Nobel and novice: Author prominence affects peer review

Jürgen Huber[a,1] 🔟, Sabiou Inoua[b] 🔟, Rudolf Kerschbamer[c] 🔟, Christian König-Kersting[a] 🔟, Stefan Palan[d] 🔟, and Vernon L. Smith[b,1] 🔟

Peer review is a well-established cornerstone of the scientific process, yet it is not immune to biases like status bias, which we explore in this paper. Merton described this bias as prominent researchers getting disproportionately great credit for their contribution, while relatively unknown researchers get disproportionately little credit [R. K. Merton, *Science* 159, 56–63 (1968)]. We measured the extent of this bias in the peer-review process through a preregistered field experiment. We invited more than 3,300 researchers to review a finance research paper jointly written by a prominent author (a Nobel laureate) and by a relatively unknown author (an early career research associate), varying whether reviewers saw the prominent author's name, an anonymized version of the paper, or the less-well-known author's name. We found strong evidence for the status bias: More of the invited researchers accepted to review the paper when the prominent name was shown, and while only 23% recommended "reject" when the prominent researcher was the only author shown, 48% did so when the paper was anonymized, and 65% did when the little-known author was the only author shown. Our findings complement and extend earlier results on double-anonymized vs. single-anonymized review [R. Blank, *Am. Econ. Rev.* 81, 1041–1067 (1991); M. A. Ucci, F. D'Antonio, V. Berghella, *Am. J. Obstet. Gynecol. MFM* 4, 100645 (2022)].

peer review | scientific method | double-anonymized | status bias

Peer review has been the key method for research validation since the first scientific journals appeared some 300 y ago (1). For researchers—and especially for young scientists, who must excel for scientific advancement—it is crucial that this process be fair and impartial. Merton (2), however, argued that "eminent scientists get disproportionately great credit for their contribution to science while relatively unknown scientists tend to get disproportionately little credit for comparable contributions." Alluding to the Gospel according to Matthew 25:29, Merton termed this pattern of the misallocation of credit for scientific contributions "the Matthew effect in science," while others called it "status bias" (3).

We measured the strength of the positive ("eminent scientists get disproportionately great credit") and the negative ("unknown scientists tend to get disproportionately little credit") components of this bias in a preregistered field experiment. Specifically, we addressed two related research questions (RQs) derived from the literature summarized in *SI Appendix*: First, is there a status bias in potential reviewers' propensity to accept the invitation to review a paper (RQ1)? Second, is there a status bias in their evaluation of the manuscript (RQ2)? To address both RQs, we used a new and unpublished research article that covered a broad range of topics in finance (to be of interest and relevance to many potential reviewers), which had been jointly written by a prominent scientist and a relatively unknown scientist. The prominent scientist is V. L. Smith, the 2002 laureate of the Nobel Memorial Prize in Economic Sciences (54,000 Google Scholar citations as of December 2021), and the relatively unknown scientist is S. Inoua, an early career research associate (42 Google Scholar citations). Both were affiliated with the Economic Science Institute at Chapman University at the time the paper was written. Holding the affiliation constant was important because previous research documented an impact of institutional affiliation on the outcome of the review process under single-anonymized evaluation (4–6). While both authors have the same gender (7), we could not hold race and name origin constant, since the Nobel laureate only had one project to write a paper about with a suitable junior researcher at the time.

The manuscript that Smith and Inoua coauthored was submitted to the *Journal of Behavioral and Experimental Finance* (*JBEF* henceforth; published by Elsevier). The journal's editor, who is a coauthor of the present paper, then sent review invitations for the paper to more than 3,300 potential reviewers. Across five conditions, we varied whether a corresponding author name was provided in the invitation email and on the manuscript, only on the manuscript, or not at all. The second author's name was never shown. Mentioning only the corresponding author is practiced by, for example, the publisher Wiley (e.g., for the *German Economic Review* and the *Journal of Public Economic Theory*).

Significance

Peer review is vital for research validation, but may be affected by "status bias," the unequal treatment of papers written by prominent and less-well-known authors. We studied this bias by sending a research paper jointly authored by a Nobel Prize laureate and a relatively unknown early career research associate to many reviewers. We varied whether the prominent author's name, the relatively unknown author's name, or no name was revealed to reviewers. We found clear evidence for bias. More than 20% of the reviewers recommended "accept" when the Nobel laureate was shown as the author, but less than 2% did so when the research associate was shown. Our findings contribute to the debate of how best to organize the peer-review process.

We designate the conditions by two capital letters, with the first representing the invitation email and the second the manuscript. In condition LL (L for low prominence), the less prominent author appeared both in the invitation mail and on the manuscript; in condition AA (A for anonymized), neither the invitation email nor the manuscript showed an author name; in condition HH (H for high prominence), the prominent author appeared both in the invitation email and on the manuscript; and in conditions AL and AH, the invitation email was anonymized, but the respective corresponding author's name appeared on the manuscript proper. Importantly, neither the paper nor a draft thereof was posted online or presented anywhere before our data collection had been completed; hence, reviewers were unable to learn the author names by searching the internet for the paper's title or abstract.

To assess the impact of author prominence on the willingness to review the paper, we compared reviewers' decisions to accept or decline the invitation between the conditions with anonymized emails (AL, AA, and AH pooled) and those with nonanonymized emails (LL and HH). To test the impact of author prominence on the assessment of the paper, we compared reviewers' publication recommendations for manuscripts that showed the corresponding author's name (conditions AL and AH) to those for manuscripts that did not show a name (condition AA). Since the invitations for these three conditions were anonymized, we observed the effect of author prominence on the recommendation decision without possible confounds caused by selection effects in the invitation stage (8). The data is available online (9).

Willingness to Review

In conditions AL, AA, and AH, no author name was given in the invitation to review the paper sent by the editor of *JBEF*; in condition LL, we added the line, "Corresponding author: Sabiou Inoua," and in condition HH, we added the line, "Corresponding author: Vernon L. Smith." The review invitation included the title and abstract of the paper, but did not allow accessing the full manuscript. We compared review-invitation acceptance rates in the anonymized setting of conditions AL, AA, and AH with those in LL and HH (Table 1 and *SI Appendix*, Table S1). Acceptance rates were based on the 2,611 researchers that responded to our invitation email by accepting or declining. (See *SI Appendix*, Table S12 for a full attrition analysis. As is policy at *JBEF*, each reviewer who submitted a report received 50 US dollars as a "token of appreciation.")

Based on results in the related literature—summarized in the preregistration document and in *SI Appendix*—we derived several hypotheses. Our preregistered ex ante hypothesis regarding the review-invitation acceptance rates had two components: The positive component (H1+) predicted that this rate was higher in the condition where the more prominent author was mentioned in the invitation letter as corresponding author than in the conditions where no name was mentioned in the invitation letter. The negative component (H1−) predicted that this rate would be lower in the condition where the less prominent author is mentioned in the invitation letter as corresponding author than in the conditions where no name is mentioned in the invitation letter. We found that the share of researchers that accepted the invitation varied with the setting, with 28.5% of those who responded in condition LL accepting the invitation listing Inoua as the corresponding author; 30.7% across conditions AL, AA, and AH accepting the anonymized invitation to review; and 38.5% in condition HH accepting the invitation showing Nobel laureate Smith as the corresponding author. While the differences between the acceptance rate in HH and the acceptance rates in each of the other two settings (AL, AA, and AH pooled; and LL) were highly significant, the difference between the anonymized conditions (pooled) and LL was not. Thus, our data are in line with Hypothesis H1+, but we get a null result on Hypothesis H1−. Alternatively, using acceptance rates based on the invitations sent (rather than on responses received) did not change our results qualitatively: Acceptance rates dropped from 28.5%, 30.7%, and 38.5% to 22.3%, 24.3%, and 31.2%, respectively, and, again, the differences between the acceptance rate in HH and the acceptance rates in each of the other two settings were highly significant, while the difference between the anonymized settings and LL was not.

Our results thus documented statistically significant evidence for the positive component of the status bias at the invitation-acceptance stage (i.e., the invitation showing the more prominent author being accepted more often than the invitation showing no author name), but did not find such evidence for the negative component (i.e., the invitation showing the less prominent author being accepted less often than the invitation showing no author name).

While the data we gathered thus permitted us to clearly document the positive status bias at the review-invitation stage, our study was not designed to elucidate the motivational drivers of this bias. Potential reviewers may have been more willing to accept an invitation to review a paper written by a prominent author because they expected, e.g., more novel insights, greater learning opportunities, or a lower probability of having to read and evaluate a low-quality paper. We hope that future research will be able to shed light on this question.

Manuscript Assessment

Upon accepting the email invitation, potential reviewers were brought to a consent website. This website informed them that their review, if they were to agree to submit one, would be part

Table 1. Invitations

	Low (LL)	Anonymized (AL, AA, AH)	High (HH)	Total	*P*
Invitations sent	781	2,011	507	3,299	
Responses received	610	1,591	410	2,611	
Invitations accepted	174	489	158	821	
Acceptance rate, %	28.52	30.74	38.54	31.44	
Anon. vs. Low					0.3243
Anon. vs. High					0.0031
Low vs. High					0.0011

Shown are the number of review invitations sent, the number of replies received (declined or accepted), the number of invitations accepted, and the fraction of invitations accepted when the review invitation listed the low-prominence author (condition LL), no corresponding author (AL, AA, and AH), or the high-prominence author (HH). Two-sided Fisher's exact tests of invitation responses between conditions were performed.

of a scientific study, without revealing details of the experimental design. Reviewers then had to actively choose whether they wanted to proceed with reviewing the paper or whether, in light of this information, they preferred to decline the invitation at this stage. Across all conditions, 81.2% gave their consent, with no significant differences between conditions (Anon. [for anonymized] vs. Low: $P = 0.2655$; Anon. vs. High: $P = 0.550$; Low vs. High: $P = 0.163$; two-sided Fisher's exact tests, corrected α-threshold = 0.0167). Subsequently, reviewers who gave their consent received the manuscript. Across our five conditions, reviewers submitted 534 written review reports (AL: 101, AA: 110, AH: 102, LL: 114, and HH: 107), enabling us to reach our preregistered target of 100 reports per condition. For the analysis of manuscript assessments, we focused on the 313 reports received in conditions AL, AA, and AH. These conditions all had an anonymized invitation, thus allowing for a clean identification of the effect of author prominence on the evaluation of the manuscript without being confounded by selection at the invitation stage. We present all results (based on all 534 reports) in *SI Appendix*, Table S2 and discuss the effects of selection on manuscript assessment below and in *SI Appendix*.

Arguably, the most important outcome of the review process is the reviewers' final recommendations to the editor. Reviewers were asked to recommend either "reject," "major revision," "minor revision," or "accept" when submitting their reports. Our preregistered ex ante hypothesis regarding the recommendation given by the reviewers had two components: The positive component (H2+) predicted that the assessment of the paper would be more favorable in the condition where the more prominent author appeared as the corresponding author of the paper than in the condition where no name was given. The negative component (H2−) predicted that the assessment of the paper would be less favorable in the condition where the less prominent author appeared as the corresponding author of the paper than in the condition where no name was given. We followed our preanalysis plan and compared the distributions of recommendations given by the reviewers using Mann–Whitney U tests. We found highly significant differences between conditions and report them in Fig. 1.

When an editor decides to follow the reviewers' assessments, "reject" recommendations end the publication process at the journal, while the other recommendations allow the manuscript to continue to the revision stage. While *JBEF* rejects many manuscripts before or as a result of the first round of review, manuscripts that receive a "revise" decision in the first round historically had a 93% probability of eventually getting published in the journal (2018 through 2020). We find stark differences in the rejection recommendations across the three conditions: While 65.4% of the reviewers recommended "reject" when shown the less prominent author, this number was 48.2% in the anonymized

version of the manuscript and 22.6% when the prominent author was shown. The pairwise differences between conditions were all significant: The P value for AL vs. AA is 0.0120 (two-sided test of proportions, $z = 2.512$), that for AA vs. AH is 0.0001 ($z = 3.885$), and that for AL vs. AH is <0.0001 ($z = 6.144$; the α-threshold corrected for three tests was 0.05). Thus, compared to the fully anonymized review, the less prominent author clearly faced a lower, and the more prominent author a higher, chance of passing the first round of peer review. The pattern of "reject" recommendations matches both hypotheses, H2+ and H2−.

Aggregating the two most positive categories, "accept" and "minor revision," of the recommendation spectrum to a single category showed that only 9.9% of the reviewers recommended a minor revision or an outright accept when shown the less prominent author, while 23.6% gave one of these recommendations in the anonymized version, and 58.8% did so when shown the prominent author. The pairwise differences between conditions were all highly significant: The P value for AL vs. AA was 0.008 (two-sided test of proportions, $z = -2.650$), that for AA vs. AH was <0.0001 ($z = -5.213$), and that for AL vs. AH was <0.0001 ($z = -7.332$; α-threshold corrected for three tests was 0.05). Thus, we also observed effects that are in line with hypotheses H2+ and H2− for the two most positive recommendation categories.

In summary, we found strong evidence for the positive component of the status bias, in that the prominent author received far more favorable assessments (and, in particular, fewer rejections) than did the anonymized version of the manuscript, and we also found strong evidence for the negative component of the status bias, in that the less prominent author received significantly less favorable assessments (and, in particular, more rejections) than did the anonymized version of the same paper.

In case that working-paper versions of a manuscript have been published prior to journal submission, these results suggest that less prominent authors might profit from anonymizing their manuscripts by giving them a new title and rewriting the abstract (so that the papers are more difficult to find on the web) and submitting them to journals with double-anonymized review. This conclusion is, however, premature. In *SI Appendix*, we show that those reviewers that accepted the invitation to review the paper, despite having been informed that the corresponding author was the less prominent researcher (in condition LL), were milder in their judgements than reviewers that responded to the anonymized invitation. This self-selection effect benefited the less prominent author and (partly) counteracted the negative component of the status bias of the less prominent author's manuscript having been evaluated less favorably. Although the net effect still pointed in the direction of less prominent authors benefiting from fully anonymized review, the difference from single-anonymized
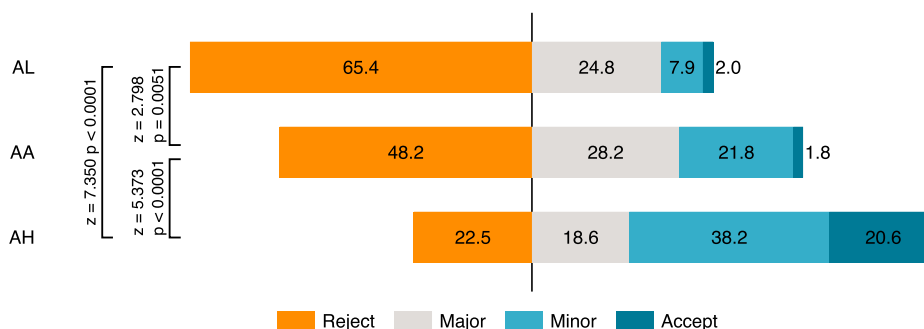

Fig. 1. Recommendation percentages by condition. In conditions AL and AH, the invitation email was anonymized, but the respective corresponding author's name appeared on the manuscript, while in AA, both the invitation and the paper were anonymized. The tests are pairwise, two-sided Mann–Whitney U tests.

review was no longer statistically significant for the less prominent author (while the reversed difference for the prominent author remained highly significant).

In addition to giving a recommendation, reviewers were also requested to assess the paper by answering six questions routinely asked of reviewers for Elsevier journals. These questions inquired whether the reviewer considered 1) the subject worthy of investigation; 2) the information new; 3) the conclusions supported by the data; 4) the manuscript suitable for *JBEF*; 5) the organization of the manuscript appropriate; and 6) the figures and tables appropriate. All six questions were answered on a Likert scale, ranging from 1) strongly disagree to 5) strongly agree.

Fig. 2 shows the distribution of ratings and individual test statistics for the six items addressing the quality of the manuscript.

For all six items, we observed a significant upward shift of the distribution of ratings when the manuscript showed the highly prominent corresponding author compared to both the anonymized manuscript and the manuscript showing the relatively unknown author (AH vs. AA/AL: all $z \geq 4.260$ and all $P \leq 0.0001$, two-sided Mann–Whitney U tests). Thus, the positive component of the status bias affected not only final recommendations, but also the individual components of the manuscript-quality assessments. Notably, while all comparisons had the expected sign, we found statistically significant evidence for the negative component of the status bias only in items 2) on the information being new, 3) on the conclusions of the paper being supported by the data, and 6) on the figures and *SI Appendix* being appropriate.

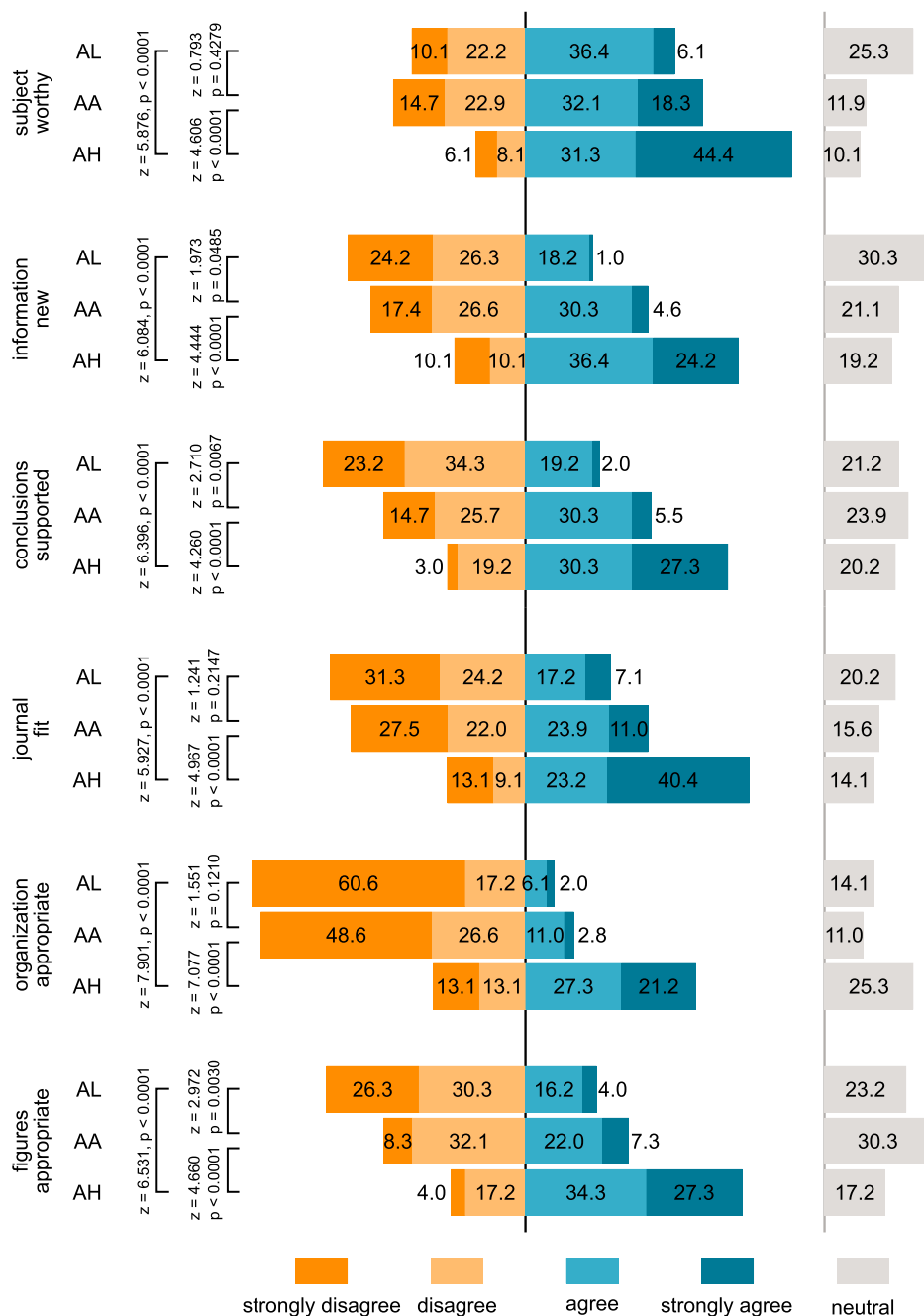**Fig. 2.** Responses to reviewer questionnaire items 1 through 6. We plot the percentage of neutral responses on the right-hand border of the figure. For each item, we conducted pairwise, two-sided Mann–Whitney U tests across conditions. In conditions AL and AH, the invitation email was anonymized, but the respective corresponding author's name appeared on the manuscript, while in AA, both the invitation and the paper were anonymized.

Limitations

We want to discuss three limitations to our study in the interest of transparency. First, to comply with the European Union's (EU's) *General Data Protection Regulation* (GDPR), as well as with university ethics guidelines, we informed researchers who clicked the accept link in the invitation to review that their review would both serve to inform the editor about the manuscript's publishability and form part of a research study into the peer-review process. Reviewers then had to actively consent to these conditions to continue. This potentially raises two issues regarding the quality of our data: 1) Researchers who gave their consent may have differed in some characteristic of interest from those who withheld their consent—a selection problem—and 2) researchers who gave their consent may have been influenced in their evaluations by the fact that they were aware that they were part of a research study—a moral-hazard problem. Both of these issues potentially affect the quality of the data used to address RQ2 (the accept/decline responses for RQ1 are unaffected). However, the clarity of our results and the fact that reviewer characteristics did not differ significantly between those who gave their consent and those who did not suggests that our findings are unlikely to be driven by reviewers being informed about their participation in a study.

Second, both authors of the paper being sent out for review, Inoua and Smith, were affiliated with the same institution (Chapman University) and had the same gender. While we held these two potential confounding factors constant, their names have different origins and may have led to—conscious or unconscious—discrimination (8). Furthermore, Inoua is dark-skinned, while Smith is light-skinned. Skin tone (i.e., colorism) has long been identified as among the most salient markers for discrimination (10, 11). An influence of discrimination based on name or skin color is thus also possible in our study. However, while available in principle, a photo of Inoua was not straightforward to find online at the time of our study.

Third, we used only one paper in one field (finance/financial economics) to study our RQs, such that we cannot unambiguously speak for other fields of science. While we personally are confident that similar effects likely are also present in other fields, we encourage follow-up studies to replicate our findings in other disciplines.

Discussion and Conclusion

We collected 534 review reports for the same manuscript, varying only whether the corresponding author shown was relatively unknown, highly prominent, or not mentioned. Our results document strong evidence for a status bias in peer review: When the prominent researcher was shown as the corresponding author, significantly more peers accepted the invitation to review the paper than in the other conditions. More importantly, the paper with the prominent author also received significantly fewer "reject" recommendations, and its content was assessed much more favorably than the manuscript showing the less prominent author or no author name.

Given that we sent out otherwise identical manuscripts and that both authors were of the same gender and affiliated with the same institution, the most plausible explanation for the very different assessments by reviewers is that they consciously or unconsciously ascribed higher quality to a paper authored by a prominent researcher. We consider it rather unlikely that our results are due to "the old boys network" (3), as arguably few of the 534 reviewers were likely to have been personal acquaintances of Nobel laureate Smith. Furthermore, reviewers had no way of

"ingratiating" themselves with Smith through a favorable report, as all reports were anonymized, and no personal information of the reviewers was shared with the authors of the paper, in line with the typical practices of both single- and double-anonymized review. The evidence, rather, suggests that many reviewers, aware of the previous work of Smith, automatically attributed high quality to this paper. This is reminiscent of the "halo effect" in social psychology (12, 13), where raters extrapolate from known to unknown information: An initial, favorable impression of a subject typically leads to a higher rating, while an unfavorable prior leads to a lower rating. Halo effects have been documented for various kinds of evaluations, ranging from performance ratings of employees (14), to licensing of university inventions (15), to student (16, 17) and instructor (18) assessments.

On the crucial question of whether peer review should be single-anonymized or double-anonymized, our results speak in support of the latter. However, as more and more working papers and preprints are made available on the internet, a truly double-anonymized review process becomes less and less realistic. Still, even if many papers today can easily be found on the internet, anonymized papers at least leave the reviewer the option to remain ignorant. Double-anonymizing the peer-review process could also help level the playing field for academics from other marginalized groups, giving them a fairer chance to succeed, which, in turn, would promote diverse points of view in journal output and eventually expand the journal peer-reviewer pool to authors coming from such backgrounds (3–5, 19–23).

That said, we only compared single-anonymized vs. double-anonymized review processes, but no other alternatives, like, e.g., a fully transparent review process, possibly even allowing open discussions between authors, reviewers, and editors. Another option are forms of "structured peer review" that prompt reviewers to help improve a manuscript, but do not ask for an accept/revise/reject recommendation. Exploring the impact of such innovative review processes is a promising avenue for future research. Other avenues to gain additional insights on the issues raised here would be to redo the same experiment again with an outstanding or a particularly bad paper to see whether the treatment differences persist. Also, rerunning the experiment at a very-high-prestige journal, rather than at the rather young journal we used, could be an interesting avenue for future research.

Materials and Methods

Preregistration and Ethics Approval. This study was preregistered on July 27, 2021. The preregistration is available on the Open Science Framework (OSF) (https://osf.io/mjh8s/). The study was approved by Elsevier (ref. IJsbrand Jan Aalbersberg, Senior Vice President of Research Integrity), the publisher of *JBEF*, and by the institutional review boards of the University of Graz (39/126/63 ex 2020/21) and the University of Innsbruck (13/2021).

Reviewer Selection. To compile the pool of potential reviewers, we started from the top 100 list of the Journal Citation Report 2019, category "Business, Finance" (24). From this list, we eliminated journals whose focus did not fit the topic of the research paper by S.I. and V.L.S., and we added to the list *JBEF*, as it is the journal with which we were collaborating for this study. The result of this exercise was the journal list given in *SI Appendix*, Table S3. The pool of potential reviewers consisted of researchers who had published in at least 1 of these 29 journals between 2018 and 2020 and for whom contact details and Google Scholar profiles were available. The result of this exercise was what we called the "adjusted list of potential reviewers," which contained information on more than 5,500 researchers, who were affiliated with more than 1,500 research institutions at the time. More information on selection and disambiguation is available in *SI Appendix*.

Randomization and Invitations in Waves. For the assignment of conditions to potential reviewers, we took the adjusted list of potential reviewers as our starting point. We first counted the number of potential reviewers by institution. We then distributed the institutions into bins guided by the number of reviewers, starting with the institutions with just one potential reviewer. We set bin cut-offs such that the number of institutions decreased monotonically and the number of institutions per bin was evenly divisible by five (corresponding to our five experimental conditions). Within each bin, experimental conditions were then randomly, but uniformly, assigned to institutions. We followed this stratified sampling method based on bins to avoid biases in the distribution of institutional affiliations of reviewers across experimental conditions. We randomly assigned conditions to institutions (and not to individual researchers) to minimize the risk of condition spillovers (i.e., reviewers at the same institution discussing the study and learning about our condition variations). Reviewers were invited in a predefined order that was determined by recency of their latest publication; i.e., we first invited potential reviewers who had more recently published in one of the relevant journals. Specifically, in the first wave, we invited one reviewer per institution, and, for institutions with more than one reviewer, it was the reviewer who had published in a relevant outlet most recently. For the second wave, we excluded bin 1 (because this bin contained only institutions with a single potential reviewer), and from the other bins, we invited, from each institution, the reviewer who had published second-most recently. Further waves followed the same pattern. We sent more than 3,300 invitations following this procedure, aiming to gather at least 100 reports for each of our five conditions. Randomization checks are provided in *SI Appendix*, Table S4. As we honored individuals' requests to delete data from our database, our analysis was based on (non)responses from 3,299 researchers (see *SI Appendix*, Table S12 for a full attrition analysis).

Data Handling and Reviewer Anonymity. Only one member of the team of researchers and one student assistant (aka "the administrators") had administrative access to the database and the review-management software we used to conduct the study. Before the start of the study, the administrators carried out the randomized assignment of institutions to the experimental conditions and subsequently triggered the waves of invitations. After the data collection had been completed, all data were first anonymized. Specifically, we employed the following procedure to ensure reviewer anonymity in the data-analysis stage: First, names, email addresses, and affiliation information were deleted from the dataset. Second, the exact information from the Google Scholar profile (number of citations, h-index, and i10-index) was recoded into categories (for instance, if a researcher had 2,867 citations, then this information was replaced by the entry "between 2,740 and 2,898 citations"; *SI Appendix*, Table S5, bin 62)—making sure that at least 20 individuals on the adjusted list of potential reviewers fell into each interval. All analyses that include reviewer characteristics are based on the categorical data. *SI Appendix*, Tables S5 and S6 provide details. Third, review reports and comments to the editors were anonymized. Any mention of reviewers' own names or references to their own works that could have revealed their identity were removed. All these steps were carried out by the administrators. Only fully anonymized datasets were then used for the analysis and shared with other members of the author team.

Informed Consent. Due to the nature of the study and the requirement for informed consent (in accordance with the EU's GDPR, as well as with university ethics guidelines), review invitations followed a two-step process. Researchers were first contacted by email and invited to review the manuscript. The email template for this initial contact is included in *SI Appendix*. The editor truthfully listed as handling this manuscript was S.P., one of the two co-editors-in-chief of *JBEF* and one of the authors of the present paper. The content of the review-invitation email followed the standard *JBEF* email template, with one key difference: In conditions LL and HH, the name of one of the authors was displayed as the corresponding author. The email contained two options: The recipients could accept the invitation to review the manuscript, or they could decline. Recipients who clicked the "decline" link in the invitation to review were brought to a website that asked for reasons for choosing to decline. A screenshot of this website is included in *SI Appendix*. Recipients who clicked the "accept" link in the invitation to review were brought to a website that informed them that the review both served to inform the editor about the manuscript's publishability and formed part of a research study into the peer-review process. Reviewers had to actively consent

to these conditions to continue. *SI Appendix*, Fig. S1 presents a screenshot of the corresponding website. If recipients declined or did nothing, the process ended, and no further communication took place with them. Recipients who accepted received immediate access to the manuscript on the same website. They also received an email informing them that they now had access to the manuscript via our platform.

Data Collection. We recorded the answers of potential reviewers to our invitation-to-review email. Specifically, we recorded 1) the date and time of the response; 2) the email address of the responder; 3) the condition the responder was in; and 4) the response (accept/decline). We asked those responders who declined the invitation to review to answer a short questionnaire, and we recorded the answers given in this questionnaire. Those responders who accepted the invitation landed on the informed-consent page, and we recorded the answers given on that page (accept/decline). On this page, we also collected participants' explicit consent to storing and processing their personal data for the purpose of the study, as required by the European GDPR. Most of the responders who accepted the invitation to review and gave consent on the informed-consent page later gave a recommendation regarding publication and provided a report. Here, we stored the date and time of the response, the recommendation given, and the report provided. In addition to asking for a recommendation regarding publication, we also elicited reviewers' opinions on six statements about the paper. Here, we stored the responses (each question had to be answered on a scale from 1, or "strongly disagree," to 5, or "strongly agree"). Those reviewers who submitted a report were finally asked to fill in a postreview questionnaire, and we recorded the answers to these questions. *SI Appendix* contains an overview of the times participants took for each stage of the experiment (*SI Appendix*, Table S7), information about attrition throughout the study (*SI Appendix*, Table S12), and the questionnaire that participants filled in when submitting their report (*SI Appendix*, Fig. S3), as well as the postreview questionnaire.

Report Handling. Before reports were shared with the full team of researchers in this project and used in the data analysis, they were checked by the administrators for elements that could reveal the reviewers' identities. All such elements were removed. Furthermore, information identifying the condition under which the report was submitted was removed where possible. Once all reports had been received, comments from all reports were aggregated into a single decision letter by S.P., one of the two co-editors-in-chief of *JBEF*. This decision letter, together with all individual reports, were shared with the authors.

Reviewer Debriefing. A debriefing email was sent to all reviewers who submitted a report. The email explained the purpose of the study and was sent to all recipients at the same time to prevent an early revelation of details of the study to parts of the sample. The template is included in *SI Appendix*.

RQs, Key Variables of Interest, and Hypotheses. We addressed two RQs: RQ1: What effect does author prominence have on the probability of a reviewer accepting the invitation to review the paper? RQ2: What effect does author prominence have on the assessment of the paper in the review report?

The key dependent variable for RQ1 (willingness of potential reviewers to assess the paper) was the frequency with which potential reviewers accepted the invitation to review the paper. Our preregistered ex ante hypothesis regarding this frequency was that the review-invitation acceptance probability would be higher in the condition where the prominent author was mentioned in the invitation letter as corresponding author than in the condition where the less prominent author was mentioned in the invitation letter as corresponding author. We also preregistered two subhypotheses:

- H1+ (positive component of the status bias in willingness to review): The reviewer-invitation acceptance rate is higher in the condition where the more prominent author is mentioned in the invitation letter as corresponding author than in the condition where no name is mentioned in the invitation letter.
- H1− (negative component of the status bias in willingness to review): The reviewer-invitation acceptance probability is lower in the condition where the less prominent author is mentioned in the invitation letter as corresponding author than in the condition where no name is mentioned in the invitation letter.

The key dependent variable for RQ2 (rating and assessment of the paper by the reviewers) was the recommendation given by the reviewer regarding publication of the paper. Our preregistered ex ante hypothesis regarding this decision was that the assessment of the paper would be more favorable in the condition where the more prominent author appeared as the corresponding author of the paper than in the condition where the less prominent author appeared as the corresponding author. Here, again, we had two subhypotheses:

- H2+ (positive component of the status bias in assessment of the paper): The assessment of the paper is more favorable in the condition where the more prominent author appears as the corresponding author of the paper than in the condition where no name is given.
- H2− (negative component of the status bias in assessment of the paper): The assessment of the paper is less favorable in the condition where the less prominent author appears as the corresponding author of the paper than in the condition where no name is given.

In addition to asking for a recommendation regarding publication, many Elsevier journals elicit reviewers' opinions on six statements about a paper [each of them has to be answered on a Likert scale, ranging from 1) strongly disagree to 5) strongly agree]. The questionnaire is included in *SI Appendix*, Fig. S4. Here, again, we had the hypothesis that the assessment would be more favorable in the condition where the more prominent author appeared as the corresponding author than in the condition where no name was given, and that the assessment of the paper would be less favorable in the condition where the less prominent author appeared as the corresponding author than in the condition where no name was given.

Identification Strategy. For RQ1 (willingness of reviewers to write a report), the distinction between AL, AA, and AH was irrelevant since potential reviewers did not see the paper or the author name before accepting or rejecting the invitation. We, therefore, pooled these three conditions to an "Anon" category and compared this category to LL and HH. We tested H1 by comparing LL to HH, H1+ by comparing Anon to HH, and H1− by comparing LL to Anon. For RQ2 (rating and assessment in the report), the main comparisons (reported in the body of the paper) involved conditions AL, AA, and AH, as these conditions allowed for a clean identification of the effect of author prominence on the recommendation stage without possible confounds caused by selection at the invitation stage. We tested H2 by comparing AL to AH, H2+ by comparing AA to AH, and H2− by comparing AL to AA. Concentrating on conditions AL, AA, and AH meant forgoing the potentially useful data from conditions LL and HH. We therefore followed our preregistration and checked whether we could pool AL and LL or AH and HH by testing for pairwise differences. Since we found statistically significant differences (AL vs. LL: $z = -2.172, P = 0.030$; AH vs. HH: $z = 0.774, P = 0.439$; two-sided Mann–Whitney U tests), we refrained from pooling, and we discuss the effects of selection in *SI Appendix*. A second set of variables of interest for RQ2 were the reviewer evaluations of the six statements about the paper. We describe a series of robustness checks in *SI Appendix*.

Deviations from the Preregistration Document. Our analysis largely followed the preregistration. In a conservative deviation from the plan, we reported two-sided tests instead of one-sided ones throughout the manuscript and *SI Appendix*. Apart from this deviation, we conducted few analyses that went beyond the preregistration: 1) We conducted additional tests of proportions on the rejection rates between conditions. 2) We aggregated the two most positive categories, "accept" and "minor revision," of the recommendation spectrum to illustrate the shift in the distributions of responses between conditions. 3) We added a robustness check to investigate the impact of removing outliers in terms of review time. 4) We conducted randomization checks to make sure reviewers' characteristics did not differ significantly between conditions. 5) We tested whether the less prominent author benefited from fully anonymized compared to single-anonymized evaluation. We plan to analyze the contents of the review reports both qualitatively and quantitatively in a separate publication. In this second manuscript, we also plan to analyze the postreview questionnaire in more detail.

Statistical Tests. All statistical tests reported stem from two-sided tests. We used four different types of tests throughout the manuscript. Fisher's exact tests were used whenever the measurement variables were binary–for example, when we tested for differences in the participants' decision to accept or decline the review invitation in the respective conditions. We used tests of proportions when we compared shares of participants across conditions–for example, when we aggregated the recommendation categories "minor revision" and "accept." We used Mann–Whitney U tests for pairwise tests on ordinal data, such as the hypothesis tests we ran on reviewers' recommendations. We used Kruskal–Wallis tests to test whether more than two samples stemmed from the same distribution.

For our preregistered hypothesis tests, we did not apply multiple hypothesis-testing corrections and took 0.05 as the α-threshold. For all other tests, we report the α-threshold after conducting Bonferroni–Holm corrections. The multiple hypothesis-testing corrections did not qualitatively affect the main results of the article.

Data, Materials, and Software Availability. All study data and the code producing the analyses we report are available in the publicly available OSF repository at https://osf.io/mjh8s/ (9). Names and affiliations of the 534 reviewers of the paper have to remain anonymous.

1. D. A. Kronick, Peer review in 18th-century scientific journalism. *JAMA* **263**, 1321–1322 (1990).
2. R. K. Merton, The Matthew effect in science. The reward and communication systems of science are considered. *Science* **159**, 56–63 (1968).
3. D. Cox, L. Gleser, M. Perlman, N. Reid, K. Roeder, Report of the ad hoc committee on double-blind refereeing. *Stat. Sci.* **8**, 310–317 (1993).
4. R. Blank, The effects of double-blind versus single-blind reviewing: Experimental evidence from the *American Economic Review*. *Am. Econ. Rev.* **81**, 1041–1067 (1991).
5. D. Peters, S. Ceci, Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behav. Brain Sci.* **5**, 187–195 (1982).
6. J. M. Garfunkel, M. H. Ulshen, H. J. Hamrick, E. E. Lawson, Effect of institutional prestige on reviewers' recommendations and editorial decisions. *JAMA* **272**, 137–138 (1994).
7. S. J. Ceci, W. M. Williams, Understanding current causes of women's underrepresentation in science. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3157–3162 (2011).
8. A. Tomkins, M. Zhang, W. D. Heavlin, Reviewer bias in single- versus double-blind peer review. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12708–12713 (2017).
9. J. Huber *et al.*, Supplement to Nobel and Novice: Author prominence affects peer review. Open Science Framework. https://doi.org/10.17605/OSF.IO/MJH8S. Deposited 23 September 2022.
10. E. Pierson *et al.*, A large-scale analysis of racial disparities in police stops across the United States. *Nat. Hum. Behav.* **4**, 736–745 (2020).
11. P. A. Riach, J. Rich, Field experiments of discrimination in the market place. *Econ. J. (Lond.)* **112**, F480–F518 (2002).
12. E. L. Thorndike, A constant error in psychological ratings. *J. Appl. Psychol.* **4**, 25–29 (1920).
13. W. H. Cooper, Ubiquitous halo. *Psychol. Bull.* **90**, 218–244 (1981).
14. R. Holzbach, Rater bias in performance ratings: Superior, self-, and peer ratings. *J. Appl. Psychol.* **63**, 579–588 (1978).
15. W. D. Sine, S. Shane, D. DiGregorio, The halo effect and technology licensing: The influence of institutional prestige on the licensing of university inventions. *Manage. Sci.* **49**, 478–496 (2003).
16. I. Dennis, Halo effects in grading student projects. *J. Appl. Psychol.* **92**, 1169–1176 (2007).
17. J. Malouff, A. Emmerton, N. Schutte, The risk of a halo bias as a reason to keep students anonymous during grading. *Teach. Psychol.* **40**, 233–237 (2013).
18. R. E. Nisbett, T. D. Wilson, The halo effect: Evidence for unconscious alteration of judgments. *J. Pers. Soc. Psychol.* **35**, 250–256 (1977).
19. M. A. Ucci, F. D'Antonio, V. Berghella, Double- vs single-blind peer review effect on acceptance rates: A systematic review and meta-analysis of randomized trials. *Am. J. Obstet. Gynecol. MFM* **4**, 100645 (2022).
20. M. Fisher, S. B. Friedman, B. Strauss, The effects of blinding on acceptance of research papers by peer review. *JAMA* **272**, 143–146 (1994).
21. M. Alam *et al.*, Blinded vs. unblinded peer review of manuscripts submitted to a dermatology journal: A randomized multi-rater study. *Br. J. Dermatol.* **165**, 563–567 (2011).
22. K. Okike, K. T. Hug, M. S. Kocher, S. S. Leopold, Single-blind vs double-blind peer review in the setting of author prestige. *JAMA* **316**, 1315–1316 (2016).
23. D. Card, S. DellaVigna, What do editors maximize? Evidence from four economics journals. *Rev. Econ. Stat.* **102**, 195–217 (2020).
24. Clarivate Analytics, Journal citation reports (2019). https://clarivate.com/webofsciencegroup/web-of-science-journal-citation-reports-2022-infographic/.