (n = 3677, 49 countries).
- A bias in peer review frequency favours reviewers from and in academically prestigious (usually, Western) countries.
- Men peer-reviewed slightly more than other genders.
- Marginalized researchers felt they were investing more time and effort than others.

1. Introduction

Peer review, designed as a key element for upholding scientific rigor in academic publishing, is fraught with challenges and problems (Smith, 2006; Tite & Schroter, 2007): the ever-increasing load of new manuscripts, the academic system not rewarding or incentivizing peer review (relying on injunctive norms to contribute to the public good), and the difficulty for editors to find enough reviewers. At the same time, peer review is the backbone of scientific innovation; ensuring the fairness and inclusivity of the peer review process is directly related to upholding high standards of rigor and fairness in the scientific process. Among the most threatening problems due to its potential to cause extensive harm to scientific integrity is the presence of systematic bias. Biases in peer review have systemic consequences for the scientific system. If certain groups are excluded from peer review, whether by lack of invitation or rejection, this increases bias through the loss of valuable expertise from the quality assurance system and risks perpetuating existing power structures (Haffar et al., 2019; Lee et al., 2013). For example, lack of incentives for peer review further exacerbates the situation of marginalized scholars, who can ill-afford to donate their time to a cause that disproportionately benefits more powerful scholars through their higher publication rates.

1.1 Bias in the Peer Review System

Although the scientific peer review system should be a globalised system making use of available peer reviewers all over the world, it is also subject to inherent power structures which impact who gets invited to conduct peer review and who gets (systematically) excluded. Some of these effects include the positive feedback loops researchers at institutions historically seen as prestigious receive and the devaluing of local research in non-Wealthy,

Educated, Industrialised, Rich, and Democratic (WEIRD) countries (Bol et al., 2023; González-Sauri & Rossello, 2023; Ross et al., 2006). Both influence editorial board participation rates (network effects, further devaluing non-WEIRD research due to lack of representation; Goyanes et al., 2022), asymmetric inclusion in citation networks, funding success, all leading to a phenomenon usually termed the Matthew Effect, the phenomenon that early-career institutional and supervisor prestige increases the likelihood of success throughout the rest of one's career (see e.g., Gonzalez-Sauri & Rosello, 2023). Those power structures thus go beyond typical trajectories of increasing invitations at higher career stages and are affected by other intersecting sociodemographic, personal, and environmental factors.

In terms of immediate social relations biasing the peer-review process, the most obvious source of bias is collaborator favouritism, or vice versa, rival hostility. Teplitskiy et al. (2018) show evidence of co-author favouritism across neuroscience papers submitted to *PLOS One*, in that manuscripts' scientific validity was judged more favourably the closer the authors were to the reviewers within their professional networks. Investigating a similar mechanism of social capital causing bias, Si et al. (2023) document an asymmetrical affiliation bias across Chinese journals, with male researchers affiliated with high-prestige institutions profiting more from being (previously) affiliated with a journal's host institution than both women and researchers at low-prestige institutions in terms of article acceptance. This illustrates how bias caused by personal relationships intersects with group-based bias mechanisms, thus reinforcing structural inequalities.

One major category of structural bias is gender. Lerback & Hanson (2017) report evidence of gender bias in peer review in journals of the American Geophysical Union, with fewer women being invited to review manuscripts, and women under the age of 60 declining their (fewer) invitations at a slightly higher rate than men in the same age group. Helmer et al. (2017) find similar results in *Frontiers* journals. Additionally, institutional affiliation, both

the specific institution and the institution's country, can have bias attached to it due to the complex mechanisms underlying the social construction of academic prestige centring Western-ness. Smith et al. (2023) reported worse review outcomes in the biological sciences (e.g., lower overall acceptance rates) for authors with institutional affiliations in Asia and whose country's primary language is not English. Similarly, Rubin et al. (2023) document European editors favouring authors from the European Union in their acceptance decisions.

Finally, Teplitskiy et al. (2018) also present evidence for an ingroup bias along "schools of thought"; that is, irrespective of closeness to the authors, reviewers tended to evaluate manuscripts more favourably if they aligned with their views on contested substantive matters in their field. This mechanism of bias interacts with the abovementioned mechanisms to create lasting consequences for the peer review system: If certain segments of the population are disproportionately excluded from the peer review system, e.g., due to national or institutional bias, this can lead to certain schools of thought remaining underrepresented among reviewers, making it more difficult for authors to publish manuscripts adhering to these schools of thought, thus creating a cycle that keeps some narratives out of the dominant scientific discourse.

1.2 Mechanisms and Consequences of Systemic Bias: Stereotyping and Competence

Competence is a central part of the content of gender stereotypes, with men being seen as more competent but less socially warm versus women being viewed as warmer but less competent, especially when working in fields that are not typically considered feminine (e.g., Abele & Wojciszke, 2007; Fiske et al., 2002). As such, women in science and especially in STEM fields may be affected more strongly by stereotypical ascriptions of low competence, which can in turn result in them being invited to review papers less frequently than their male counterparts. These perceptions may also contribute to lower review

acceptance rates, as internalising these stereotypes can result in women experiencing impostor phenomenon, thus doubting their own competence (Clance & Imes, 1978). Women experiencing impostor phenomenon may feel unfit to judge someone else's manuscript where their male counterparts feel more confident in the scope of their expertise, which could lead to them declining requests more frequently.

Similarly, stereotypes about different nationalities and geographic regions prominently feature assumptions about competence. A large-scale study of 38 geographic regions found that ethnic groups from warmer climates and from lower wealth countries - i.e., typically non-Western countries - were viewed as less competent (Grigoryev, 2022). Generally, groups seen as having a lower status than the ingroup are assigned higher stereotypic warmth but lower stereotypic competence (Fiske, 2017). This asymmetry, in conjunction with stereotypes about nations and ethnic groups, may explain the link between lower institutional prestige and fewer invitations for peer review, as researchers who are, by association, assumed to be less competent will be invited to review less frequently.

We previously mapped peer review practices across 61 countries to investigate the status quo from the peer reviewers' perspective (BLINDED). We found that while scholars in North America and Western Europe received significantly more invitations to review, other regions demonstrated a higher ratio of invitations accepted to reviews completed. These findings hint at different peer reviewing practices related to burden and power. Thus, further exploration of the geographical and socio-demographical distribution of these phenomena is warranted.

The goal of the present study is to extend this work and shed light on the distribution of power and burden in the peer review system, and explore how this relates to current power structures, focusing on different marginalizations in researchers across the globe.

1.3 Research Questions in the Present Study

We aim to map the relationships between multiple facets of societal marginalization and involvement in the international, academic peer review system. Specifically, we investigate the relationships of academic prestige, geography (both in terms of nationality and institutional affiliation), gender identity, academic legacy, discriminatory experiences, perceived disadvantage, and degree of disadvantage in everyday life, with peer review involvement.

2. Methods

We conducted a large-scale survey among international scholars in the social sciences. The study and all hypotheses were preregistered on the OSF before data collection: https://osf.io/t4a63/?view_only=e23a13ee8c1647bb97f15477413d05cf. For details on the larger study, see [BLINDED]. All participants provided informed consent for participation in the study and anonymous collection of their data. Ethics approval was granted by the ethics committee of the [BLINDED].

2.1 Deviations from Preregistration

There is one notable deviation from the preregistration that results in an addition, rather than omission, of results: We preregistered the comparison between Western and non-Western countries as a central divide throughout the paper. However, in discussions between the authors during the analysis phase, a different divide was identified which we now believe to be closer to the core of the issue we seek to investigate: The divide between countries that are, by virtue of a mix of cultural prejudice, post-colonial thought, and reputation of individual institutions in the countries, considered to be academically prestigious vs. not prestigious. While there were multiple definitions of "the West" to choose from, we believe

the group of "prestigious countries" to be smaller than even the narrow definitions of the West. We use the Times Higher Education (THE) ranking as a proxy for this differentiation, as detailed below. All results on geographical differences remain fully reported in footnotes, but the main structure of the text now follows the differentiation between prestigious and non-prestigious countries instead. Note that we understand "prestige" not as a normative term implying that the country is deserving or undeserving of recognition, but as a descriptive term. Our concept of an academically prestigious country means that, in absence of other information, the (academic) public is likely to think an academic institution must be good because it is based in said country - regardless of whether this judgement is rooted in factual information, prejudice, or a mix of both. The term is italicized in the following text to highlight this specific usage.

2.2 Hypotheses

We tested several preregistered hypotheses in this study that relate nationality and institutional affiliation, gender identity, academic legacy, discriminatory experiences, perceived disadvantage, and degree of disadvantage in everyday life with peer review involvement. Note that we rearranged the order of hypotheses to improve the narrative flow of the paper; see the Table in Supplement S1 for a translation of hypothesis numbers in the text to hypothesis numbers in the preregistration.

2.2.1 National Academic Prestige

We divided the countries in our sample into *prestigious* (P) and *non-prestigious* (NP) countries. We used the 2022 THE ranking as a proxy for the most influential institutions per country, selecting the first 15 or, if fewer than 15 are listed, the first 10 entries. If fewer than 10 universities were listed, the search was supplemented with a Google search for "[country] university". The number of institutions in the USA was oversampled ($n_{institutions}$ = 20). The

resulting list of institutions were grouped *prestigious* (P) and *non-prestigious* based on the countries they are located in, by taking the mean rank of the institutions we sampled from that country on the THE 2022 ranking (i.e., the country ranking is calculated as: $\frac{\sum_i^n r_i}{n}$), where r_i is the rank of an institution *i* and *n* is the number of institutions sampled in that country. This method was chosen to reflect that university rankings are in large part privilege and prestige rankings (indeed, many include a prestige metric explicitly). We acknowledge that our method is imperfect like any dichotomisation of this spectrum would be, but deem it appropriate in the context of this work. The resulting *prestige (P)/non-prestige* list: Australia (P), Belgium (P), Benin, Brazil, Canada (P), Chile, China (P), Colombia, Costa Rica, Croatia, Czech Republic, Denmark (P), Egypt, Ethiopia, France, Germany (P), Ghana, Hong Kong (P), Hungary, India, Indonesia, Iran, Israel, Italy, Japan, Kazakhstan, Kenya, Lebanon, Lithuania, Malta, Mexico, Netherlands (P), New Zealand, Nigeria, Norway, Peru, Philippines, Poland, Russia, Singapore, South Africa, South Korea (P), Spain, Sweden (P), Switzerland (P), Tanzania, Tunisia, Turkey, Uganda, United Kingdom (P), United States (P), and Uruguay. See supplement S2 for a description of the preregistered geographic hypotheses and how the related categorisation of countries differs from the one presented above.

2.2.2 Gender Influence

We hypothesized that gender would affect the number of invitations received (H5), either to the effect that male scholars receive more invitations than female scholars (female incompetence beliefs hypothesis; H5a), or to the effect that female scholars receive more invitations (female unpaid labour hypothesis, H5b). Additionally, we explored whether there were differences in invitation frequency between cis men, cis women, and other gender identities (H5c). We alternatively hypothesized that, relative to the number of invitations received, male scholars review either more often than female scholars (superior competence

beliefs hypothesis, H6a), or female scholars review more often than male scholars (higher readiness to take on unpaid labour/compensation hypothesis, H6b). In terms of subjective perceptions, we hypothesized that male scholars on average believe they are performing a higher number of reviews than comparable scholars, as compared to what scholars with other gender identities believe (H7a), and that male scholars believe they are spending more time on their reviews than comparable peers, compared to other gender identities (H7b).

2.2.3 Academic Family Legacy

We hypothesized that first-generation academics are less likely to be invited to peer-review than scholars with at least one academic parent (H8). We further hypothesized that legacy scholars would either be reviewing more frequently than first-generation scholars (confidence, feeling entitled to judge others' level of scholarship hypothesis; H9a), or less frequently than first-generation scholars (compensation hypothesis, H9b). In terms of subjective perceptions, we hypothesized that scholars with at least one academic parent on average believe they are performing a higher number of reviews and spending more time on reviews than comparable scholars, as compared to what first-generation academic scholars believe (H10).

2.2.4 Discriminatory Experiences and Perceived Disadvantage

Self-identification as a member of a group who is discriminated against is negatively related to frequency of invitation to peer-review (discrimination < no discrimination; H11). Moreover, scholars who select multiple grounds for discrimination (intersectional discrimination) are invited to peer-review less frequently than those who select a single ground for discrimination (intersectional < single; H11a). We tested whether being part of a discriminated group and being discriminated against are related to the frequency of accepting peer review requests (H12). We hypothesized that those who self-identify as a member of a

discriminated group review more often than those who don't identify as such (H12a), and that those who experience multiple disadvantages review more frequently than those who do not (H12b).

We hypothesized that the perceived disadvantage index is negatively related to the frequency of invitation to peer review, with higher disadvantage being associated with lower invitations (H13). Moreover, the perceived disadvantage index is positively related to frequency of accepting invitations to peer review relative to the frequency of invitations received, with higher disadvantage being associated with a higher ratio of acceptance (i.e., number of accepted invitations divided by total invitations, resulting in the percentage of accepted invitations; H14a). Regarding subjective perceptions, we tested competing hypotheses: the perceived disadvantage index is either negatively related to the perceived number of reviews completed in relation to comparable peers (low confidence hypothesis; H15a), or positively related to the same perception (compensation hypothesis; H15b).

We further hypothesized a relationship between the perceived impact of one's disadvantage(s) on everyday life and acceptance ratio would be negative (H14b). Regarding subjective perceptions, we hypothesized that the perceived impact of disadvantage(s) on everyday life would either be negatively (H17a) or positively (H17b) related to how much time scholars thought they were investing in their reviews compared to comparable peers.

2.3 Sample

We implemented a four-stage stratified sampling approach to gather contact data from researchers working in the social sciences. First, we identified 49 countries which included the three most populous countries in each of seven regions (Africa, Asia, Western Europe, Central and Eastern Europe, North America, South America, and Oceania) and additional countries with average and low population sizes in each region (though not below 200,000

inhabitants) to ensure representation of different levels of population diversity in the sample. Countries with active totalitarian regimes and with no universities ranked in the Time Higher Education (THE) ranking were excluded from recruitment (since the THE was used for the systematic identification of universities). We originally selected 50 countries, but there was no reliable information available about universities in Botswana, so the country had to be excluded. Within these countries, we sampled 10 universities or similar research institutions. In each institution (university, research centre), all departments, teaching units and research clusters related to the main research disciplines covered ("social sciences", broadly defined[1]) were manually identified, and their websites listed in a masterfile. Approximately 3600 such departments were identified. We then scraped the contact details of all members of these departments and reached out to this contact list with survey invitations. After removing 2,000 respondents from the list who were used in a pilot study, the final contact list included 155,439 potential respondents. These respondents were contacted and invited to participate in the survey in early 2024. The survey was additionally circulated on social media to reach a wider audience. The newly sampled countries that were not specified in the preregistration were omitted in the geographical hypothesis tests, but were included in the analyses of *prestigious* vs. *non-prestigious* countries.

A total of $n = 3677$ researchers participated ($n = 1645$ cis female, $n = 1869$ cis male, $n = 37$ other, $n = 126$ no disclosure). Respondents were overwhelmingly first-generation academics ($n = 2936$, versus n = 548 who had one academic parent and $n = 172$ who had two academic parents) of *prestigious* nationalities, working in *prestigious* countries ($n = 1611$, versus $n = 1584$ *non-prestigious* at *non-prestigious* universities, $n = 170$ *prestigious* at *non-*

[1] Anthropology, Archeology, Communication and Media, Demography and Life-Course-Studies, Economics and Econometrics, Education, Finance, Gender Studies, Geography, Planning and Development, History, Law, Linguistics and Language, Marketing, Organizational Behavior and Human Resources, Philosophy, Political Science and International Relations, Psychology, Public Administration, Social Work, Sociology, Strategy and Management, Transportation and Logistics, Urban Studies and Architecture.

prestigious universities, and n = 312 *non-prestigious* at *prestigious* universities). The majority of the sample, around 30%, were between 35 and 45 years old (see Supplement S3). The clear majority of participants was tenured (or equivalent; 56.19%, n = 2066), with another 8.59% (n = 314) on tenure track and 4.92% (n = 181) already retired. 13.92% (n = 512) were post doctoral researchers without any direct tenure prospects, 11.88% (n = 437) were PhD students, and 3.18% (n = 117) of participants were undergraduate students.

2.4 Measures

The measures reported in the present article were part of a larger survey; the full questionnaire can be found in the OSF project (https://osf.io/hqvar/?view_only=2b1a9120a8de4e89b0ea3150955ff20c).

2.4.1 Interactions with Peer Review

For our main outcome, participants were asked to estimate the number of review invitations for scientific journals they received in 2022 (the last complete year at the time of the study; M = 10.39, SD = 16.41) and how many article reviews they performed in the same year (M = 5.66, SD = 7.01). They also provided estimates of their own perceived participation in the peer review process relative to comparable peers in terms of career stage within their discipline (specifically, whether they believed the extent to which they were performing reviews (M = 2.95, SD = 1.05) and investing time in these reviews (M = 3.45, SD = .82) was much higher or much lower than the average, on a 5-point scale ranging between these options).

2.4.2 Gender Identity and Academic Family Legacy

We also asked participants to provide information on their gender identity (single-choice question, response options: female; male; non-binary; agender; genderfluid;

transgender; other (specify); prefer not to say) and whether either of their parents holds a PhD and/or has worked for an academic institution in a research or teaching role (academic family legacy; response options: yes, one parent; yes, both parents; no, neither).

2.4.3 Geographic Indicators

We further asked what country's nationality participants held and which country their institutional affiliation was in, in each case prompting them to select the main one if they held multiple.

2.4.4 Perceived Disadvantage Index

The Perceived Disadvantage Index is an author-created summative score of ways in which a person's life is harder than the average, resulting in hindrances to the person's scientific career. We created this score based on some of the identities listed in the Academic Wheel of Privilege (Middleton et al., 2025). The index includes caregiving responsibilities (we assessed whether such responsibilities existed for children under 10 in the person's household or for other family members, such as sick or old relatives, and whether they were shared or borne alone), academic family background (whether the person is a first-generation academic), teaching workload (whether the person's role allows a focus on research or has a higher teaching load), and experiences of intersectional discrimination. The latter was measured by presenting participants with the following list of features and counting due to how many of them they considered themselves regularly disadvantaged by or discriminated against: nationality or origin, ethnicity, race, or skin colour, immigration status, religion or values, political beliefs, language, social class or socioeconomic background, education, age, gender, sexual orientation, relationship or family status, appearance, mental health and ability, physical health and ability. The resulting PDI ranges between 0 and 20, with higher scores indicating higher disadvantage ($M = 2.74$, $SD = 1.99$).

2.4.5 Degree of Disadvantages in Everyday Life

Lastly, following the list of characteristics that could be grounds for disadvantages or discrimination, participants were asked to rate on a 0 - 100 slider scale to what extent they felt, all things considered, that their everyday life was affected. The slider was labelled "It affects me…", with the extremes labelled as "not at all" (0) and "a great deal" (100) (M = 44.00, SD = 26.94).

2.4.6 Free Associations with Peer Review

At the very beginning of the survey, participants were asked to provide spontaneous associations with the phrase "peer review". To ensure that these remained spontaneous, the task was limited to 20 seconds. Participants were instructed as follows: "What comes first to your mind when you think about "peer review" in your discipline? How do you feel about it? On the next page we will ask you to write the first word associations that come to your mind when thinking about peer review. Single words are enough. You will have 20 seconds to write your thoughts, before the survey automatically moves to the next question. A timer will show how much time you have left. Do not worry about making typos." Participants then had the option to provide up to ten individual entries. We conduct an exploratory sentiment analysis on these associations to illustrate how they differ between peer reviewers from prestigious/non-prestigious countries.

2.4.7 Reasons for Rejecting Review Requests

Participants were asked to indicate their likelihood of rejecting a review request on a scale from 1 (very unlikely) to 5 (very likely) for various reasons: lack of time (M = 4.12, SD = 1.18), disliking the journal (M = 3.17, SD = 1.36), having recently been rejected by the journal (M = 2.08, SD = 1.15), lack of interest in ever publishing in the journal (M = 2.24, SD = 1.21), doubting the journal's scientific reputation (M = 4.63, SD = .94), doubting one's

expertise concerning the article (M = 4.59, SD = .83), fundamentally disagreeing with central claims of the article (e.g., philosophically or morally; M = 2.76, SD = 1.20), and recognising the authors (M = 3.49, SD = 1.29). Additionally, participants were asked to indicate what other reasons had caused them to reject a review request in the past, using an open-ended response format.

2.5 Analyses

2.5.1 Preregistered Analyses

Deviating from the preregistration, analyses were conducted in R (version 4.4.1) instead of JASP for improved reproducibility. The following packages were used: *dplyr* (Wickham et al., 2025), *psych* (Revelle, 2025), *haven* (Wickham, Miller & Smith, 2025), *ggplot2* (Wickham, 2016), and *BayesFactor* (functions *anovaBF* and *generalTestBF*; Morey et al., 2024). Frequentist ANOVAs with post hoc Tukey's *HSD* contrasts were calculated to test any group comparison hypotheses. General linear models were calculated to answer questions about perceived disadvantages predicting engagement with the peer review system. All analyses controlled for researchers' level of seniority ("career stage").

The full results of the Bayesian analyses are reported in Supplement S5 and referenced below for our confirmatory hypotheses as BF_{10} (= evidence for the alternative vs. the null hypothesis). In interpreting these values, a BF_{10} = 1-3 has been suggested to indicate weak evidence, a BF_{10} = 3-20 positive evidence, a BF_{10} = 20-150 strong evidence, and BF_{10} > 150 very strong evidence for the alternative hypothesis, while BF_{10} < 1 indicating more evidence for the null vs. the alternative hypothesis (the smaller the value, the more evidence for the null; see also Jarosz and Wiley, 2014). Note that the referenced BFs relate to the likelihood of the main effect and not post-hoc comparisons.

2.5.2 Exploratory Analyses

As mentioned above, we opted for centering the exploratory analysis of differences between *prestigious* and *non-prestigious* countries instead of the preregistered analysis of Wealthy Western vs. non-Wealthy Western countries. This new classification was based on the Times Higher Education (THE) ranking 2022 (see Sample section).

Additionally, we explored group differences in the contents of free associations with peer review ($n_{associations}$ = 11,191). To see whether there were quantifiable differences between groups in those associations, we conducted sentiment analysis. The sentiment analysis was conducted using an open-source model from the Hugging Face platform (Hartmann et al., 2021), which assigns a probability to one of three labels: positive, neutral, and negative, for each response. To ensure the validity of these findings, we first performed conservative preprocessing steps, removing any low-effort or non-responses. We excluded variations of "N/A", " ", single/two letter words, excluding "ok", indistinguishable three-letter combinations (e.g., "fka", "rew"), but finishing a few that were unambiguous (e.g., "hel" = help, "fak" = fake, "goo" = good, "lik" = like; $n_{associations}$ = 10,935). As a measure of sentiment, we take the top-probability label for each document.

Further insights into participants' peer review burden were given through analysis of their reasons for rejecting review requests. We analysed responses to a closed and an open-ended question, using BERTopic topic modelling for the latter (Grootendorst, 2022). BERTopic outperforms classical topic modelling methods such as Latent Dirichlet Allocation on shorter documents (e.g., one or two sentences of feedback). Classical topic modelling approaches use word co-occurrence within documents to cluster them, while BERTopic uses embeddings (i.e., multidimensional vectors) based on tokenised text to represent the meaning of short documents (here: usually a paragraph). BERTopic then uses dimensionality reduction, in our case via uniform manifold approximation and projection, clusters those

dimensions, in our case using hierarchical density-based spatial clustering of applications with noise. As a result, the approach is more robust for short documents (where other models would struggle due to low co-occurrence of words within documents), spelling errors, uncommon but important words, multigrams (e.g., "peer review" is a bigram), and stopwords. Therefore, the preprocessing requirements are also reduced. As such, our only preprocessing step for the topic modelling was to remove answers shorter than three words.

To analyse differences between groups, we used a class-based topic modelling approach to see how likely topics were based on the group the author belonged to. Those frequency tables were then used for exploratory chi-square tests.

3. Results

3.1 What is the Relationship Between Institutional Prestige, Geography, and Peer Review?

Current peer review in the social sciences is strongly affected by academic prestige. Researchers at universities in academically prestigious countries received more peer review invitations ($F(1, 3233) = 26.46$, $p < .001$; M = 11.79 versus M = 8.88 invitations per year). This effect was further amplified when taking into account the researchers' nationalities ($F(3, 3231) = 18.78$, $p < .001$): Researchers coming from non-prestigious countries, but working in prestigious countries received fewer invitations (M = 8.45) to peer review than those coming from and working in prestigious countries (Tukey's HSD(3, 3231) = -3.92, $p < .001$; M = 12.43), but also fewer than those coming from prestigious countries but now working in a non-prestigious country (HSD(3, 3231) = -5.36, $p = .01$; M = 13.67). Those coming from and working in non-prestigious countries were invited to review least (M = 8.35), both compared to those who had migrated from prestigious to non-prestigious countries (HSD(3, 3231) = -

5.53, $p < .001$) and to those who were from and working in prestigious countries (HSD(3, 3231) = -4.10, $p = < .001$).[2]

Relative to the number of invitations received, researchers at universities in non-prestigious countries performed more reviews ($F(1, 563) = 12.15$, $p < .001$).[3] In absolute terms, researchers with non-prestigious affiliations reviewed M = 4.99 articles per year, compared to M = 6.27 for researchers with prestigious affiliations.

Finally, researchers' beliefs about the number of reviews they perform relative to comparable peers in the field did not differ depending on nationality ($F(1, 2715) = 0.90$, $p = .343$)[4], nor country of affiliation ($F(1, 2715) = 0.01$, $p = .945$)[5], but their beliefs about the amount of time and effort they put in their reviews depend on both. Researchers with a non-prestigious nationality ($F(1, 2614) = 14.27$ $p < .001$)[6] and researchers affiliated with a non-prestigious university ($F(1, 2614) = 5.03$, $p = .03$)[7], respectively, believe they are investing more effort than comparable peers.

[2] Note that this pattern parallels the Western vs. non-Western divide we tested in line with the preregistration: Researchers at Western universities received more peer review invitations than those at non-Western universities (confirming H1; $F(1, 3176) = 34.43$, $p < .001$, $BF_{10} > 150$; $M = 11.40$ invitations/year ($SD = 18.00$) at Western universities and $M = 7.62$ invitations/year ($SD = 10.95$) at non-Western universities). This effect was further amplified when taking into account the researchers' nationalities ($F(3, 3151) = 15.65$, $p < .001$, $BF_{10} > 150$): Westerners at Western universities received more peer review invitations ($M = 11.74$, $SD = 18.59$) than non-Westerners at Western universities ($HSD(3, 3151) = -3.77$, $p = .005$; $M = 7.98$, $SD = 8.81$) and non-Westerners at non-Western universities ($HSD(3, 3151) = -4.38$, $p < .001$; $M = 7.37$, $SD = 10.48$). Notably, there was no significant difference to the number of invitations received by Westerners at non-Western universities ($HSD(3, 3151) = -2.07$, $p = .630$; $M = 9.58$, $SD = 14.30$), leaving H2 partly confirmed.

[3] Similarly, researchers at non-Western universities performed more reviews relative to the number of invitations ($F(1, 3502) = 12.88$, $p < .001$, $BF_{10} = 16.64$; undirected difference hypothesis H3 - $M = .58$ and $SD = .36$ for Western universities, $M = .62$ and $SD = .39$ for non-Western universities) than those at Western universities.

[4] Similarly, there is no difference for Western vs. non-Western nationalities: $F(1, 2663) = 1.50$, $p = .221$.

[5] Similarly, there is no difference for Western vs. non-Western affiliation: $F(1, 2670) = 0.01$, $p = .911$.

[6] Similarly to researches from non-Western countries: ($F(1, 2563) = 17.36$, $p < .001$); undirected hypothesis H4a.

[7] Similarly to researchers at non-Western universities: ($F(1, 2573) = 9.62$, $p = .002$); undirected hypothesis H4b.

3.2 How do Affiliations with Academic Prestige Affect Representations and Evaluation of the Peer Review System?

Despite these differences, researchers' primary spontaneous associations with peer review were similar across prestige-categories, as shown in Figure 1.

Figure 1.

Word cloud associations with peer review.

Note. Orange indicates negative valence, blue indicates positive valence, grey indicates neutral valence, determined by sentiment analysis. Relative word size represents the word frequency per group.

Across groups, narratives of peer review as a burdensome task were prominent, as illustrated by relatively common descriptors like "time consuming", "difficult", "slow" or "long", and "annoying". To a lesser extent, there were narratives of systemic bias, exemplified by associations like "unfair", "biased", "arbitrary" and "luck". This stands opposite the very prominent narrative focused on the necessity of peer review for maintaining

trustworthy science, exemplified by descriptors like “quality”, “necessary”, “feedback”, “important”, and “evaluation”. In addition to this important function of peer review, participants emphasized the importance of a positive communication culture in the peer review process, indicated by descriptors like “constructive”, “useful”, “interesting”, and “help”. Respondents from less prestigious affiliations were less likely to use “constructive” (15 occurrences out of 4985 total vs 48/5951), $\chi^2(1) = 5.160$, $p < .01$, while those from more prestigious nationalities were, counterintuitively, more likely to use “gate keeping/gatekeeping” (13/4579 vs 39/6361), $\chi^2(1) = 5.423$, $p = .02$.

As a proxy for review effort, we examined the length of reviews researchers shared in the survey ($n_{reviews} = 1367$). The mean review length differed across groups, the non-prestigious affiliation and nationality group submitting a mean review length of 748.43 characters ($SD = 521.79$, $n_{reviews} = 522$), while reviewers with prestige in either or both of the categories submitted longer reviews ($M = 974.45$, $SD = 580.69$, $n_{reviews} = 845$).

3.2.1 Being Invited vs. Choosing to Participate: Does Prestige Affect Reasons for Declining Review Invitations?

Yes. An initial quantitative analysis showed that researchers benefitting from academic prestige through affiliation and nationality indicated a higher likelihood of rejecting a peer review request due to self-interest (specifically, lack of time: $HSD(3, 3652) = -.25$, $p < .001$; disliking the journal: $HSD(3, 3639) = -.38$, $p < .001$; recent rejection by the journal: $HSD(3, 3639) = -.11$, $p = .037$; no interest in publishing in the journal: $HSD(3, 3643) = -.38$, $p < .001$) or fundamental concerns regarding the journal (dubious reputation, potentially predatory journal: $HSD(3, 3641) = -.16$, $p < .001$) than those from and in non-prestigious countries (see Table S6 in the Supplement). Inversely, those from and in non-prestigious countries indicated a higher likeliness to reject a review request if they had fundamental disagreements with core tenets of the article, e.g., on a philosophical or moral level ($HSD(3, 3633) = .20$, $p < .001$). By contrast, across groups, participants indicated no

differences in their likelihood to reject an article because they recognised the authors *(F*(3, 3637) = 1.01, p = .386) or because they felt their expertise was lacking (F(3, 3641) = 1.58, p = .195). An additional analysis of the open-ended responses participants provided when elaborating upon further reasons for declining review requests, using topic modelling, provides a more differentiated view: Using BERTopic to analyse reasons for declining review requests ($n_{reasons}$ = 658), we found thirteen topics with a coherence score of .557 (see Table S7 in the Supplement). The topics were clearly defined, corresponding to the different reasons why people may decline to review (Tite & Schroter, 2007). For example, researchers enjoying an element of prestige were less likely to mention the methodology is outside their area of expertise (Topics 7 & 8 combined; 7.2%) than those without prestige (15.7%, $\chi^2(1) = 10.869$, $p < .001$), and more likely to mention poorly written abstracts or poor methods (as described in the abstract) as reasons for rejection (12.8%, $\chi^2(1) = 4.690$, $p = .03$).

3.3 Does Gender Affect how Researchers Interact with the Peer Review System?

Minimally. Cis men, cis women, and researchers with other gender descriptions did not differ significantly from each other in terms of review invitations when controlling for seniority level (F(2, 3117) = 1.57, p = .207, BF_{10} = 0.06; thus not confirming H5a - c; M_{women} = 9.90, M_{men} = 10.90, M_{other} = 9.89). Notably, men and women were distributed similarly across the career stages in our sample, with the other gender descriptions category too small to make comparisons (see Supplement S4). While the number of invitations received did not differ, there were minimal, but statistically significant differences in review performance (F(2, 3440) = 24.58, $p < .001$, $BF_{10} > 150$). Specifically, men completed more reviews (on average, 6.18 publications per year) relative to their number of invitations than women (on average, 5.07 publications per year; *HSD*(2, 3440) = .08, $p < .001$, confirming H6a, but not H6b) and those with other gender identities completed slightly fewer reviews (on average, 4.54 publications per year) than both men (*HSD*(2, 3440) = -.22, p =.001) and women

(*HSD*(2, 3440) = -.14, p = .050).[8] Somewhat in line with these reports, men believed they were completing a higher number of reviews than comparable peers (F(2, 2626) = 3.72, p = .024, confirming H7a). Contrary to H7b, there were no systematic gender differences regarding how much time and effort people believed they were investing in peer review compared to comparable peers (F(2, 2524) = 1.03, p = .356).

3.4 Do First-Generation Academics Interact Differently with the Peer Review System than Others?

No. We found no significant differences in terms of number of invitations received (F(2, 3213) = .20, p = .822, BF_{10} = 0.01; contrary to H8), nor in terms of reviews completed relative to number of invitations (F(2, 3543) = .50, p = .607, BF_{10}'s = 0.01/0.03; contrary to H9a - b).[9] In line with these reports, first-generation researchers neither believed that they were completing more reviews than comparable peers (F(2, 2703) = 1.29, p = .276; contrary to H10), nor that they were investing more time and effort in their reviews than comparable peers (F(2, 2602) = .173, p = .842; contrary to H10).

3.5 Does Experiencing Systemic Disadvantages or Discrimination Affect how Researchers Interact with the Peer Review System?

Minimally. Those who reported experiencing discrimination or structural disadvantages on at least one dimension did not receive fewer peer review invitations than those who reported no such discrimination (M = 10.37 vs. M = 10.42; F(1, 3233) = .002, p = .968, BF_{10} = 0.04, contrary to H11), nor were they more or less likely to accept peer review requests (M = 5.58 vs. M = 5.77; F(1, 3187) = .47, p = .493, BF_{10}'s = 0.05/1.41, contrary to

[8] To test this result's robustness, the analysis was rerun with the absolute number of reviews completed instead of ratio, yielding the same pattern of results: F(2, 3120) = 9.95, p < .001.
[9] To test this result's robustness, the analysis was rerun with the absolute number of reviews completed instead of ratio, yielding the same pattern of results: F(2, 3543) = 0.47, p = .493.

H12).[10] The Perceived Disadvantage Index, contrary to H13, was not significantly related to the number of invitations received - that is, people were not less likely to be invited as a function of how many intersecting structural disadvantages otherwise impact their lives (β = -.02, t = -.14, p = .890, BF_{10} = 0.05). Those facing a higher number of intersecting disadvantages completed slightly fewer reviews relative to the number of invites received, contrary to H14a (β = -.01, t = -2.10, p = .036, BF_{10} = 0.98; $F(6, 3486) = 42.94$, $p < .001$, *Adj. R2* = .067).[11] The perceived severity with which those disadvantages impacted participants' lives overall was not related to the ratio of review request acceptance, contrary to H14b (β = -.00, t = -1.16, p = .247, BF_{10} = 0.09).[12] Researchers' perception of the amount of reviews they were completing compared to comparable peers was not related to the amount of intersecting disadvantages they experienced (contrary to H15a; β = -.00, t = -.45, p = .656, BF_{10} = 0.07), nor to the extent to which these disadvantages impacted their lives (contrary to H15b; β = -.00, t = -.56, p = .578, BF_{10} = 0.09). Researchers who considered themselves part of a discriminated-against group did not think they were completing more reviews than comparable peers ($F(1, 2715) = .00$, p = .955), but did think they were putting more effort into their reviews than comparable peers ($F(1, 2614) = 8.38$, p = .004; H16). In line with this, the perceived impact of structural discrimination and disadvantages on researchers' lives was positively related to perceiving the effort they put into reviewing as higher than that put in by comparable peers (in line with H17b and contrary to H17a; β = .003, t = 3.42, p = .001, BF_{10} = 22.07).

[10] To test this result's robustness, the analysis was rerun with the absolute number of reviews completed instead of ratio, yielding the same pattern of results: $F(1,3187) = 0.47$, p = .493.

[11] To test this result's robustness, the analysis was rerun with the absolute number of reviews completed instead of ratio, yielding the same pattern of results: $F(6, 3129) = 18.13$, p <.001.

[12] To test this result's robustness, the analysis was rerun with the absolute number of reviews completed instead of ratio, yielding the same pattern of results: $\beta = 4.55$, $t = 5.18$, $p < .001$)

4. Discussion

We conducted a large-scale international survey of academics representing different backgrounds, career stages, and institutions, about their experiences with and conduct in the peer review system in 2022. We surveyed 3677 researchers in total and were able to demonstrate an imbalance favouring - typically wealthy Western - countries which hold a certain academic prestige, resulting in more invitations to participate in peer review. Those affiliated with non-prestigious institutions were invited to review 8.88 articles per year, compared to 11.79 invitations for scholars at prestigious institutions. This closely mirrored the Western bias we found during our preregistered analyses, with 11.40 invitations per year for those affiliated with Western universities compared to an average of 7.62 invitations for those affiliated with non-Western universities. That is, academic peer review in the social sciences is a disproportionately wealthy Western endeavour shaped by socially constructed ideas of academic prestige.

Conducting reviews for journals is a more than double-edged sword; a review conducted properly, with all the thought, thoroughness and care one would hope to receive, takes up substantial amounts of cognitive effort and time - time many academics do not have, but actively need to carve out, often out of their spare time or the time that would otherwise be going into furthering their own research. Reviewing is, although normatively expected, typically not financially compensated and does not receive much professional recognition. Many perform reviews for the public good with much frustration and little career-relevant merit. On the flipside, as any researcher who has ever received a rejection they perceived as unfair or has felt pressured to make changes to a manuscript they did not agree with can attest to, reviewers hold a certain power in the academe. This extends beyond the individual article: Researchers who have established themselves as experts in a certain niche will be invited to review many related articles, enabling them to control and steer the narratives that enter the

scientific discourse around their topic. By virtue of their expertise and editors deferring to the chosen reviewers' expertise, they have the power to accelerate, delay, or entirely halt new methodological approaches, theoretical arguments and changes of perspectives to their area of expertise. While some academics exert this power with purpose and even abuse it by engaging in questionable reviewing practices (see our prior work; (BLINDED)), such as favouring articles by authors one associates with or dismissing articles based on personal vendetta or simple competition, a large portion of reviewers are presumably uncritically unaware of the power structures underlying the practice and their epistemic implications. The recent growth in the field of gender medicine and the field's efforts to close centuries old gaps in medical research conducted without women illustrates the detrimental effects of individual population segments harbouring epistemic control over other, marginalized groups, excluding them from the creation of knowledge and narratives. Notably, even small bias can result in substantial consequences: A study of peer reviews of grant applications found that bias as low as just under 3% of the total evaluation score was enough to significantly affect funding rates (Day, 2015).

Our finding that those from and in academically prestigious countries are invited to review more frequently, yet complete a lower ratio of reviews than those from and in non-prestigious countries, combined with different favoured rationales for rejecting review requests, provides insight into the complex interaction of power, burden, and academic prestige in the peer review system. The countries considered academically prestigious based on the THE ranking are overwhelmingly Western countries. Stereotypes attributing higher competence, academic skills, or rigour to those in WEIRD countries, along with post-colonial historical narratives about Western cultural supremacy and academic dominance, may play a part in the bias we observe in terms of invitations to review. This would explain why the prestige of individual high-achieving institutions appears to extend to the entire country

associated with them. Additionally, differences in financial and other resources as well as differences in political systems contribute, as they result in researchers at some institutions having more opportunities to develop and perform excellent research than others - that is, a reputation of academic excellence is at least partly the product of sufficient funds. We further found that while nationality and affiliation interact, it appears that nationality takes precedence - that is, researchers from prestigious countries moving to institutions in non-prestigious countries are still invited to review more frequently than is the case vice versa. This may be because as researchers move, their social capital carries over to the new location: Established networks and connections do not disappear when switching universities, which may counter the loss of visibility to editors and other researchers that might otherwise occur when moving to a non-prestigious country. Additionally, stereotypes may continue to play a role in that a name associated with a prestigious nationality may continue to draw review invitations at a non-prestigious university.

On the side of acceptance as opposed to invitation, we found that those from and in non-prestigious countries accepted a higher ratio of requests relative to those from and in prestigious countries. Group differences in the reasons why participants chose to reject review requests point towards different views on duty and burden in this context: Those from and in prestigious countries considered themselves more likely to reject review requests out of self-interest or when the invitation did not benefit them, e.g., when they disliked or had no interest in publishing in the journal or had recently been rejected by the journal themselves. They also cited concerns about standards of scientific quality more frequently, e.g., the journal having a dubious, potentially predatory reputation or, as shown in the topic modelling analysis, poor abstract quality in terms of methodology or writing style. This may indicate those with less academic prestige taking a less selective approach to peer review, complying with review requests out of normative principle rather than consideration for own interests.

Those from and in non-prestigious countries cited concerns about their methodological expertise for reviewing an article more frequently, while overall concerns about expertise did not differ between nationalities and affiliations. Underlying systematic differences in type and range of methods used in research may be related to differences in available resources for training as well as potential differences in epistemologies. Finally, while those in and from non-prestigious countries accepted more review requests relative to the number of invitations, those from and in prestigious countries on average provided longer reviews (with a difference of 226 characters), perhaps indicating different approaches towards or expectations of sufficient participation in the peer review system.

4.1 Limitations

There are a few limitations worth mentioning. Although our age range was quite broad, the sample was slightly biased towards older and/or tenured colleagues from Western countries. Future work might therefore focus stronger on early career researchers, such as (post)graduate students, PhD students and non-tenured personnel from the majority world. This is relevant, as those are most affected by power/burden issues of peer review. In a way, this sample also showcases exactly this problem of trying to reach those individuals that suffer the most from the current system. Moreover, we had a specific sampling strategy for each country, but relied in part on convenience sampling and of course on individuals interested to participate. In addition, the THE ranking we used for our classification is likely biased towards research-heavy institutions, especially in countries like the US and UK, leading to stronger skews in power/participation in peer review. In contrast, the average institution sampled in smaller countries might be more "standard" and teaching focused. Future work might try to use an alternative sampling strategy for better comparability. Lastly, our division based on prestige can be considered somewhat arbitrary, as it could be done in many other ways and might be dependent on time and context. This is why we see our

findings as a starting point for people to build and expand upon. We highly recommend that other researchers try to validate and replicate our methods and results.

4.2 Implications

This study has potential implications to address invitation biases in peer reviewing that relate to the different parties involved in this process. Universities may strive to improve the visibility of their institution and individual researchers, including searchable websites and easy-to-access e-mail addresses. Similarly, individual researchers may create their own websites or online presence where their expertise is clearly outlined, to facilitate peer review invitations. Editors who are searching for reviewers may actively look beyond WEIRD countries and "prestigious" universities; systematically expanding the criteria and extent of their search could contribute to reducing this bias.

Moreover, certain measures need to be considered in the future that balance out burden and inequalities related to peer reviewing, such as providing incentives for review work (although the exact design of incentive structures is not trivial; Squazzoni et al., 2013, experimented with reviewing incentives and found a decline in review quality). Lastly, providing opportunities for review training that go beyond "piggy-backing" on the supervisors' shoulders during their peer review and towards international, free and systematic workshops and courses that teach the art of peer review may help to level the playing field.

4.3 Conclusion

Peer review in the social sciences is affected by bias introduced by geography and academic institutional prestige. These biases shape both the number of review invitations received and the rate of invitation acceptance, suggesting multiple, interacting mechanisms producing these biases. Consequently, a set of multiple intervention approaches is required to correct this systemic issue and improve peer review.

5. Data statement

We provide the complete survey materials, open data, and analysis code on the Open Science Framework [https://osf.io/t4a63/?view_only=e23a13ee8c1647bb97f15477413d05cf].

6. References

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*(5), 751–763.

Bol, J. A., Sheffel, A., Zia, N., & Meghani, A. (2023). How to address the geographical bias in academic publishing. *BMJ Global Health, 8*(12).

Clance, P. R., & Imes, S. A. (1978). The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychotherapy: Theory, research & practice*, *15*(3), 241.

Day, T. E. (2015). The big consequences of small biases: A simulation of peer review. *Research Policy*, *44*(6), 1266-1270.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878–902.

Fiske, S. T. (2017). Prejudices in cultural contexts: Shared stereotypes (gender, age) versus variable stereotypes (race, ethnicity, religion). *Perspectives on psychological science, 12*(5), 791-799.

González-Sauri, M., & Rossello, G. (2023). The role of early-career university prestige stratification on the future academic performance of scholars. *Research in Higher Education, 64*(1), 58-94.

Goyanes, M., De-Marcos, L., Demeter, M., Toth, T., & Jordá, B. (2022). Editorial board

interlocking across the social sciences: Modelling the geographic, gender, and institutional representation within and between six academic fields. *PloS one, 17*(9), e0273552.

Grigoryev, D. (2022). Ethnic stereotype content beyond intergroup relations within societies: Exploring the North-South hypothesis for competence and warmth. *Cross-Cultural Research*, *56*(4), 345-384.

Haffar, S., Bazerbachi, F., & Murad, M. H. (2019). Peer review bias: a critical review. *Mayo Clinic Proceedings, 94*(4), 670-676.

Hartmann, J., Heitmann, M., Schamp, C., & Netzer, O. (2021). The Power of Brand Selfies. *Journal of Marketing Research*, *58*(6), 1159-1177. https://doi.org/10.1177/00222437211037258

Helmer, M., Schottdorf, M., Neef, A., & Battaglia, D. (2017). Gender bias in scholarly peer review. *eLife, 6*, e21718.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving, 7*(1), 2.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2-17.

Lerback, J., & Hanson, B. (2017). Journals invite too few women to referee. *Nature, 541*(7638), 455-457.

Malcom, D. (2018). It's time we fix the peer review system. *American Journal of Pharmaceutical Education, 82*(5).

Middleton, S. L., Iley, B., Sulik, J., Elsherif, M., & Azevedo, F. (2025). The academic wheel of privilege: An equity-based tool for authorship order. *Unpublished manuscript.* Retrieved August 7th 2025, from https://forrt.org/awop/

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2024).

BayesFactor: Computation of Bayes Factors for common designs. R package version 0.9.12-4.7, https://doi.org/10.32614/CRAN.package.BayesFactor.

William Revelle (2025). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.5.6, https://CRAN.R-project.org/package=psych.

Ross, J. S., Gross, C. P., Desai, M. M., Hong, Y., Grant, A. O., Daniels, S. R., ... & Krumholz, H. M. (2006). Effect of blinded peer review on abstract acceptance. *Jama, 295*(14), 1675-1680.

Rubin, A., Rubin, E., & Segal, D. (2023). Editor home bias?. *Research Policy*, *52*(6), 104766.

Si, K., Li, Y., Ma, C., & Guo, F. (2023). Affiliation bias in peer review and the gender gap. *Research Policy*, *52*(7), 104797.

Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*(4), 178-182.

Smith, O. M., Davis, K. L., Pizza, R. B., Waterman, R., Dobson, K. C., Foster, B., ... & Davis, C. L. (2023). Peer review perpetuates barriers for historically excluded groups. *Nature Ecology & Evolution, 7*(4), 512-523.

Squazzoni, F., Bravo, G., & Takács, K. (2013). Does incentive provision increase the quality of peer review? An experimental study. *Research Policy*, *42*(1), 287-294.

Teplitskiy, M., Acuna, D., Elamrani-Raoult, A., Körding, K., & Evans, J. (2018). The sociology of scientific validity: How professional networks shape judgement in peer review. *Research Policy*, *47*(9), 1825-1841.

Tite, L., & Schroter, S. (2007). Why do peer reviewers decline to review? A survey. *Journal of Epidemiology & Community Health, 61*(1), 9-12.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wickham H, François R, Henry L, Müller K, Vaughan D (2025). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, https://dplyr.tidyverse.org.

Wickham H, Miller E, Smith D (2025). *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. R package version 2.5.5, https://haven.tidyverse.org.

Supplementary Materials

Supplement S1. Table of Hypothesis Numbers in Text vs. Preregistration

Hypothesis no. in text	Hypothesis no. in pre-registration and analysis code	Text in preregistration
H1	H1	Scholars whose institutional main affiliation is currently in a WW country are invited to peer review articles more frequently across levels of professional seniority than those in non-WW countries.
H2	H2	Country of main affiliation and nationality interact, such that scholars with both nationality and affiliation of WW countries are invited to peer-review most frequently, followed by those with non-WW nationalities but WW affiliations, then those with WW nationalities but non-WW affiliations, and finally those with both nationality and affiliation of non-WW countries, who are invited least frequently.
H3	H6 (ex.)	Relative to the number of invitations received, do scholars with WW affiliations review more or less frequently than scholars with non-WW affiliations?
H4a	H16a (ex.)	Do beliefs about the number of reviews produced and amount of time spent on them compared to comparable scholars differ depending on, (a) WW vs. non-WW nationality, (b) WW vs. non-WW institutional affiliation, and (c) self-identification as a member of a discriminated group?
H4b	H16b (ex.)	Do beliefs about the number of reviews produced and amount of time spent on them compared to comparable scholars differ depending on, (a) WW vs. non-WW nationality, (b) WW vs. non-WW institutional affiliation, and (c) self-identification as a member of a discriminated group?

H5a	H3a	Male scholars receive more invitations to peer review than female scholars (female incompetence beliefs hypothesis).
H5b	H3b	Female scholars receive more invitations to peer review than male scholars (female unpaid labour hypothesis).
H5c	H3c (ex.)	Are there systematic differences between cis male, cis female, and all other gender identities in terms of (i) representation across career stages relative to group size and (ii) frequency of being invited?
H6a	H7a	Relative to the number of invitations received, male scholars review more often than female scholars (superior competence beliefs hypothesis).
H6b	H7b	Relative to the number of invitations received, female scholars review more often than male scholars (higher readiness to take on unpaid labour/compensation hypothesis).
H7a	H14a	Male scholars on average believe they are performing a higher number of reviews than comparable scholars, as compared to what scholars with other gender identities believe.
H7b	H14b	Male scholars on average believe they are spending more time on reviews than comparable scholars, as compared to what scholars with other gender identities believe.
H8	H5	First-generation academics are less likely to be invited to peer-review than scholars with at least one academic parent.
H9a	H9a	Relative to the number of invitations received, scholars with at least one academic parent review more frequently than first-generation academics (confidence, feeling entitled to judge others' level of scholarship hypothesis).
H9b	H9b	Relative to the number of invitations received, first-generation academics review more frequently than scholars with at least one academic parent

		(compensation hypothesis).
H10	H15	Scholars with at least one academic parent on average believe they are performing a higher number of reviews and spending more time on reviews than comparable scholars, as compared to what first-generation academic scholars believe.
H11	H4	Self-identification as a member of a group who is discriminated against is negatively related to frequency of invitation to peer-review (discrimination < no discrimination).
H11a	H4a	Scholars who select multiple grounds for discrimination (intersectional discrimination) are invited to peer-review less frequently than those who select a single ground for discrimination (intersectional < single).
H11b	H4b (ex.)	Are there systematic differences in frequency of invitation based on the group feature that is subject of discrimination?
H12	H8	Being part of a discriminated group and discrimination are related to the frequency of accepting peer review requests.
H12a	H8a	Relative to the number of invitations received, those who self-identify as a member of a discriminated group review more often than those who don't.
H12b	H8b	Relative to the number of invitations received, those who are discriminated against on multiple grounds review more often than those who don't.
H13	H10	The Perceived Disadvantage Index is negatively related to frequency of invitation to peer review, with higher disadvantage being associated with lower invitations.
H14a	H11a	The Perceived Disadvantage Index is positively related to frequency of accepting invitations to peer review relative to the frequency of invitations received, with higher disadvantage being associated with a higher ratio of invited vs. accepted peer reviews (i.e., number of accepted invitations divided by total invitations, resulting in the percentage of accepted invitations).

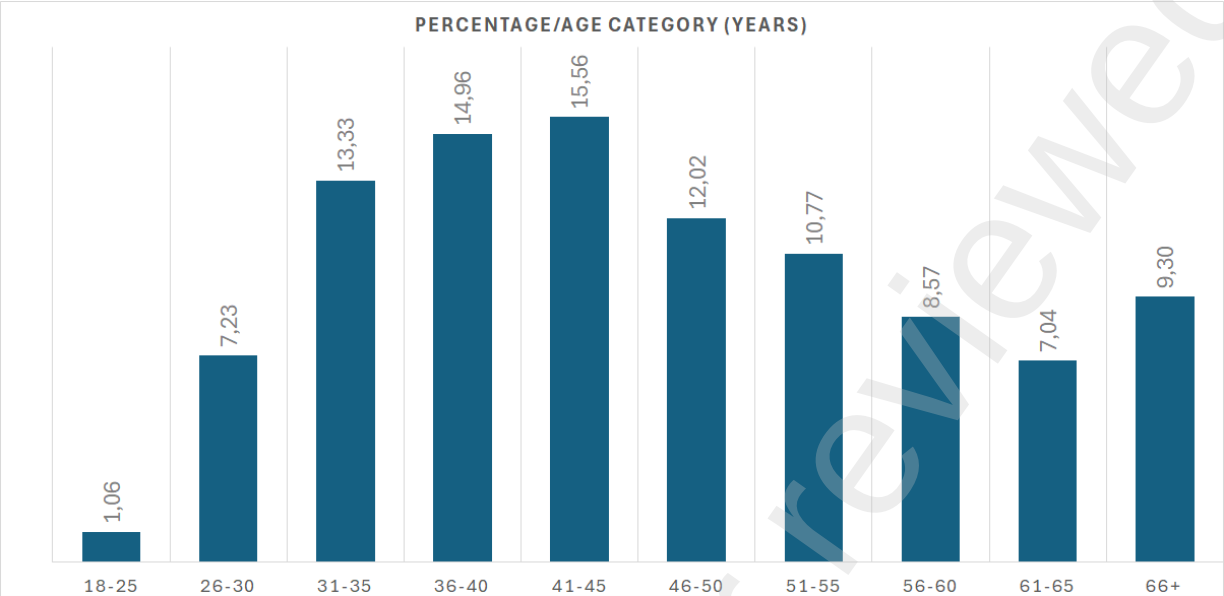
H14b	H11b	The perceived impact of one's disadvantage(s) on everyday life is negatively related to frequency of accepting invitation to peer review relative to frequency of invitations received, with higher impact being associated with a lower ratio of invited vs. accepted peer reviews.
H15a	H12a	The Perceived Disadvantage Index is negatively related to the perceived number of reviews compared to other scholars, with higher disadvantage being associated with a lower perceived amount of accepted peer reviews (missing resources/low confidence hypothesis).
H15b	H12b	The Perceived Disadvantage Index is positively related to the perceived number of reviews compared to other scholars, with higher disadvantage being associated with a higher perceived amount of accepted peer reviews (compensation hypothesis).
H16	H16c (ex.)	Do beliefs about the number of reviews produced and amount of time spent on them compared to comparable scholars differ depending on, (a) WW vs. non-WW nationality, (b) WW vs. non-WW institutional affiliation, and (c) self-identification as a member of a discriminated group?
H17a	H13a	The perceived impact of disadvantage(s) on everyday life is negatively related to the perceived effort during peer review compared to other scholars, with higher disadvantage being associated with a lower effort (missing resources/low confidence hypothesis).
H17b	H13b	The perceived impact of disadvantage(s) on everyday life is positively related to the perceived effort during peer review compared to other scholars, with higher disadvantage being associated with a higher effort (compensation hypothesis).

Note. "ex." stands for an exploratory preregistered question rather than a confirmatory hypothesis.

Supplement S2. Methodological note on Western vs. non-Western countries (preregistered comparison)

Regarding *nationality* and *affiliation*, we hypothesized that scholars whose institutional main affiliation is currently in a Wealthy Western (WW) country are invited to peer review articles more frequently across levels of professional seniority than those in non-WW countries (H1). Moreover, the country of main institutional affiliation and nationality interact, such that scholars with both nationality and affiliation of WW countries are invited to peer-review most frequently, followed by those with non-WW nationalities but WW affiliations, then those with WW nationalities but non-WW affiliations, and finally those with both nationality and affiliation of non-WW countries, who are invited least frequently (H2). We further explored whether those with Western institutional affiliations performed more or less reviews relative to the number of invitations received (H3), and whether the subjective beliefs about how many reviews scholars were performing and how much time they were investing in them relative to comparable peers differed between Western vs. Non-Western nationalities (H4a) and institutional affiliations (H4b).

Supplement S3. Age Distribution of the Sample



Supplement S4. Table with Distribution of Genders Across Career Stages

	Cis female	Cis male	Other
n	1645	1869	37
Undergraduate student	57 / 3.47 %	57 / 3.05 %	1
PhD student	248 / 15.08 %	164 / 8.77 %	8
Post Doc with no Tenure Prospects	234 / 14.22 %	248 / 13.27 %	11
On Tenure Track	141 / 8.57 %	157 / 8.40 %	2
Tenured	889 / 54.04 %	1099 / 58.80 %	15
Retired	53 / 3.22 %	119 / 6.37 %	0
NA	23	25	0

Supplement S5. Results of the Bayesian Analyses

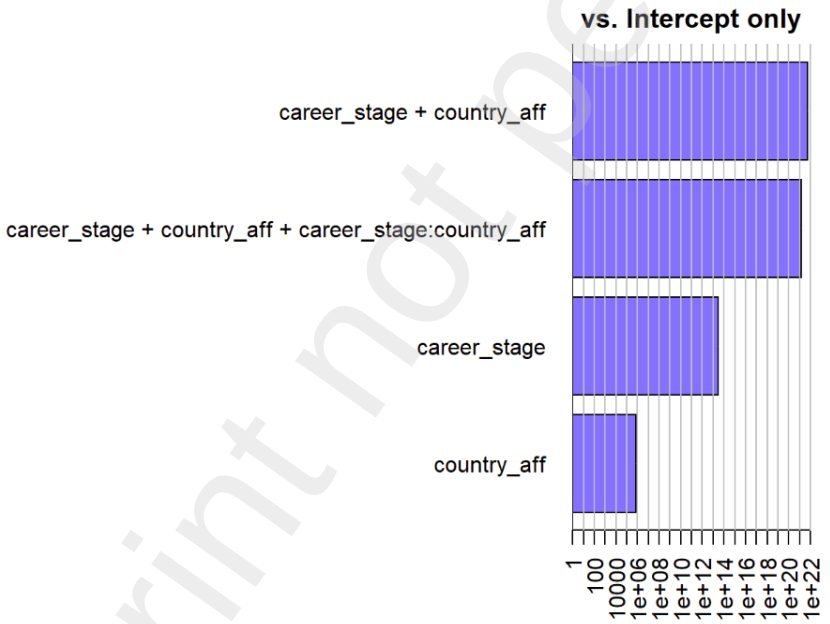
Notes on the results:

- Single variables denote the Bayes Factor for one main effect
- + denotes the Bayes Factor for two main effects
- + and : denote the Bayes Factor for both main effects and an interaction effect

H1 - Country affiliation x number of invites

Model and included variables	Bayes Factor
career_stage	3.04^{e+13}
country_aff	694033.30
career_stage + country_aff	5.71^{e+21}
career_stage + country_aff + career_stage:country_aff	1.38^{e+21}

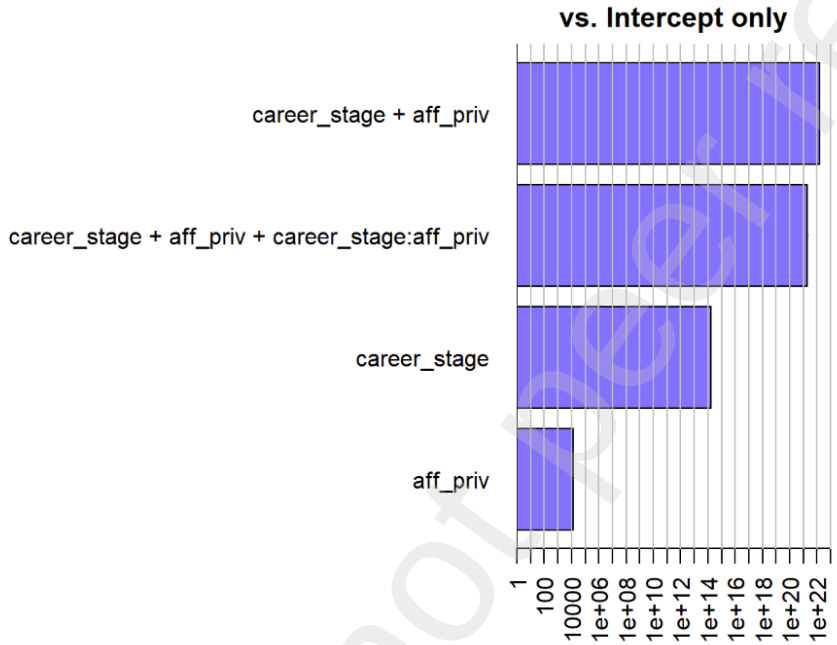
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H1 - Country affiliation x number of invites (privilege variant)

Model and included variables	Bayes Factor
career_stage	$1.59e^{+14}$
aff_priv	13225.82
career_stage + aff_priv	$1.63e^{+22}$
career_stage + aff_priv + career_stage:aff_priv	$2.03e^{+21}$

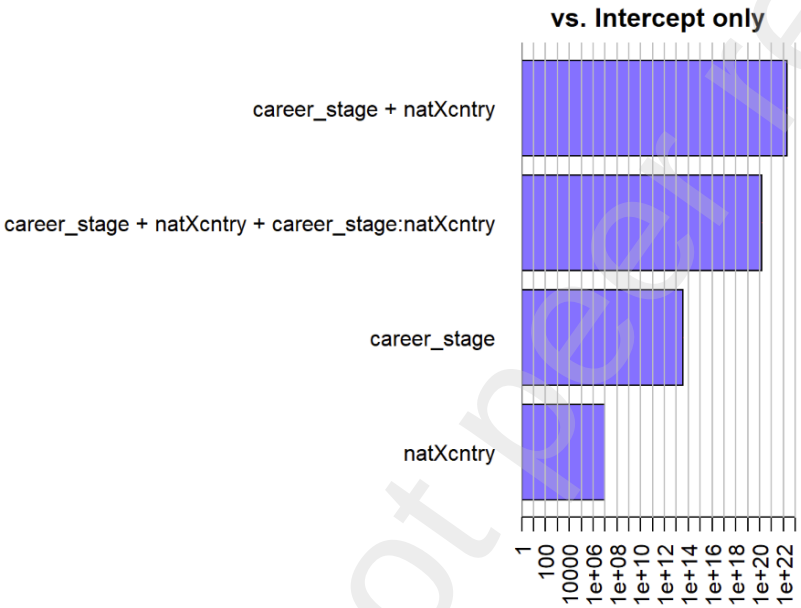
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H2 - Country affiliation x nationality x number of invites

Model and included variables	Bayes Factor
career_stage	$3.77e^{+13}$
natXcntry	$9.72e^{+6}$
career_stage + natXcntry	$1.94e^{+22}$
career_stage + natXcntry + career_stage:natXcntry	$1.44e^{+20}$

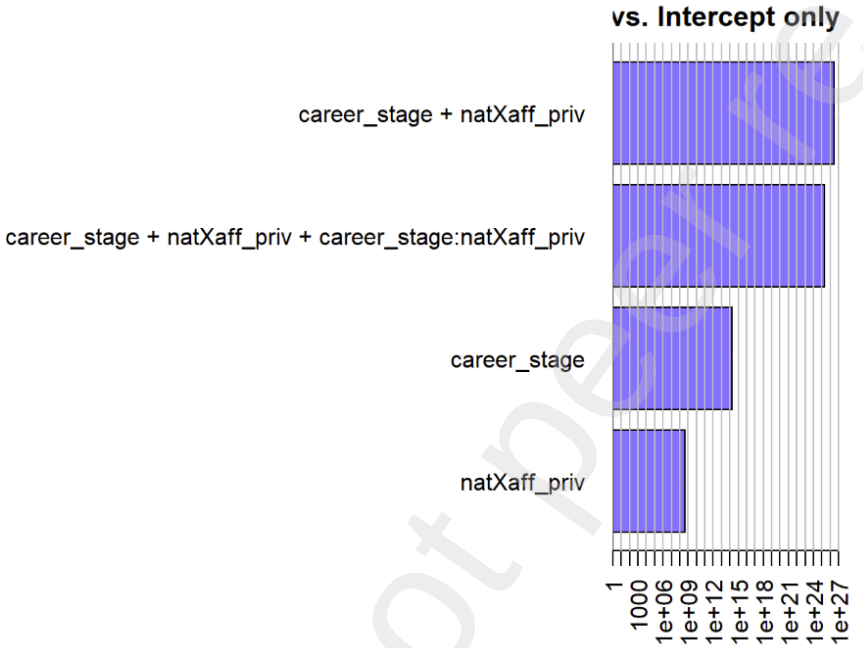
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H2 - Country affiliation x nationality x number of invites (privilege variant)

Model and included variables	Bayes Factor
career_stage	$1.59e^{+14}$
natXaff_priv	$4.12e^{+8}$
career_stage + natXaff_priv	$2.69e^{+26}$
career_stage + natXaff_priv + career_stage:natXaff_priv	$2.12e^{+25}$

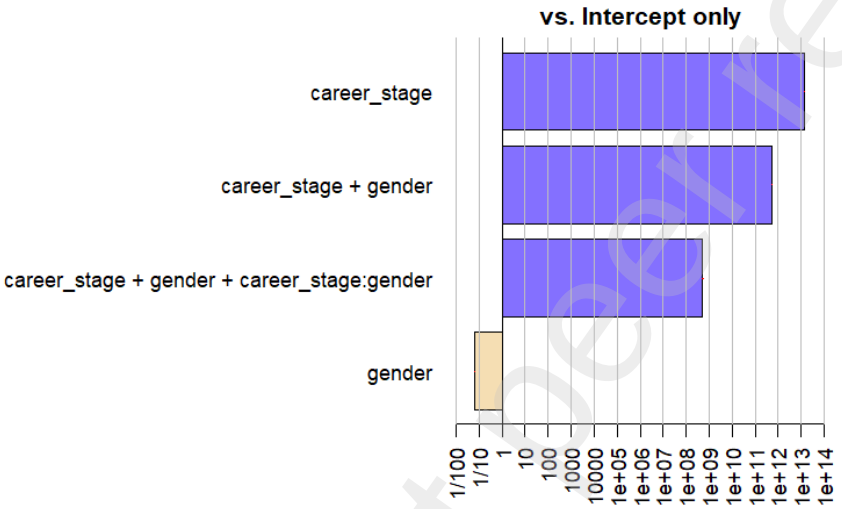
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H3 - Gender x number of invites

Model and included variables	Bayes Factor
career_stage	$1.32e^{+13}$
gender	0.06
career_stage + gender	$5.36e^{+11}$
career_stage + gender + career_stage:gender	$5.24e^{+8}$

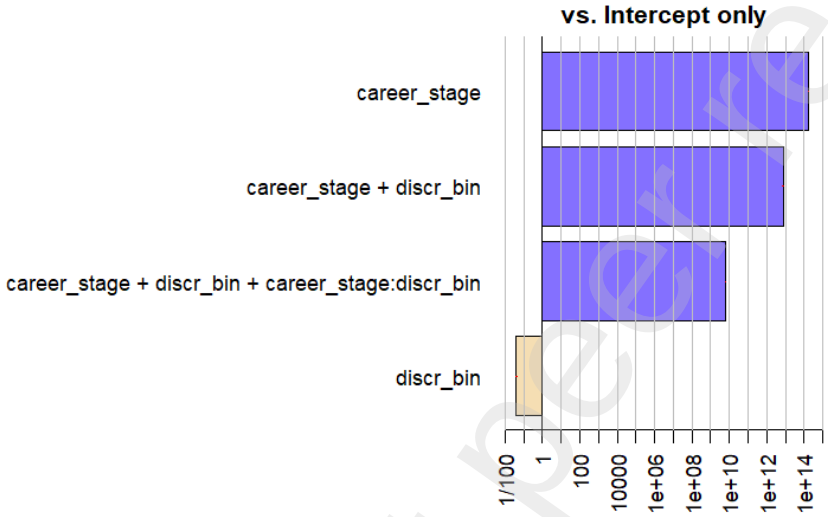
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H4 - Discrimination (binary) x number of invites

Model and included variables	Bayes Factor
career_stage	$1.59e^{+14}$
discr_bin	0.04
career_stage + discr_bin	$7.42e^{+12}$
career_stage + discr_bin + career_stage:discr_bin	$6.11e^{+9}$

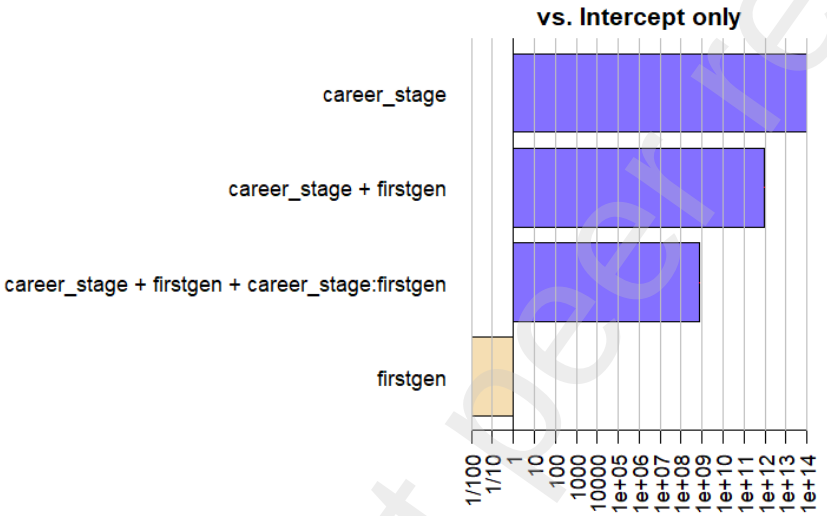
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H5 - First-generation academic status x number of invites

Model and included variables	Bayes Factor
career_stage	$9.31e^{+13}$
firstgen	0.01
career_stage + firstgen	$9.53e^{+11}$
career_stage + firstgen + career_stage:firstgen	$7.71e^{+8}$

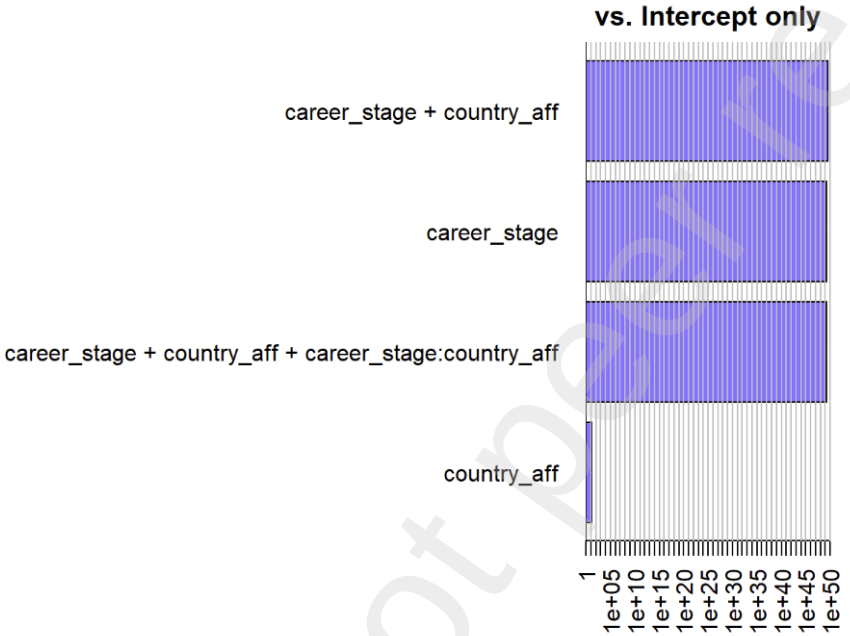
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H6 - Country affiliation x ratio invited vs. accepted

Model and included variables	Bayes Factor
career_stage	$2.01e^{+49}$
country_aff	16.64
career_stage + country_aff	$3.97e^{+49}$
career_stage + country_aff+ career_stage:country_aff	$1.64e^{+49}$

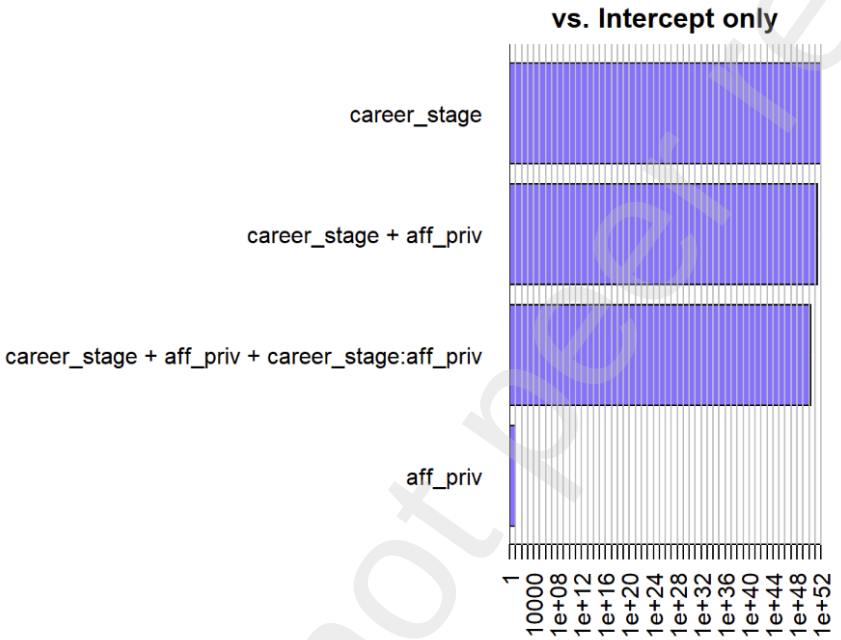
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H6 - Country affiliation x ratio invited vs. accepted (privilege variant)

Model and included variables	Bayes Factor
career_stage	$9.30e^{+51}$
aff_priv	10.48
career_stage + aff_priv	$2.53e^{+51}$
career_stage + aff_priv + career_stage:aff_priv	$2.09e^{+50}$

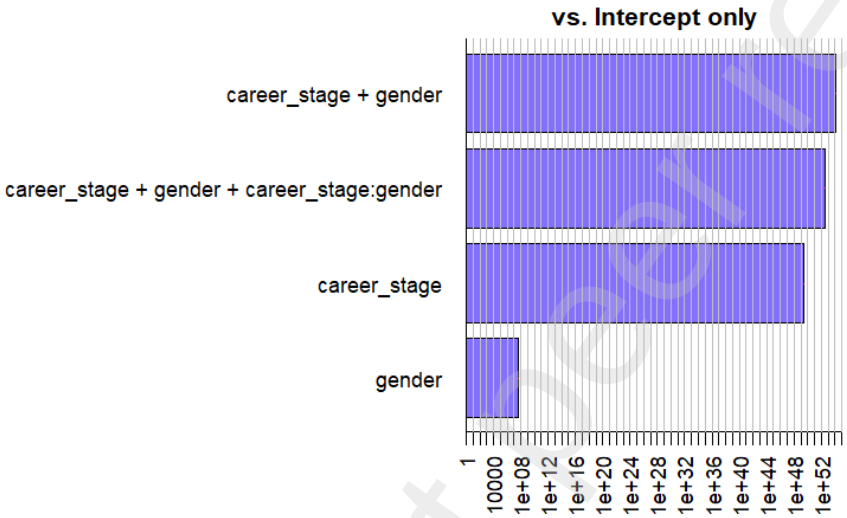
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H7 - Gender x ratio invited vs. accepted

Model and included variables	Bayes Factor
career_stage	$2.73e^{+49}$
gender	$4.95e^{+7}$
career_stage + gender	$1.08e^{+54}$
career_stage + gender + career_stage:gender	$3.84e^{+52}$

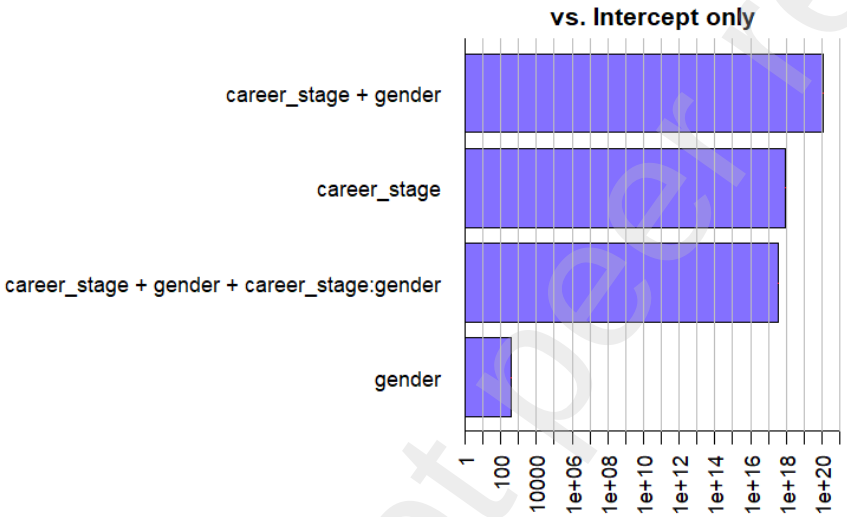
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H7 - Gender x number of reviews

Model and included variables	Bayes Factor
career_stage	$9.54e^{+17}$
gender	379.41
career_stage + gender	$1.13e^{+20}$
career_stage + gender + career_stage:gender	$3.73e^{+17}$

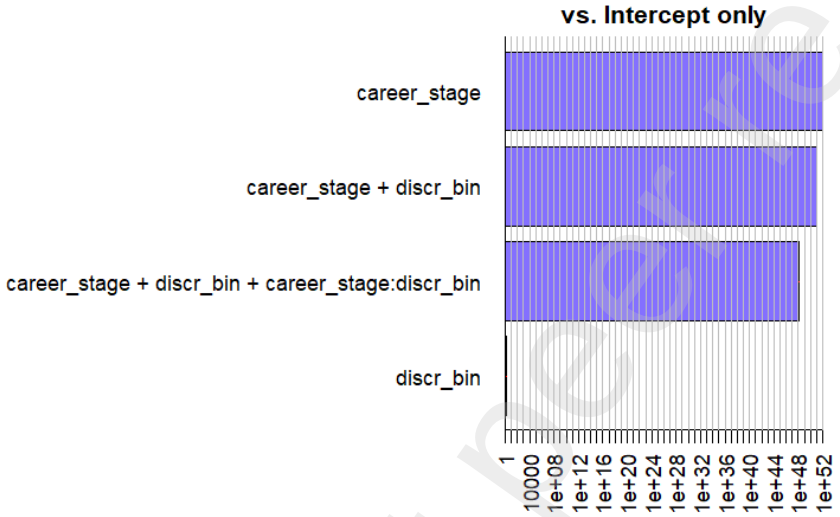
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H8 - Discrimination (binary) x ratio invited vs. accepted

Model and included variables	Bayes Factor
career_stage	$9.30e^{+51}$
discr_bin	1.41
career_stage + discr_bin	$1.17e^{+51}$
career_stage + discr_bin + career_stage:discr_bin	$1.46e^{+48}$

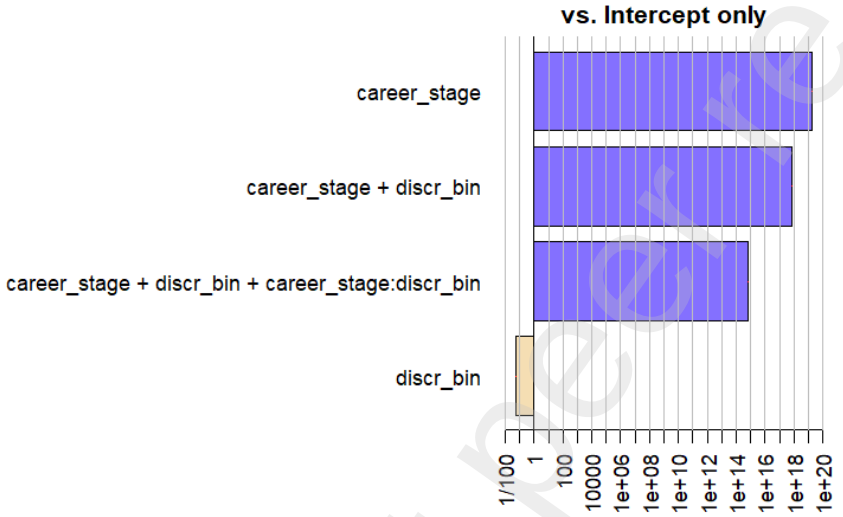
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H8 - Discrimination (binary) x number of reviews

Model and included variables	Bayes Factor
career_stage	$1.72e^{+19}$
discr_bin	0.05
career_stage + discr_bin	$6.70e^{+17}$
career_stage + discr_bin + career_stage:discr_bin	$6.67e^{+14}$

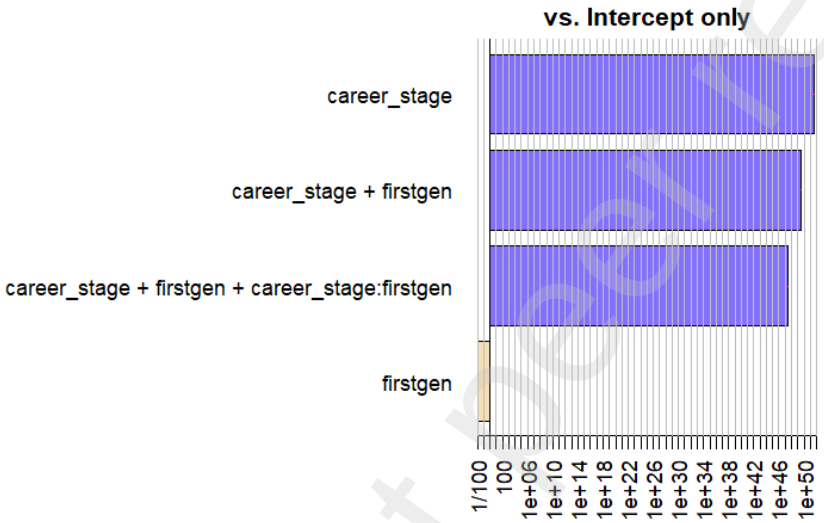
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H9 - First-generation academic status x ratio invited vs. accepted

Model and included variables	Bayes Factor
career_stage	$3.42e^{+51}$
firstgen	0.01
career_stage + firstgen	$3.02e^{+49}$
career_stage + firstgen + career_stage:firstgen	$2.85e^{+47}$

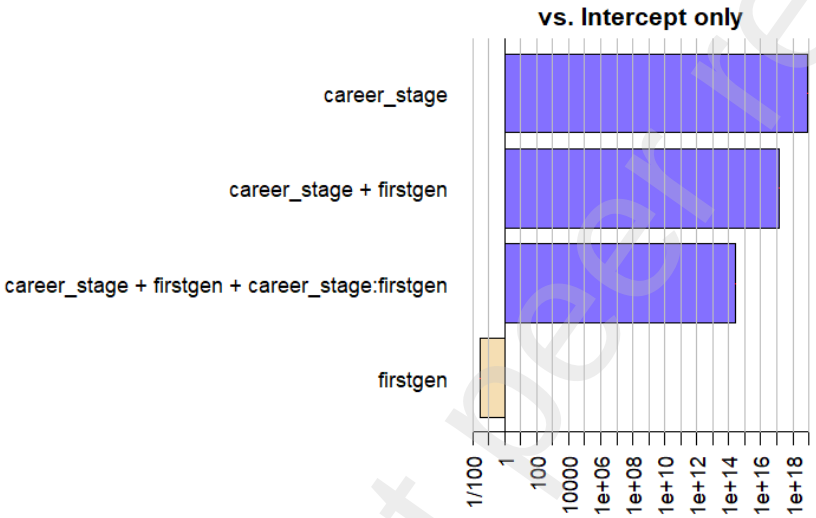
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H9 - First-generation academic status x number of reviews

Model and included variables	Bayes Factor
career_stage	$8.57e^{+18}$
firstgen	0.03
career_stage + firstgen	$1.42e^{+17}$
career_stage + firstgen + career_stage:firstgen	$2.48e^{+14}$

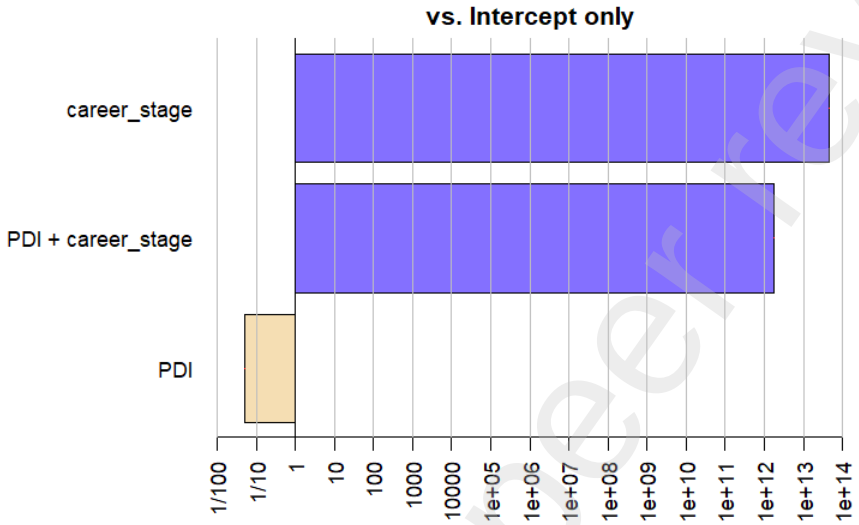
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H10 - Perceived Disadvantage Index (PDI) x number of invites

Model and included variables	Bayes Factor
PDI	0.05
career_stage	$4.45e^{+13}$
PDI + career_stage	$1.73e^{+12}$

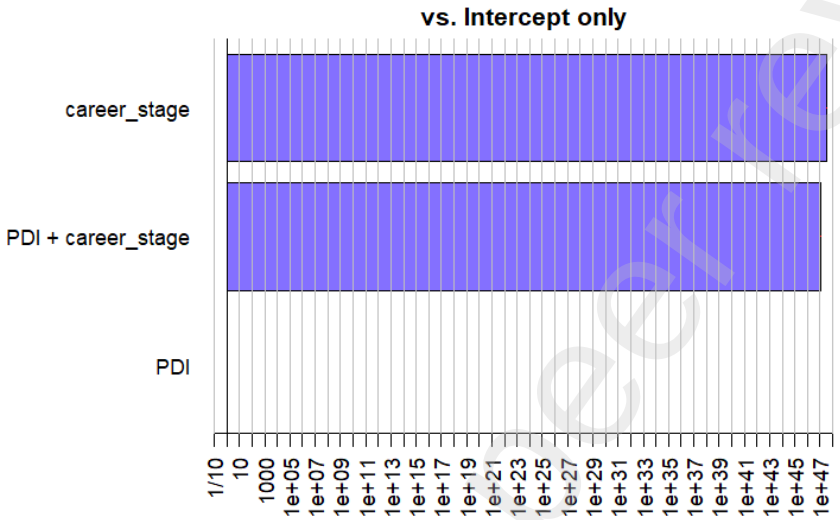
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H11a - Perceived Disadvantage Index (PDI) x ratio invited vs. accepted

Model and included variables	Bayes Factor
PDI	0.98
career_stage	$3.29e^{+47}$
PDI + career_stage	$1.09e^{+47}$

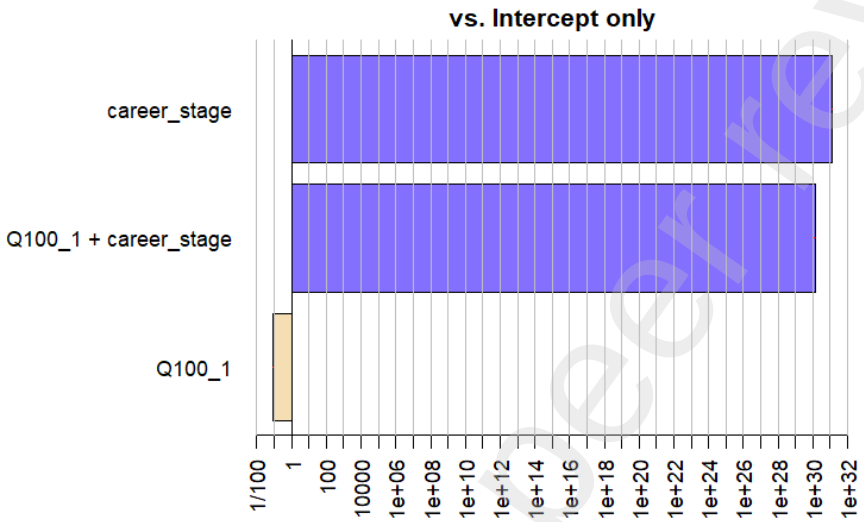
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H11b - Discrimination impact x ratio invited vs. accepted

Model and included variables	Bayes Factor
Impact (Q100_1)	0.09
career_stage	$1.36e^{+31}$
Impact (Q100_1) + career_stage	$1.29e^{+30}$

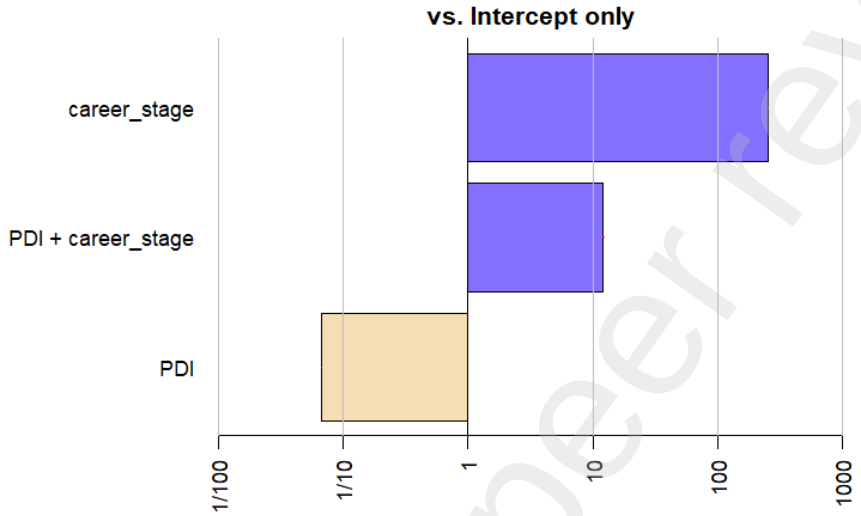
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H12a - Perceived Disadvantage Index (PDI) x number of acceptances

Model and included variables	Bayes Factor
PDI	0.07
career_stage	253.76
PDI + career_stage	12.05

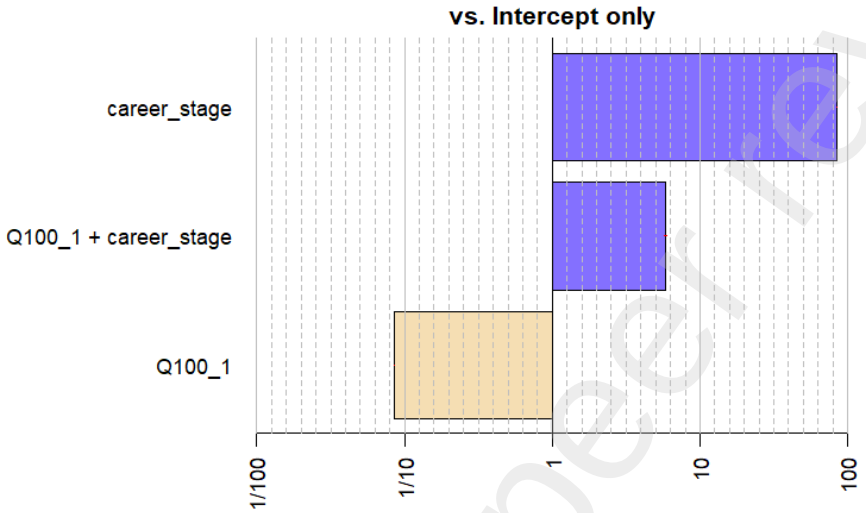
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H12b - Discrimination impact x number of acceptances

Model and included variables	Bayes Factor
Impact (Q100_1)	0.09
career_stage	84.13
Impact (Q100_1) + career_stage	5.82

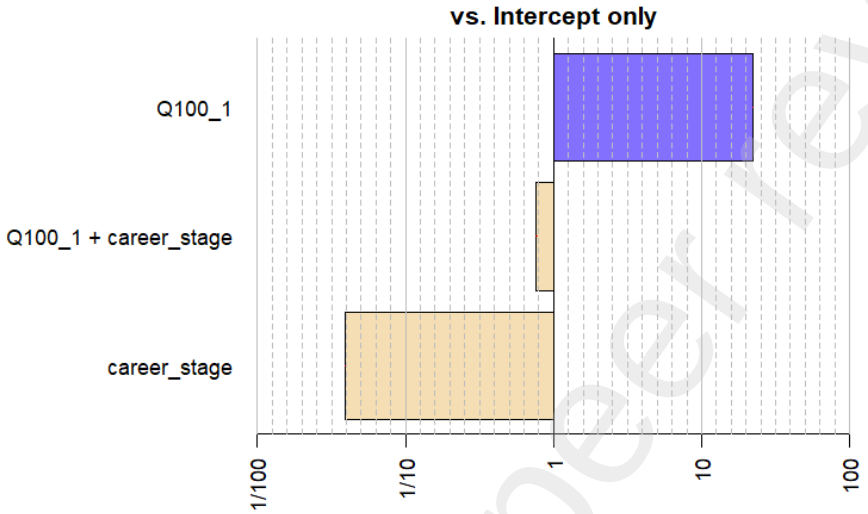
Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



H13 - Discrimination impact x effort

Model and included variables	Bayes Factor
Impact (Q100_1)	22.07
career_stage	0.04
Impact (Q100_1) + career_stage	0.76

Note. All models tested against a model with only the intercept. Bayes factor type: BFlinearModel. Evidence for evidence for H_1 vs. H_0. Career stage was included as a control variable.



Supplement S6. ANOVAs and Tukey's HSD contrasts comparing reasons for review request rejection across academic prestige groups.

						Tukey's HSD (p)					
	1	2	3	4	ANOVA	2-1	3-1	4-1	3-2	4-2	4-3
M(SD) Lack of time	4.25(1.06)	4.18(1.14)	4.03(1.22)	4.00(1.26)	$F(3, 3652) = 12.88$, $p < .001$	-.07(.889)	-.22(.012)	-.25(<.001)	-.15(.514)	-.18(.218)	-.03(.980)
M(SD) Dislike the journal	3.34(1.30)	3.42(1.32)	3.16(1.34)	2.96(1.40)	$F(3, 3639) = 22.83$, $p <.001$	.07(.904)	-.19(.115)	-.38(<.001)	-.26(.179)	-.45(<.001)	-.19(.097)
M(SD) Recent Rejection	2.14(1.14)	2.38(1.20)	1.87(1.11)	2.03(1.17)	$F(3, 3639) = 9.51$. $p <.001$	.23(.057)	-.27(.001)	-.11(.037)	-.50(<.001)	-.34(.001)	.16(.120)
M(SD) No interest in journal	2.42(1.25)	2.29(1.20)	2.20(1.22)	2.05(1.15)	$F(3, 3643) = 26.21$, $p <.001$	-.13(.539)	-.23(.012)	-.38(<.001)	-.10(.825)	-.25(.053)	-.15(.189)
M(SD) Dubious journal reputation	4.70(.84)	4.68(.92)	4.71(.86)	4.54(1.03)	$F(3, 3641) = 8.67$, $p <.001$	-.01(.999)	.01(.994)	-.16(<.001)	.03(.992)	-.15(.211)	-.17(.016)
M(SD) Lack of expertise	4.62(.76)	4.63(.85)	4.60(.78)	4.56(.89)	$F(3, 3641) = 1.58$, $p = .195$	-	-	-	-	-	-
M(SD) Fundamental disagreement	2.67(1.12)	2.88(1.18)	2.64(1.22)	2.87(1.27)	$F(3, 3633) = 9.08$, $p <.001$	.21(.137)	-.03(.982)	.20(<.001)	-.24(.165)	-.01(.999)	.23(.011)
M(SD) Recognised the authors	3.47(1.25)	3.36(1.25)	3.53(1.26)	3.52(1.34)	$F(3, 3637) = 1.01$, $p = .386$	-	-	-	-	-	-

Note. 1 = prestigious nationality and prestigious affiliation, 2 = prestigious nationality and non-prestigious affiliation, 3 = non-prestigious nationality and prestigious affiliation, 4 = non-prestigious nationality and non-prestigious affiliation

Supplement S7. Topic modelling results for reasons researchers declined their last review.

There were 13 Topics. We describe the content of these topics here.

Topic 1 - Conflicts of Interest

Participants note they recognise the authors and have published with them or know them in a personal capacity, and decline to review to ensure objectivity.

Topic 2 - Time is Short

Topic 2 is a clearly defined topic, wherein participants indicate that timelines for the review turnaround were too short, causing them to decline the invitation.

Topic 3 - Reviewed Elsewhere

Topic 3 is a narrow topic wherein participants describe they recently reviewed the paper in question for another journal. Often participants note they recommended rejection of the paper for the previous journal.

Topic 4 - Review Frequency and Unpaid Work

Topic 4 encompasses both issues around capacity, usually due to review invitation frequency. They also note it is unpaid labour and that many invitations come from predatory journals.

Topic 5 - Lack of Time

Topic 5 is predominantly a single-issue topic; many responses are simply “lack of time”. Some participants describe the reasons why they lack time to review.

Topic 6 - Poorly Written Abstracts

Topic 6 centres around the quality of the writing of the abstract and how that is indicative of the quality of the paper. Some also mention that the abstract is difficult to get through due to the poor writing.

Topic 7 - Topic Expertise

Topic 7 are mostly short responses stating the participant did not have expertise in the subject area of the paper.

Topic 8 - Methodological Expertise

Topic 8 is very similar to Topic 7, but emphasises that the topic is not the issue, but rather the methodology is not within their field of expertise.

Topic 9 - Journal-Related Concerns

This topic straddles two issues: too many requests from the same journal and spam requests from predatory journals.

Topic 10 - Lack of Interest

Topic 10 is a broadly defined topic with often lengthy reasons for declining to review. Participants mention declining to review papers they have no interest in reading, articles that should have been obvious desk rejects for the editor, or if the journal has a poor reputation.

Topic 11 - Open Access & Not-For-Profit

In Topic 11, participants detail how they prefer to review for not-for-profit or open access journals. They do not like to review for free for private companies.

Topic 12 - Shady Journals

In Topic 12, participants emphasise poor journal reputation and singular profit motives.

Topic 13 - I Do Enough & The Editor Didn't Do Their Homework

Topic 13 is a wide-ranging topic that emphasises effort of the participant (e.g., "I review enough papers already") and a lack of understanding on the part of the editor what the participant's field of work is.

Group	Topic	Words	Frequency	Proportion within group
Prestigious	1	conflict, author, authors, know, paper	56	0.131
	2	short, review, reviews, editors, given	35	0.082
	3	reviewed, journal, paper, previously, rejected	39	0.091

	4	invitations, invitation, accept, decline, journal	30	0.070
	5	lack, lack time, reason, work, main	25	0.058
	6	abstract, quality, paper, ask, read	32	0.075
	7	expertise, outside, feel, area, topic	13	0.030
	8	expertise, methodology, research, area expertise, area	18	0.042
	9	requests, asked, reviews, times, number	22	0.051
	10	decline, comments, reject, poor, review	23	0.054
	11	free, access, journals, open, refuse	15	0.035
	12	journal, journals, open, predatory, access	15	0.035
	13	papers, editorial, journals, review, reviewing	16	0.037
	Unassigned	review, peer, journals, articles, peer review	90	0.210
Non-Prestigious	1	author, authors, conflict, know, working	21	0.092
Non-Prestigious	2	short, review, given, time review, journal	23	0.100
Non-Prestigious	3	reviewed, journal, different, paper, previously	16	0.070

Non-Prestigious	4	invitations, invitation, accept, did, just	22	0.096
Non-Prestigious	5	lack time, lack, work, reason, topic	18	0.079
Non-Prestigious	6	abstract, feel, quality, research, wrong	10	0.044
Non-Prestigious	7	expertise, area, outside, outside expertise, topic	22	0.096
Non-Prestigious	8	research, methodology, familiar, area, expertise	14	0.061
Non-Prestigious	9	requests, reviewing, number, times, asked	8	0.035
Non-Prestigious	10	reject, likely, field, decline, refuse	6	0.026
Non-Prestigious	11	access, refuse, open, practice, journals	9	0.039
Non-Prestigious	12	journal, predatory, access, open, journals	9	0.039
Non-Prestigious	13	send, papers, review, journals, topic	6	0.026
Non-Prestigious	Unassigned	peer, review, editor, articles, peer review	45	0.197

Note. Data grouped by any prestige (affiliation/nationality) or none. The topics are in descending order of total frequency across groups. The topic words differ slightly between groups to show that they represent somewhat different perspectives within the topic.