8-16-2022

Nobel and Novice: Author Prominence Affects Peer Review

Jürgen Huber
University of Innsbruck

Sabiou M. Inoua
Chapman University, inoua@chapman.edu

Rudolf Kerschbamer
University of Innsbruck

Christian König-Kersting
University of Innsbruck

Stefan Palan
University of Graz

See next page for additional authors

Nobel and Novice: Author Prominence Affects Peer Review

Comments

ESI Working Paper 22-15

This paper later underwent peer review and was published as:

Huber, J., Inoua, S., Kerschbamer, R., König-Kersting, C., Palan, S., & Smith, V. L. (2022). Nobel and novice: Author prominence affects peer review. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(41), e2205779119. https://doi.org/10.1073/pnas.2205779119

Authors

Jürgen Huber, Sabiou M. Inoua, Rudolf Kerschbamer, Christian König-Kersting, Stefan Palan, and Vernon L. Smith

Nobel and novice:
Author prominence affects peer review.

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

[a]*Dep. of Banking and Finance, University of Innsbruck, Universitätsstraße 15, 6020 Innsbruck, Austria*
[b]*Economic Science Institute, Chapman University, 1 University Drive, Orange, CA 92866, USA*
[c]*Department of Economics, University of Innsbruck, Universitätsstraße 15, 6020 Innsbruck, Austria*
[d]*Institute of Banking and Finance, University of Graz, Universitätsstraße 15, 8010 Graz, Austria*

Abstract

Peer-review is a well-established cornerstone of the scientific process, yet it is not immune to status bias. Merton identified the problem as one in which prominent researchers get disproportionately great credit for their contribution while relatively unknown researchers get disproportionately little credit.[1] We measure the extent of this effect in the peer-review process through a pre-registered field experiment. We invite more than 3,300 researchers to review a paper jointly written by a prominent author – a Nobel laureate – and by a relatively unknown author – an early-career research associate –, varying whether reviewers see the prominent author's name, an anonymized version of the paper, or the less well-known author's name. We find strong evidence for the status bias: while only 23 percent recommend "reject" when the prominent researcher is the only author shown, 48 percent do so when the paper is anonymized, and 65 percent do so when the little-known author is the only author shown. Our findings complement and extend earlier results on double-anonymized vs. single-anonymized review[2,3,4,5,6,7] and strongly suggest that double-anonymization is a minimum requirement for an unbiased review process.

Peer review has been the key method for research validation since the first scientific journals appeared some 300 years ago.[8] For researchers – and especially for young scientists, who must excel for scientific advancement – it is crucial that this process be fair and impartial. Merton, however, argues 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"*.[1] Alluding to the Gospel according to Matthew 25,29, Merton terms this pattern of the misallocation of credit for scientific work "the Matthew effect in science", while others call it "status bias".[4]

Email addresses: juergen.huber@uibk.ac.at (Jürgen Huber), inoua@chapman.edu (Sabiou Inoua), rudolf.kerschbamer@uibk.ac.at (Rudolf Kerschbamer), christian.koenig@uibk.ac.at (Christian König-Kersting), stefan.palan@uni-graz.at (Stefan Palan), vsmith@chapman.edu (Vernon L. Smith)

Preprint submitted to an undisclosed journal *August 16, 2022*

We measure 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 pre-registered field experiment. Specifically, we address two related research questions: First, is there a status bias in potential reviewers' propensity to accept the invitation to review a paper? Second, is there a status bias in their evaluation of the manuscript? To address both research questions, we use a finance research article which was jointly written by a prominent scientist and a relatively unknown scientist. The prominent scientist is Vernon 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 Sabiou Inoua, an early-career research associate (42 Google Scholar citations). Both are affiliated with the Economic Science Institute at Chapman University, USA. Holding the affiliation constant is important because previous research documents an impact of institutional affiliation on the outcome of the review process under single-anonymized evaluation.[2,3,9]

The manuscript Smith and Inoua co-authored was submitted to the *Journal of Behavioral and Experimental Finance* (*JBEF* henceforth; published by Elsevier). The journal's editor 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, or 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* or the *Journal of Public Economic Theory*).

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 appears both in the invitation mail and on the manuscript; in condition AA (A for anonymized) neither the invitation email nor the manuscript shows an author name; in condition HH (H for high prominence) the prominent author appears both in the invitation email and on the manuscript; and in conditions AL and AH the invitation email is anonymized but the respective corresponding author's name appears on the manuscript proper. Importantly, neither the paper nor a draft thereof was posted online or presented anywhere before our data collection was complete, hence reviewers could not 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 compare reviewers' decisions to accept or decline the invitation between the conditions with anonymized emails (AL, AA and AH pooled) and those with non-anonymized emails (LL and HH). To test the impact of author prominence on the *assessment of the paper*, we compare reviewers' publication recommendations for manuscripts that show the corresponding author's name (conditions AL and AH) to those for manuscripts that do not show a name (condition AA). Since the invitations for these three conditions were anonymized, we observe the effect of author prominence on the recommendation decision without possible confounds caused by selection effects in the invitation stage.[6]

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 compare review invitation acceptance rates in the anonymized setting of conditions AL, AA and AH with those in LL and HH (see Table 1). Acceptance rates are based on the 2,611 researchers that responded to our invitation email by accepting or declining (see Table 2 in the Supplementary Information for a full attrition analysis).

We find that the share of researchers that accepted the invitation varies with the setting, with 28.5% of those who responded in condition LL accepting the invitation listing Sabiou 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 listing Nobel laureate Vernon 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; LL) are highly significant, the difference between the anonymized conditions (pooled) and LL is not. Alternatively using acceptance rates based on the invitations sent (rather than on responses received) does not change our results qualitatively: Acceptance rates drop 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 are highly significant, while the difference between the anonymized settings and LL is not.

Hence, we find statistically significant evidence for the positive component of the status bias in 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 do 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).

Table 1: Invitations

	Low (LL)	Anonymized (AL, AA, AH)	High (HH)	Total
Invitations sent	781	2011	507	3299
Responses received	610	1591	410	2611
Invitations accepted	174	489	158	821
Acceptance rate	28.52%	30.74%	38.54%	31.44%
Anon. vs. Low		$p = 0.3243$		
Anon. vs. High		$p = 0.0031$		
Low vs. High		$p = 0.0011$		

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

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 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. vs. Low: $p = 0.2655$; Anon. vs. High: $p = 0.5496$; Low vs. High: $p = 0.1634$; 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, HH: 107), enabling us to reach our pre-registered target of 100 reports per condition. For the analysis of manuscript assessments, we focus 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 Extended Data Table 2 and discuss the effects of selection on manuscript assessment below and in the Supplementary Information.

Arguably the most important outcome of the review process are the reviewers' final recommendations to the editor. Reviewers were asked to recommend either "reject", "major revision", "minor revision", or "accept" when submitting their reports. We follow our pre-analysis plan and compare the distributions of recommendations given by the reviewers using Mann-Whitney U tests. We find highly significant differences between conditions and

report them in Figure 1.

When the 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 have a 93 percent probability of getting published in the journal (2018–2020). We find stark differences in the rejection recommendations across the three conditions: While 65.4% of the reviewers recommend "reject" when shown the less prominent author, this number drops to 48.2% in the anonymized version of the manuscript and to 22.6% when the prominent author is shown. The pairwise differences between conditions are all significant: the p-value for AL vs. AA is 0.0120 (two-sided test of proportions, $z = 2.5117$), that for AA vs. AH is 0.0001 ($z = 3.885$), and that for AL vs. AH is $<$0.0001 ($z = 6.1444$; α-threshold corrected for three tests is 0.05). Thus, compared to fully anonymous review, the less prominent author clearly faces a lower and the more prominent author a higher chance of passing the first round of peer review.

Aggregating the two most positive categories "accept" and "minor revision" of the recommendation spectrum to a single category shows that only 9.9% of the reviewers recommend a minor revision or an outright accept when shown the less prominent author, while 23.6% give one of these recommendations in the anonymized version, and 58.8% do so when shown the prominent author. The pairwise differences between conditions are all highly significant: the p-value for AL vs. AA is 0.0081 (two-sided test of proportions, $z = -2.6495$), that for AA vs. AH is $<$0.0001 ($z = -5.2133$), and that for AL vs. AH is $<$0.0001 ($z = -7.3323$; α-threshold corrected for three tests is 0.05).

In summary, we find strong evidence for the positive component of the status bias in that the prominent author receives far more favorable assessments (and in particular, fewer rejections) than does the anonymized version of the manuscript; and we also find strong evidence for the negative component of the effect in that the less prominent author gets significantly worse assessments (and in particular more rejections) than 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 the Supplementary Information, we show that those reviewers that accept the invitation to review the paper despite being informed that the corresponding author is the less prominent researcher (in condition LL) are milder in their judgements than reviewers that respond to the anonymized invitation. This self-selection effect benefits the less prominent author and (partly) counteracts the negative component of the status bias of the less prominent author's manuscript being evaluated less favorably. Although the net effect still points in the direction of less prominent authors benefiting from fully anonymized review, the difference to single-anonymized review is no longer statistically significant for the less prominent author (while the reversed difference for the prominent author remains highly significant).

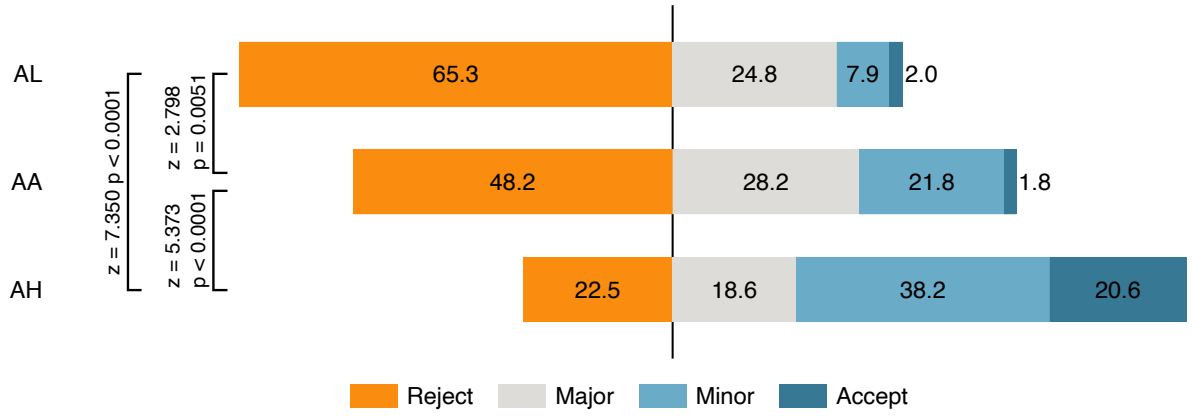


Figure 1: Recommendation percentages by condition. L stands for the relatively unknown author, A stands for anonymized and H stands for the highly prominent author. In conditions AL and AH, the invitation email is anonymized, but the respective corresponding author's name appears on the manuscript, while in AA both the invitation and the paper are anonymized. The tests are pairwise, two-sided Mann-Whitney U tests.

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.

Figure 2 shows the distribution of ratings and individual test statistics for the three items addressing the quality of the manuscript. Extended Data Figure 1 shows the corresponding data for items (4) to (6). For all six items, we observe a significant upwards shift of the distribution of ratings when the manuscript shows 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 affects not only final recommendations, but also the individual components of the manuscript quality assessments. Notably, while all comparisons have the expected sign, we find 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 supplementary material being appropriate.

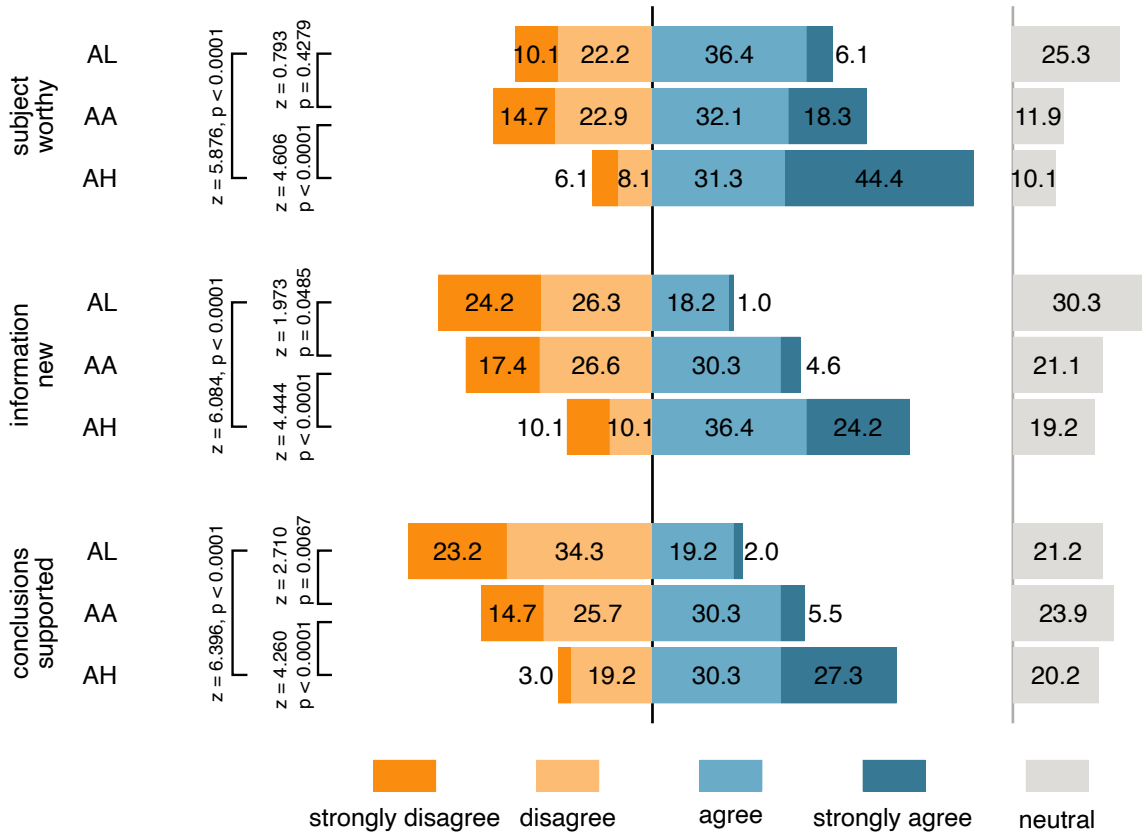


Figure 2: Responses to reviewer questionnaire items 1–3. We plot the percentage of neutral responses on the right-hand border of the figure. For each item, we conduct pairwise, two-sided Mann-Whitney U tests across conditions.

Discussion

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 or no author name.

Given that we sent out otherwise identical manuscripts and that both authors were 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",[4] as arguably few of the 534 reviewers are likely to be personal acquaintances of Nobel laureate Vernon Smith. Furthermore, reviewers had no way of 'ingratiating' themselves with Vernon 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 Vernon Smiths' previous work, automatically attributed high quality to this paper. This is reminiscent of the 'halo effect' in social psychology[10,11] where raters extrapolate from known to unknown information: an initial, favorable impression of a subject typically leads to a higher rating while an unfavourable prior leads to a lower rating. Halo effects have been documented for various kinds of evaluations, ranging from performance ratings of employees[12], to licensing of university inventions[13], to student[14,15] and instructor[16] assessments.

Several recent papers[17,18,19] argue that increasing transparency in the review process by, e.g., publishing the (anonymized) reports, is an important step in improving the review process and boosting trust in science. Yet our results suggest the best solution would be "perfect" double-anonymization of the whole process. Thus, on the crucial question of whether peer review should be single-anonymized or double-anonymized, our results clearly speak in support of the latter. However, especially since more and more working papers and preprints are made available on the internet, the trend points in the other direction. Still, even if many papers can today easily be found on the internet, anonymized papers at least leave the reviewer the option to remain ignorant. Double-anonymizing the peer review process would likely also level the playing field for academics from marginalized groups, giving them a fairer chance to succeed, which in turn promotes diverse points of view in journal output and eventually expands the journal peer reviewer pool to authors coming from these backgrounds.

References

[1] R. K. Merton, The Matthew effect in science, Science 159 (1968) 56–63.

[2] D. Peters, S. Ceci, Peer-review practices of psychological journals: The fate of published articles, submitted again, Behavioral and Brain Sciences 5 (1982) 187–195.
[3] R. Blank, The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review, American Economic Review 81 (1991) 1041–1067.
[4] D. Cox, L. Gleser, M. Perlman, N. Reid, K. Roeder, Report of the ad hoc committee on double-blind refereeing, Statistical Science 8 (1993) 310–317.
[5] K. Okike, K. Hug, K. Kocher, S. Leopold, Single-blind vs double-blind peer review in the setting of author prestige, Journal of the American Medical Association 316 (2016) 1315–1316.
[6] A. Tomkins, M. Zhang, W. Heavlin, Reviewer bias in single- versus double-blind peer review, PNAS 114 (2017) 12708–12713.
[7] D. Card, S. DellaVigna, What do editors maximize? Evidence from four economics journals, Review of Economics and Statistics 102 (2020) 195–217.
[8] D. A. Kronick, Peer review in 18th-century scientific journalism, Jama 263 (1990) 1321–1322.
[9] J. Garfunkel, M. Ulshen, H. Hamrick, E. Lawson, Effect of institutional prestige on reviewers' recommendations and editorial decisions, Journal of the American Medical Association 272 (1994) 137–138.
[10] E. L. Thorndike, A constant error in psychological ratings, Journal of Applied Psychology 4 (1920) 25–29.
[11] W. H. Cooper, Ubiquitous halo, Psychological Bulletin 90 (1981) 218–244.
[12] R. Holzbach, Rater bias in performance ratings: Superior, self-, and peer ratings, Journal of Applied Psychology 63 (1978) 579–588.
[13] W. D. Sine, S. Shane, D. DiGregorio, The halo effect and technology licensing: The influence of institutional prestige on the licensing of university inventions, Management Science 49 (2003) 478–496.
[14] I. Dennis, Halo effects in grading student projects, Journal of Applied Psychology 92 (2007) 1169–1176.
[15] J. Malouff, A. Emmerton, N. Schutte, The risk of a halo bias as a reason to keep students anonymous during grading, Teaching of Psychology 40 (2013) 233–237.
[16] R. E. Nisbett, T. D. Wilson, The halo effect: Evidence for unconscious alteration of judgments, Journal of Personality and Social Psychology 35 (1977) 250–256.
[17] J. Polka, R. Kiley, B. Konforti, R. Vale, Publish peer-reviews, Nature 560 (2018) 545–547.
[18] G. Bravo, F. Grimaldo, E. López-Iñesta, B. Mehmani, F. Squazzoni, The effect of publishing peer review reports on referee behavior in five scholarly journals, Nature Communications 10 (2019) 322.
[19] F. Squazzoni, P. Ahrweiler, T. Barros, F. Bianchi, A. Birukou, H. Blom, G. Bravo, S. Cowley, V. Dignum, P. Dondio, F. Grimaldo, L. Haire, J. Hoyt, P. Hurst, R. Lammey, C. MacCallum, A. Marušić, B. Mehmani, H. Murray, D. Nicholas, G. Pedrazzi, I. Puebla, P. Rodgers, T. Ross-Hellauer, M. Seeber, K. Shankar, J. V. Rossum, M. Willis, Unlock ways to share data on peer-review, Nature 578 (2020) 512–514.
[20] Clarivate Analytics, Journal citation reports (2019). URL: `https://clarivate.com/webofscience group/web-of-science-journal-citation-reports-2021-infographic/`.

Methods

Pre-registration and ethics approval. This study was pre-registered on 27 July 2021. The pre-registration is available on OSF at `https://osf.io/mjh8s/?view_only=e898002f08864bfd94b0328508f1d54b`. The study was approved by Elsevier (ref. IJsbrand Jan Aalbersberg, Senior Vice President of Research Integrity), the publisher of the *Journal of Behavioral and Experimental Finance*, and by the IRBs 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".[20] From this list we eliminated journals whose focus does not fit the topic of the research paper by Sabiou Inoua and Vernon Smith and we added to the list the *Journal of Behavioral and Experimental Finance* (henceforth, *JBEF*) as it is the journal with which we are collaborating for this study. The result of this exercise was the journal list given in Extended Data Table 1. The pool of potential reviewers consists of researchers who had published in at least one of these 29 journals between 2018 and 2020 and for whom contact details and Google Scholar profiles were available. The result of this exercise is what we call the *adjusted list of potential reviewers*, which contains information of more than 5,500 researchers who were affiliated with more than 1,500 research institutions at the time.

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 decreases monotonically and the number of institutions per bin is evenly divisible by 5 (since we have 5 experimental conditions, as explained in the body of the paper). 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 contains 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 Extended Data Table 3.

As we honored individuals' requests to delete data from our database, our analysis is based on (non-)responses from 3,299 researchers (see Table 2 in the Supplementary Information 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 was 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, i10-index) was replaced by "class information" (for instance, if a researcher had 2,867 citations then this information was replaced by the entry "between 2,000 and 5,000 citations") – making sure that at least 20 individuals on the adjusted list of potential reviewers fall into the same class. Third, review reports and comments to the editors were anonymized. Any mention of reviewers' own names or references to their own works that could reveal 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 *General Data Protection Regulation* or 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 Section 1.1 of the Supplementary Information. The editor truthfully listed as handling this manuscript was Stefan Palan, one of the two co-editors-in-chief of *JBEF*. 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 which asked for reasons for choosing to decline. A screenshot of the respective website is included in the Supplementary Information. Recipients who clicked the accept link in the invitation to review were brought to a website which 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. Figure 1 in the Supplementary Information presents a screenshot of the corresponding website. If recipients declined or did nothing, the process ended and no further communication took place with them. Only recipients who accepted received access to the manuscript. 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 (i) the date and time of the response, (ii) the email address of the responder, (iii) the condition the responder was in, and (iv) 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 post-review questionnaire and we recorded the answers to these questions. The Supplementary Information contains an overview of the times participants took for each stage of the experiment, information about attrition throughout the study, the questionnaire participants filled in when submitting their report, as well as the post-review 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 the report was submitted under were removed where possible. Once all reports had been received, comments from all reports were aggregated into a single decision letter by Stefan Palan, 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 Section 1.4 of the Supplementary Information.

Research questions, key variables of interest and hypotheses. We address two research questions: 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) is the frequency with which potential reviewers accept the invitation to review the paper. Our pre-registered ex-ante hypothesis regarding this frequency is that the review invitation acceptance probability is higher in the condition where the prominent author is mentioned in the invitation letter as corresponding author than in the condition where the less prominent author is mentioned in the invitation letter as corresponding author. We also pre-registered two sub-hypotheses:

- 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) is the recommendation given by the reviewer regarding publication of the paper. Our pre-registered ex-ante hypothesis regarding this decision is that 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 the less prominent author appears as the corresponding author. Here again we have two sub-hypotheses:

- 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 Figure 3 in the Supplementary Information. Here again we have the hypothesis that the assessment 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, and that the assessment of the paper is less favorable in the condition where the less prominent author appears as the corresponding author than in the condition where no name is given.

Identification strategy. For RQ1 (willingness of reviewers to write a report) the distinction between AL, AA and AH is 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 test 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) involve conditions AL, AA and AH, as these conditions allow 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 test 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 means forgoing the potentially useful data from conditions LL and HH. We therefore followed our pre-registration and checked whether we can pool AL and LL or AH and HH by testing for pairwise differences. Since we found statistically significant differences (AL vs LL: $z = -2.172$, $p = 0.0467$; AH vs. HH: $z = 0.774$, $p = 0.5049$; two-sided Mann-Whitney U tests), we refrained from pooling and we discuss the effects of selection in Section 4 of the Supplementary Information. A second set of variables of interest for RQ2 are the reviewer evaluations of the six statements about the paper.

Statistical tests. All statistical tests reported are two-sided tests. We use three different types of tests throughout the manuscript. Fisher's exact tests are used whenever the measurement variables are binary, for example when we test for differences in the participants' decision to accept or decline the review invitation in the respective conditions. We use tests of proportions when we compare shares of participants across conditions, for example when we aggregate recommendation categories "minor revision" and "accept". Mann-Whitney U tests are used for tests on ordinal data, such as the hypothesis tests we run on reviewers' recommendations.

For our pre-registered hypothesis tests we do not apply multiple hypothesis testing corrections and take 0.05 as the $\alpha-$threshold. For all other tests, we report the α-threshold after conducting Bonferroni-Holm corrections. The multiple hypothesis testing corrections do not qualitatively affect the main results of the article.

Robustness checks. In our first set of robustness checks, we run additional regressions (reported in Extended Data Table 4) to investigate the impact of reviewer experience (operationalized by calculating ranks based on the number of Google Scholar citations) on the results regarding our two research questions. For RQ1 (willingness of reviewers to write a report) we find that experience decreases the propensity to accept the review invitation (the main effect of the number of citations is negative) and we also find that more experienced authors are more willing to accept the invitation to review the paper displaying the more prominent author's name (the interaction effect of the number of citations with 'High' is positive). While the former effect is probably due to the fact that more experienced authors have higher costs (in terms of the opportunity cost of time) and lower benefits (e.g., they do not need to convince editors that they are experts in the field) of reviewing, the latter effect is consistent with more experienced potential reviewers being more likely to recognize the name of the more prominent author. For RQ2 (rating and assessment in the report) experience has neither a main effect nor an interaction effect with our experimental conditions and its inclusion leaves all of our findings qualitatively unchanged. The results regarding reviewer experience remain qualitatively unchanged if we replace the number of Google Scholar citation by the h-index, the i10-index or academic age.

In our second set of robustness checks, we investigate the impact of removing outliers in terms of review time. Excluding the top and bottom 5 percent, respectively, of the reports in terms of how long the reviewers took to submit their reports, leaves all of our

findings qualitatively unchanged. We also searched for differences between our experimental conditions in (i) the time it took reviewers to reply to the invitation to review, (ii) the time between the reply and the consent and (iii) the time between the consent and the report submitted. We find no relevant or statistically significant differences. The timings are reported in Table 1 of the Supplementary Information.

End notes

Acknowledgements

The authors thank Elsevier for permitting them to conduct this research through the *Journal of Behavioral and Experimental Finance*. Special thanks are due to IJsbrand Jan Aalbersberg, Sumantha Alagarsamy, Jürgen Fleiß, Anita Gantner, Felix Holzmeister, Yana Litkovsky, Bahar Mehmani, Dana Niculescu, Andrea Schertler, Marco Schwarz, Swetha Soman, Thomas Stöckl, Erik Theissen and Daniel Woods for valuable comments and suggestions. Financial support from the Austrian Science Fund (FWF) through SFB F63 is gratefully acknowledged.

Author contributions

Jürgen Huber acquired the funding and – together with Rudolf Kerschbamer – came up with the main concept for this study. Jürgen Huber, Rudolf Kerschbamer, Christian König-Kersting and Stefan Palan developed the detailed study protocol in close collaboration with Elsevier. Sabiou Inoua and Vernon Smith wrote the manuscript submitted for review. Christian König-Kersting programmed and managed the software to conduct the study. He also curated the data and conducted the formal analysis. Stefan Palan was in charge of the editorial process and provided a detailed summary of all reports to Sabiou Inoua and Vernon Smith. Jürgen Huber and Rudolf Kerschbamer wrote the original draft, and all authors were involved in validation, reviewing and editing.

Data availability

The data set is available on OSF at `https://osf.io/mjh8s/?view_only=e898002f08864bfd94b0328508f1d54b`.

Code availability

The code producing the analyses we report is available on OSF at `https://osf.io/mjh8s/?view_only=e898002f08864bfd94b0328508f1d54b`.

Additional information

Supplementary Information is available for this paper.
Correspondence and requests for materials should be addressed to Jürgen Huber.

Extended data

Extended Data Table 1: List of journals reviewers were selected from

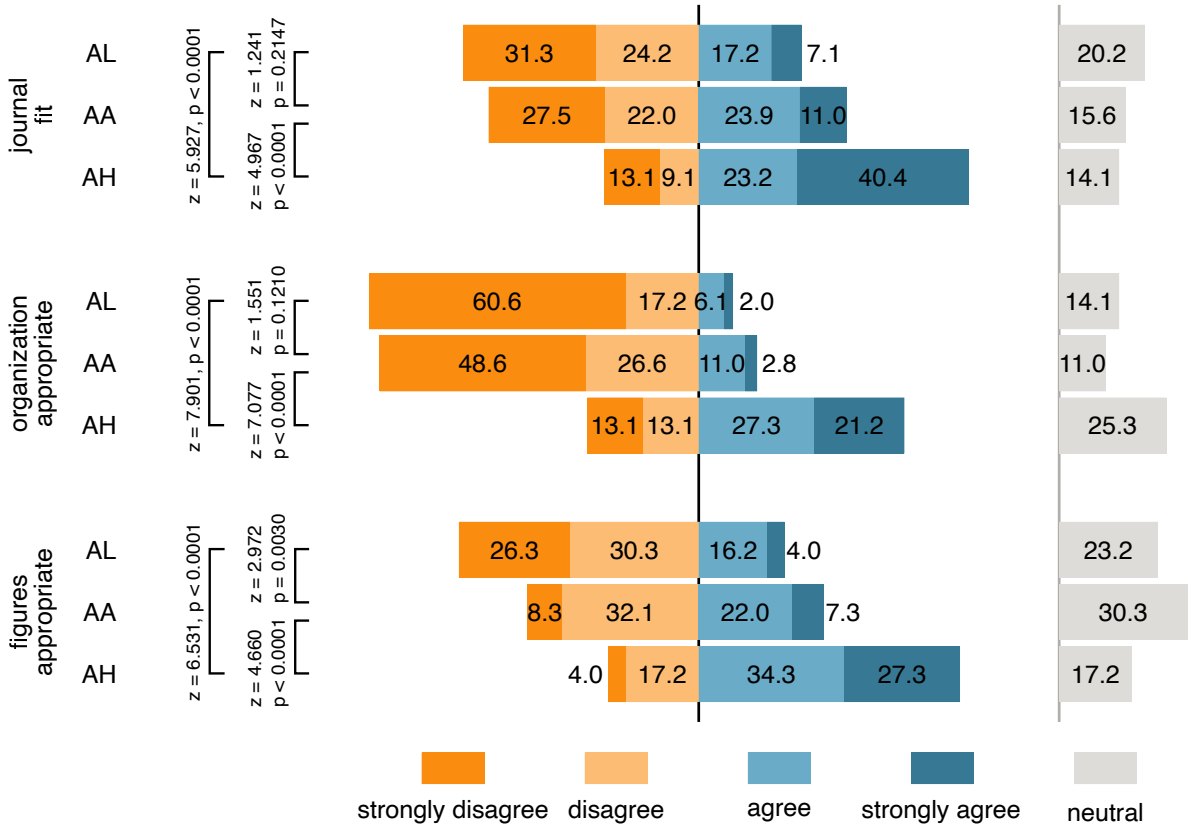
Name	Rank in category	Impact factor (2019)
Journal of Finance	2	6.813
Journal of Financial Economics	3	5.731
Review of Financial Studies	4	4.649
Finance Research Letters	9	3.527
Review of Finance	16	2.885
Journal of Financial and Quantitative Analysis	18	2.707
International Review of Financial Analysis	23	2.497
Pacific-Basin Finance Journal	25	2.382
Journal of Banking and Finance	30	2.269
Mathematical Finance	31	2.250
Borsa Istanbul Review	33	2.130
Finance and Stochastics	38	2.048
Journal of International Money and Finance	40	2.014
Abacus	41	1.975
International Review of Economics & Finance	48	1.818
Research in International Business and Finance	49	1.801
Journal of Risk and Uncertainty	51	1.756
Journal of Financial Markets	55	1.677
Journal of Empirical Finance	61	1.566
North American Journal of Economics and Finance	62	1.535
Quantitative Finance	64	1.491
Journal of Financial Research	74	1.263
European Journal of Finance	76	1.217
International Review of Finance	78	1.177
International Journal of Finance & Economics	87	0.943
Journal of Behavioral Finance	88	0.930
International Finance	91	0.848
Mathematics and Financial Economics	92	0.792
Journal of Behavioral and Experimental Finance[a]	n/a	n/a

[a] The Journal of Behavioral and Experimental Finance was only founded in 2014. It will receive its first impact factor in 2022.

Extended Data Table 2: Recommendations by condition.

condition	N	Accept %	Minor %	Major %	Reject %
AL	101	1.98	7.92	24.75	65.35
AA	110	1.82	21.82	28.18	48.18
AH	102	20.59	38.24	18.63	22.55
LL	114	3.51	14.91	29.82	51.75
HH	107	14.95	36.45	29.91	18.69
Total	534	8.43	23.78	26.40	41.39

L=author prominence low, A=anonymized, H=author prominence high, first letter is for the invitation letter, second letter is for the manuscript.



Extended Data Figure 1: Responses to reviewer questionnaire items 4–6. We plot the percentage of neutral responses on the right-hand side of the figure. For each item, we conduct pairwise, two-sided Mann-Whitney U tests across conditions.

Extended Data Table 3: Randomization checks: all invited participants.

	Citations	h-index	i10-index	Years active
AL	0.1173	0.3097	0.1021	0.114
	(0.0967)	(0.1342)	(0.1189)	(0.1158)
AH	0.06081	0.1845	0.0829	0.0231
	(0.0977)	(0.1376)	(0.1205)	(0.1186)
LL	-0.0004	-0.0167	-0.1229	0.08613
	(0.0945)	(0.138)	(0.1199)	(0.115)
HH	-0.0450	0.1362	0.0632	0.1154
	(0.1053)	(0.1489)	(0.1301)	(0.1268)
Observations	3284	3284	3284	3279
LR chi2(4)	3.395	9.111	5.284	1.623
Prob. $>$ chi2	0.494	0.0584	0.2593	0.8047

Ordered logistic regressions, odds ratios reported; standard errors in parentheses.

Extended Data Table 4: Reviewer characteristics.

	invitation acceptance			recommendations		
	H1	H1c	H1i	H2	H2c	H2i
Low	-0.106	-0.147	-0.146	-1.087	-1.058	-1.009
	(0.105)	(0.108)	(0.157)	(0.218)	(0.218)	(0.282)
High	0.346	0.346	0.112	1.106	1.121	1.155
	(0.115)	(0.121)	(0.173)	(0.228)	(0.232)	(0.322)
Citations		-0.0238	-0.0256		-0.0020	-0.0012
		(0.0021)	(0.0027)		(0.0046)	(0.0055)
Low × Citations			-0.0001			-0.0030
			(0.0053)			(0.0114)
High × Citations			0.0097			-0.0020
			(0.0052)			(0.0145)
Constant	-0.813	-0.201	-0.158			
	(0.0543)	(0.0730)	(0.0817)			
Observations	2611	2601	2601	534	528	528

Models H1, H1c, and H1i show odds rations of logit regressions and standard errors pertaining to Hypothesis 1. The dependent variable is the decision to accept (1) or decline (0) the invitation to review the manuscript. Anonymized invitation conditions (AL, AA, AH) serve as the base category. Here, Low and High denote conditions LL and HH, respectively. Models H2, H2c, and H2i show odds ratios of ordered logit regressions and standard errors pertaining to Hypothesis 2. The dependent variable is the reviewer's recommendation: reject (1), major revision (2), minor revision (3), or accept (4). Condition AA is the base category. Here, Low and High denote conditions AL and AH, respectively. Citations represents the number of times the reviewer's total body of work had been cited according to Google Scholar at the time of data collection. For reasons of data protection, the variable consists of 85 bins, containing at least 20 observations each. The bins are ordered from the lowest to the highest number of citations. All models are estimated with Huber and White (robust) standard errors.

Supplementary Information to Nobel and novice: Author prominence affects peer review.

This document consists of 7 Sections. In Section 1 (Supplementary methods) we provide the template of the invitation email, a screenshot of the website reviewers saw when they accepted the invitation to review (consent page), a screenshot of the website reviewers saw when they declined the invitation to review, a screenshot of the questionnaire eliciting a reviewer's opinion on six statements about the paper, the post-review questionnaire and the template of the debriefing email. Section 2 (Supplementary discussion of limitations) deals with possible limitations of our study. In Section 3 (Supplementary discussion of deviations from the pre-registration document) we explore deviations from the pre-registered protocol. Section 4 (Supplementary discussion of selection effects) discusses the effects of reviewers' self-selection on their evaluation of the manuscript. In Section 5 (Supplementary discussion of ethical considerations) we discuss potential ethical concerns with this project. Section 6 (Supplementary discussion of related literature) expands on the related literature. The final section (Supplementary tables) provides additional data on the time participants spent in each stage of the experiment, and on attrition over the course of the study.

1. Supplementary methods

1.1. Invitation email

(text in italics only present in conditions LL and HH):

Subject:
Invitation to review for Journal of Behavioral and Experimental Finance

Email body:
Manuscript Number: 21-00864

Title: Re-tradable Assets, Speculation, and Economic Instability
Corresponding author: {{ author_name }}

Dear {{ first_name }} {{ last_name }},
I would like to invite you to review the above referenced manuscript, as I believe it falls within your expertise and interest. The abstract for this manuscript is included below.

You should treat this invitation, the manuscript and your review as confidential. You must not share your review or information about the review process with anyone without the agreement of the editors and authors involved, even after publication.

Please respond to this invitation at your earliest opportunity.

If you would like to review this paper, please click this link:
{{ accept_link }}

If you have a conflict of interest or do not wish to review this paper, please click this link: {{ decline_link }}

Since timely reviews are of utmost importance to authors, I would appreciate receiving your review within 30 days of accepting this invitation.

As a mark of appreciation for your timely review, we would be pleased to send you a reviewer reward amounting to $50. Please note that the reward is on a personal title and not transferable to an organization. Those reviewers that are not able to receive the reward on a personal level are kindly requested to waive it. The transfer will be made through the payment platform WISE.

I hope you will be able to review this manuscript.
Thank you in advance for your contribution and time.

Kind regards,

Stefan Palan
Editor-in-Chief
Journal of Behavioral and Experimental Finance

Title: Re-tradable Assets, Speculation, and Economic Instability
Corresponding author: {{ author_name }}

Abstract: This paper examines asset markets in which the key distinguishing characteristic of the goods is that they can be purchased for resale. Although the distinction between consumption durables and non-durables is clear and universally recognized, less evident is whether asset re-tradability accounts for economic instability. Market instability is strongly associated with goods that can be re-traded; stability with those that are bought for consumptive use. We emphasize the centrality of asset re-tradability in financial theory through a reinterpretation of the fundamental theorem of asset pricing: an arbitrage-free asset market is a market in which there is no advantage to re-trade any asset holdings. This result illustrates the inherent nature of the no-trade problem of neoclassical finance and suggests exploration of a different framework when it comes to dealing with asset re-tradability and speculation. We develop a relatively simple model of speculative asset price dynamics that generates excess, fat-tailed, and clustered volatility, three well-established empirical properties of financial volatility.

More information and support

You will find guidance and support on reviewing, as well as information including details of how Elsevier recognises reviewers, on Elsevier's Reviewer Hub: `https://www.elsevier.com/reviewers`

---

contact the editor if you have any questions.

1.2. Screenshots

Figure 1: Screenshot of consent page after accepting the review invitation

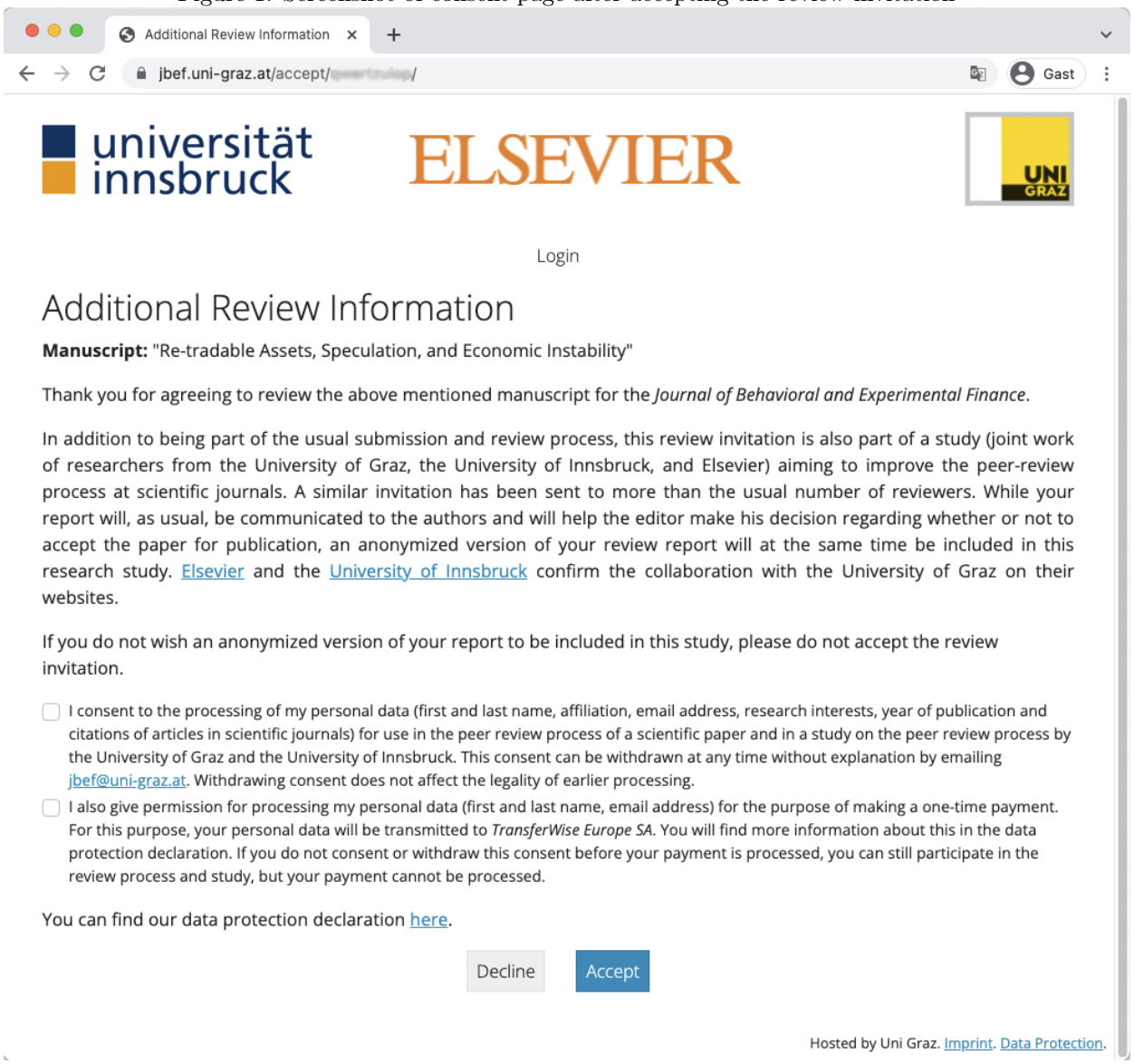


Figure 2: Screenshot of questionnaire page after declining the review invitation

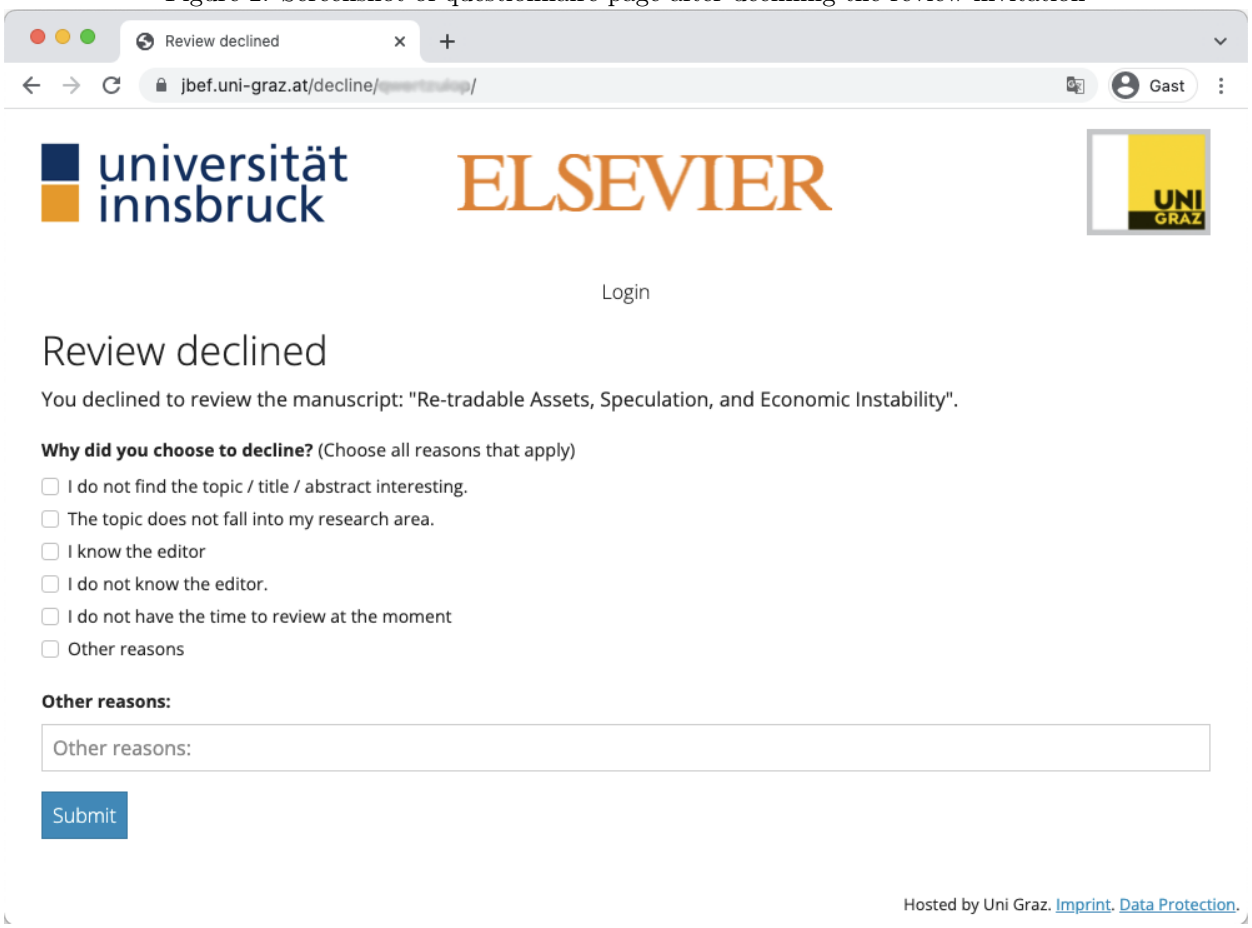
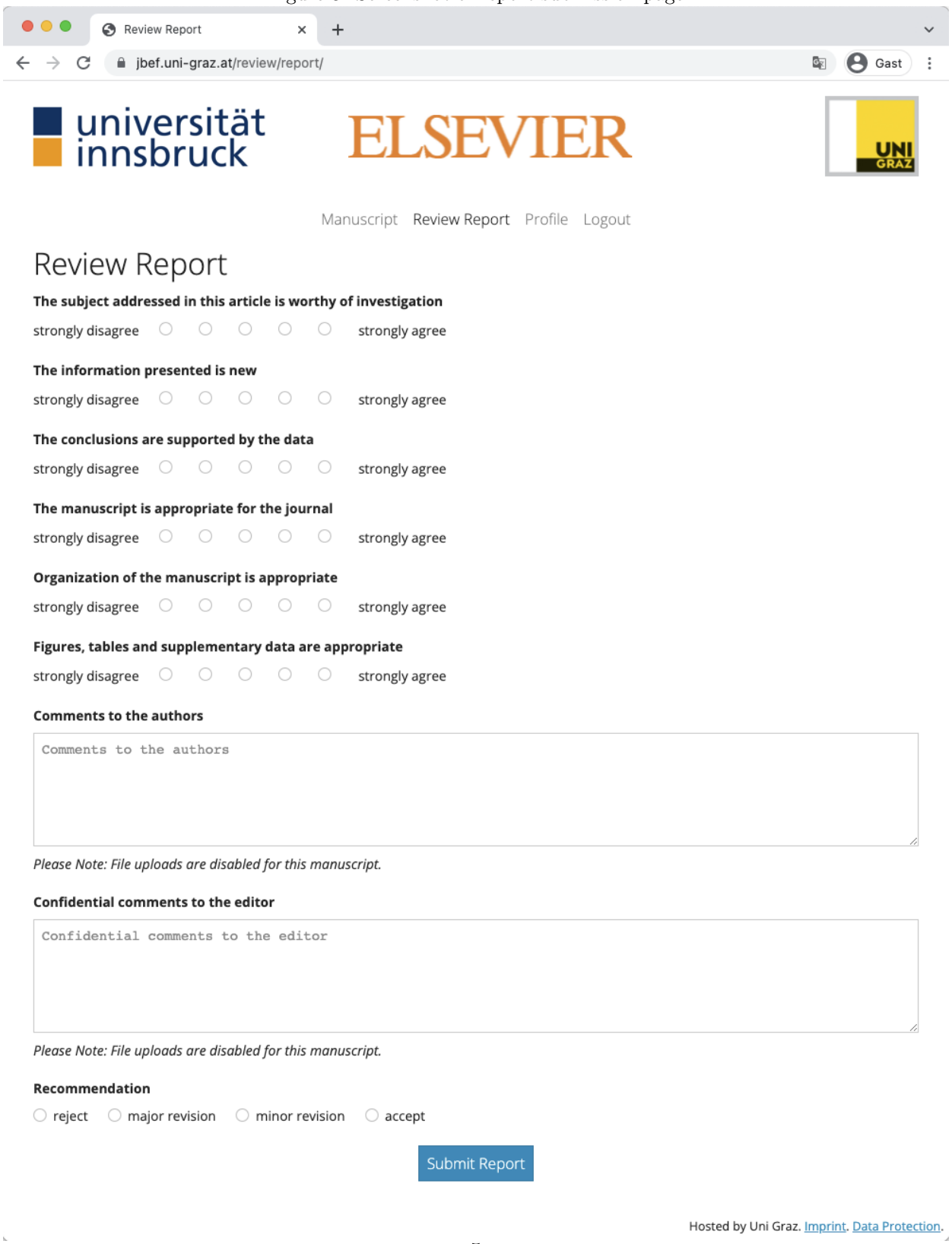


Figure 3: Screenshot of report submission page



1.3. Post-review questionnaire

General introduction

Thank you for having reviewed the paper "Re-tradable Assets, Speculation, and Economic Instability". We kindly ask that you complete your valuable contribution by answering the following (no more than 11) questions. Your answers are anonymous – we have no way of connecting you to the answers given in this questionnaire, nor do we wish to.

Questionnaire condition AA

1. How many papers do you review on average per year (including reviews of revised versions of papers)? ... papers

2. How long did you spend altogether on reading this paper and writing your review? ... hours

3. Would you say that you spent more, less, or the same amount of time on this review report compared to the amount of time you usually spend on preparing a review report?

- More
- Less
- The same
- Cannot say / do not wish to answer / not applicable

[page break]

4. What were the main motivations for you to accept the review invitation? (check all that apply)

- I found the topic/title/abstract interesting
- The topic falls into my research area
- I know the editor
- I feel a moral obligation to support science by reviewing
- Other: ...

[page break]

Questionnaire conditions AH and AL

Questions 1-4 are the same as in condition AA, in addition we ask:

5. Did you notice the corresponding author's name on the title page of the manuscript?

- Yes
- No

[page break]

[Only if "Yes" in question 5]

6. Learning that {{ author_name }} is the corresponding author of the manuscript:

- made me devote more time/effort than usual to reviewing the paper

- made me devote less time/effort than usual to reviewing the paper
- did not change the time/effort I devoted to reviewing the paper

7. Learning that {{ author_name }} is the corresponding author of the manuscript:

- made me assess the paper more positively
- made me assess the paper less positively
- did not change my assessment of the paper

Questionnaire conditions HH and LL
Questions 1-3 are the same as in condition AA, in addition we ask:

4. What were the main motivations for you to accept the review invitation? (check all that apply)

- I found the topic/title/abstract interesting
- The topic falls into my research area
- I know the editor
- I know the author
- I feel a moral obligation to support science by reviewing
- Other: ...

[page break]

[Only if NOT "I know the author" in question 4]
5. Did you notice the corresponding author's name in the review invitation?

- Yes
- No

[page break]

[Only if either "Yes" in question 5 or "I know the author" in question 4]
6. The corresponding author's name is {{ author_name }}. Before I accepted the review invitation,

- I looked up the corresponding author's track record
- I did not look up the corresponding author's track record but was nonetheless aware of it
- I was neither aware of nor did I look up the author's track record

7. Learning that {{ author_name }} is the corresponding author of the manuscript:

- made me more likely to accept the invitation
- made me less likely to accept the invitation
- did not affect my decision whether or not to accept the invitation
- Cannot say / do not wish to answer

[page break]

[Only if "No" in question 5.]
8. The corresponding author's name was listed on the title page of the manuscript. Were you aware of the corresponding author's name prior to submitting your review?

- Yes
- No

[page break]

[Only if either ("Yes" in question 5 and "I was neither aware of nor did I look up the author's track record." in question 6) or "Yes" in question 8.]
9. Before I submitted my review,

- I looked up the corresponding author's track record
- I did not look up the corresponding author's track record but was nonetheless aware of it
- I was neither aware of nor did I look up the author's track record

10. Learning that {{ author_name }} is the corresponding author of the manuscript:

- made me devote more time/effort than usual to reviewing the paper
- made me devote less time/effort than usual to reviewing the paper
- did not change the time/effort I devoted to reviewing the paper

11. Learning that {{ author_name }} is the corresponding author of the manuscript:

- made me assess the paper more positively
- made me assess the paper less positively
- did not change my assessment of the paper

1.4. Template of the debriefing email

Subject:
Information regarding your recent review for the Journal of Behavioral and Experimental Finance

Email body:
Dear {{ first_name }} {{ last_name }},
Recently, you reviewed a manuscript entitled "Re-tradable Assets, Speculation, and Economic Instability". We would like to thank you very much for participating and contributing your valuable time to the advancement of scientific progress. As we have written to you before you started reviewing the paper, this was part of a study (but your report is also used in the usual editorial process and is sent to the authors of the paper).

Let us take this opportunity to give you some background information on the study. The main research questions are: 1) How does author prominence affect the probability of accepting the invitation to review a manuscript? 2) How does author prominence affect the assessment of the manuscript, i.e., the recommendation to accept, revise, or reject? As any

systematic biases in the peer-review process are sand in the gears of science, answering these questions is of vital importance to the scientific community as a whole.

To answer our two research questions, it was necessary to invite more reviewers than usual to review the manuscript. Depending on the treatment group you were in, you received a combination of 1) an invitation email that revealed or did not reveal the corresponding author's name; 2) a non-anonymized or an anonymized manuscript, which either mentioned the corresponding author's name or did not. Depending on the treatment combination, you might have either seen Nobel prize laureate Vernon Smith or ESI research associate Sabiou Inoua (who both consented to this study) as the corresponding author of the manuscript. Be assured that both are the actual authors of the paper and there were no additional co-authors involved in writing the manuscript you reviewed. Our analysis will be conducted strictly on the aggregate level, relying on differences in aggregate behavior between treatment groups. Furthermore, the involved researchers did never get names, but only anonymized data and reports, hence, can and could never identify individuals. We will at no point and in no way reveal personally identifiable information of reviewers. We will also not link personally identifiable information to the report or the responses to the questionnaire. Safeguarding your anonymity is one of our top priorities.

Prior to conducting this study, the Internal Review Boards of the University of Graz and the University of Innsbruck were given the opportunity to review the project proposal and arrived at positive evaluations. The study protocol was developed in close collaboration with Elsevier.

We anticipate that you may be curious about our findings and we will therefore happily email you a copy of our research article reporting on this study as soon as it is available. If you wish us to do so (or have any other questions), please contact Stefan Palan, editor-in-chief of the Journal of Behavioral and Experimental Finance, via jbef@uni-graz.at.

Once again, we thank you very much for your effort in reviewing the manuscript and participating in this study. Your contribution is invaluable not only to us, but to every member of the scientific community.

Kind Regards,
Jürgen Huber, Rudolf Kerschbamer, Christian König-Kersting, and Stefan Palan

2. Supplementary discussion of limitations

To comply with the EU's *General Data Protection Regulation* as well as with university ethics guidelines, we informed researches 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 the data used to address our RQ2 regarding the rating and assessment of the paper by the reviewers (the data used to address our RQ 1 – willingness

of potential reviewers to assess the paper – was collected before asking for informed consent; this data is therefore not affected by these problems): First, researchers who gave their consent may differ in some characteristic of interest to those who withheld their consent – a selection problem. Second, 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. Nevertheless, ethical and legal considerations ruled out foregoing the informed consent stage. Furthermore, the clarity of our results and the fact that reviewer characteristics do not differ significantly between those who gave their consent and those who did not suggest that our findings are unlikely to be driven by reviewers' being informed about their participation in a study.

A second limitation is that we use only one paper in one field (finance) to study our research questions, such that we can not 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.

Third, while both authors of the paper being sent out for review, Sabiou Inoua and Vernon Smith, are affiliated with the same institution (Chapman University), holding one potential confounding factor constant, Inoua is dark-skinned while Smith is light-skinned. Skin-tone has long been identified as among the most salient markers for discrimination (Pierson et al., 2020; Riach and Rich, 2002). An influence of such discrimination (i.e., colorism) is thus also possible in our study.

A fourth limitation of our study is that while we present strong evidence for the existence of a status bias in reviewing we are unable to pin down the underlying mechanisms. A possible explanation for the identified positive component of the status bias in the invitation acceptance stage is that potential reviewers are more curious about reading a paper written by a prominent author than written by an anonymous or a comparatively unknown author. A plausible explanation for our main results regarding the evaluation of the manuscript is the 'halo effect' discussed in the body of the paper. While these explanations are plausible, our data does not permit testing such conjectures empirically. We see exploring the forces driving our main results as a fruitful area for further research.

3. Supplementary discussion of deviations from the pre-registration document

Our analysis largely follows the pre-registration. In a conservative deviation from the plan, we report two-sided tests instead of one-sided ones throughout the manuscript and the supplementary information. Apart from this deviation, we conduct few analyses that go beyond the pre-registration: 1) We conduct additional tests of proportions on the rejection rates between conditions. 2) We aggregate 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 add a robustness check to investigate the impact of removing outliers in terms of review time. 4) We conduct randomisation checks to make sure

reviewers' characteristics do not differ significantly between conditions. 4) We test whether the less prominent author benefits from fully anonymous compared to single-anonymous evaluation.

4. Supplementary discussion of selection effects

In discussing the impact of author prominence on the assessment of the manuscript, our focus in the body of the paper was on conditions AL, AA and AH, as these conditions all had an anonymized invitation. This allowed for a clean identification of the effect of author prominence on the evaluation of the manuscript without possible confounds caused by self-selection of reviewers at the invitation stage. Our methods section showed that the assessments in conditions AL and LL differ significantly ($z = -2.172, p = 0.0299$), while the assessments in condition AH are statistically indistinguishable from those in HH ($z = 0.774$, $p = 0.4387$). This suggests that selection may be an issue in the data of the less prominent author while it does not seem to have a material effect in the data of the more prominent author. Looking at Extended Data Table 2 we see that the main effect of selection is that those reviewers that accept the invitation to review the paper after having been informed that the corresponding author is the less prominent researcher are milder in their judgement than 'uninformed' reviewers (that is, reviewers that responded to the anonymous invitation): Informed reviewers recommend rejection in only 51.75 percent of the cases while uninformed reviewers reject the manuscript with the less prominent researcher as the corresponding author in 65.35 percent of the cases. Conversely, informed reviewers recommend accept or minor revision in 18.42 percent of the cases while uninformed reviewers do so in only 9.90 percent of the cases. This self-selection effect benefiting the less prominent author (partly) counteracts the negative status bias of the less prominent author's manuscript being evaluated less favorably. An important question then is whether less prominent authors still benefit from being evaluated anonymously as compared to a single-anonymized procedure. We can answer this question by comparing LL to AA. While the data still points in the same direction (anonymized review is better for less prominent authors than review with revelation of author names), the differences are no longer significant (for the rejection rate we have LL: 51.8 percent vs. AA: 48.2 percent; $z = 0.695$, $p = 0.4873$; all tests in this paragraph are two-sided Mann-Whitney U tests, the adjusted α-threshold for three tests is 0.0167).

This begs the question of whether our study was underpowered or, more generally, which magnitudes of effect sizes our study was designed to test. While we pre-registered a 'reasonable shift' in the proportions of accept, minor revision, major revision and reject recommendations from the journal's historical averages of $(.04, .22, .38, .36)$ to $(.08, .36, 0.30, .26)$ in section 4.2.2 of the pre-registration document, we did not explicitly state the corresponding effect size, raising the potential concern that our study may have been underpowered to detect small but relevant effect sizes. This is not the case. The standardized effect size of the 'reasonable shift' we pre-registered is $r = 0.175$, which is typically considered a small effect (Cohen, 1992). While the ex-post effect sizes we find between our treatments AL and AA and between AA and AH as well as between AA and HH are greater (AA vs. AL:

$r = 0.193$; AH vs. AA: $r = 0.376$; HH vs. AA: $r = 0.357$), the effect size between treatments LL and AA is too small to be detected with reasonable power given our sample size. (Note, however, that we only study the effect size between treatments LL and AA in a robustness check, since our pre-registered research questions and analysis plan do not rely on, or make statements about, this particular effect size.)

5. Supplementary discussion of ethical considerations

We are aware of several potential ethical concerns with this project. First, the study involves what may be seen as an "excessive" use of reviewer time, as we invited a significantly larger number of reviewers than the usual 2–4. We are aware that this is a substantial cost/burden to the scientific community. However, we took every possible step to ensure that the reviewers' time was used economically and only for the stated purpose. Most researchers regularly and voluntarily review papers, and (as is the norm at *JBEF*) every reviewer was offered the monetary compensation that is offered all *JBEF* reviewers for their reports (i.e., USD 50). Also, we obtained reviewers' informed consent, compelling nobody to participate. We consider the costs these reports put on the scientific community to be outweighed by the potential benefit to science overall (see last paragraph of this discussion). Note also that all reports were sent to the authors and were taken into account in reaching a decision on the eventual publication of the paper, hence no report/work was performed purely for the purposes of this study, or "wasted" in any way.

A second ethical concern is that we showed only one of two author names in some of the conditions. Here, we think this a minor concern that is far outweighed by the potential benefit to the scientific community. Also note that we referred to the shown author as "corresponding author" (which does not preclude there being other authors) and never listed an author who is not actually an author of the paper (i.e., Smith and Inoua are the sole two authors of the paper and we only gave their names). Furthermore, there are scientific journals where it is the norm to get review invitations that name only the corresponding author (for example some journals of the publisher Wiley, like the *German Economic Review* or the *Journal of Public Economic Theory*), which is exactly what we did. Once all review reports were in, we also sent a debriefing email to all reviewers, informing them about the design and research questions of the project.

A third ethical concern is that reviewers might not be comfortable with the fact that they are being used as participants in an experiment for fear of their integrity as researchers being tested. In this regard, we first highlight that this is not a study on researcher integrity. We believe that the biases we have identified in this study are most likely due to an unconscious reaction to the experimental conditions we applied and not due to conscious discrimination. Second, and at least as important, given our across-subjects design in which reviewers were exposed to only one experimental condition, all conclusions we draw are based on a comparison of the aggregate behavior of reviewers in different conditions. Neither we nor anyone else could measure a potential bias at the individual level. Third, only consenting reviewers participated in the study. Fourth, we ensured that the anonymity of our participants was

preserved to the greatest extent possible. All data and reports were anonymized prior to even being shared with the full team of researchers.

Regarding benefits, the results of this project shed light on the important question of which role author prominence has for the readiness of researchers to accept a review invitation (RQ1) and, more importantly, for the assessment of the paper (RQ2). Uncovering the extent of the status bias in peer review is important because any bias in the review process is sand in the gears of science. For instance, our result that younger researchers are systematically disadvantaged in the review process indicates that science progresses more slowly than it could, as it is often younger researchers who contribute innovative ideas and drive major breakthroughs. The outcomes of this study have major policy implications. The most obvious contribution of this project is probably the one to the highly relevant and hotly debated question of whether peer-review should be double-anonymized, single-anonymized or even fully transparent (open peer review).

6. Supplementary discussion of related literature

Here we discuss the relation of our work to Peters and Ceci (1982), Blank (1991), Garfunkel et al. (1994), Madden and DeWitt (2006), Tung (2006), Okike et al. (2016), Tomkins et al. (2017) and Card and DellaVigna (2020).

The paper closest to ours in terms of research questions is Tomkins et al. (2017). The authors investigate the impact of making author information available to reviewers (single-anonymized versus double-anonymized) in two stages of the process of reviewing submissions to a prominent computer science conference, first in a preliminary "bidding" stage in which reviewers express interest in papers to review and second in a reviewing stage in which reviewers give a recommendation regarding acceptance for presentation at the conference. Four expert committee members review each submission; two of the four receive access to author information while the other two do not. For the bidding stage, the authors find that reviewers in the single-anonymized condition typically bid for fewer papers and preferentially bid for papers from top universities and companies. No clear 'famous author' bias is identified in this stage. For the reviewing stage, the authors find that single-anonymized reviewers are significantly more likely than their double-anonymized counterparts to submit a positive review for papers from famous authors, top universities, and top companies.

In our view this is an elegant study that is clearly related to our project because it also investigates two stages of the review process, the bidding stage, which is related to our RQ1, and the review stage, related to our RQ2. An advantage of our design is that (in the single-anonymized condition) we vary the prestige of the author while keeping everything else (incl. the institution of the author and the quality of the manuscript) constant, while in Tomkins et al. (2017) the prestige bias can only be inferred indirectly (across many different manuscripts). As a consequence, there is some leeway in the interpretation of their results. For instance, the finding regarding RQ1 that reviewers preferentially bid for papers from top universities could be confounded with the papers submitted by authors from top places

simply being better manuscripts. By keeping the paper quality constant and varying the author name revealed to reviewers we can cleanly identify whether there is a bias in the willingness to accept that is purely related to the prominence of the author. Besides the fact that we have more control, there are other subtle differences in the designs of the experiments. An important one is that theirs is an experiment on the performance of a conference review process, while we are interested in the review process for academic journals. An important difference between those review processes is that for conferences, reviewers typically have to review several (sometimes even many) papers with a single deadline for completing all reviews. A consequence is that reviewers are typically under time pressure and dedicate only limited time to each paper. Here it is quite plausible that the prominence of the author influences the decision of a reviewer. It is not at all clear that this result translates to the journal review process. Another important issue regarding RQ2 is that Tomkins et al. (2017) cannot control for selection at the first stage (in their study, the pools of reviewers in the two conditions are not identical, as they are the result of a paper allocation mechanism based on the decisions in the bidding stage), while we can (by comparing AA to AL and AH; in those conditions the reviewers all receive a blinded invitation in the first stage).

While we are not aware of any other previous research addressing our RQ1 (effect of author prestige on the willingness to accept the invitation to act as a reviewer), there is some literature addressing RQ2 (effect of author prestige on the recommendation regarding publication). A famous and rather controversial study addressing RQ2 is Peters and Ceci (1982). For this study, the authors selected 12 research articles already published in highly regarded and widely read psychology journals. They then changed the author names (from real names to fictitious ones) and the institutional affiliations of the authors (from prestigious American psychology departments to fictitious institutions). The so manipulated manuscripts were then resubmitted to the journals that had originally reviewed and published them 18 to 32 months earlier. Of the 12 papers only 3 were detected as resubmissions. Of the 9 papers that continued through the review process, 8 were rejected. Peters and Ceci (1982, p. 192) interpret the change from acceptance to rejection as suggestive of bias based on authors' affiliations. However, there are several details in this study that weaken the interpretation of the results. First, the authors change author names and institutional names at the same time; second, regarding institutions, they not only change the relative prestige but also change from universities to non-academic institutions, from existing institutions to non-existing ones, etc; third, the field may have moved on since the papers were originally submitted and papers that were innovative at the time of the original submission may no longer have been innovative at the time of the experiment; fourth, and most importantly, with their design (involving resubmissions of already published papers) the authors cannot distinguish bias in the review process from pure randomness in this process (i.e., if paper acceptance was a purely random process, selecting already published papers and resubmitting them would also lead to some of them being accepted and some rejected – a result that is indistinguishable from the one the authors report).

Another prominent study addressing RQ2 is Blank (1991). This experimental study was originally initiated due to concerns about gender bias and in the end found some evi-

dence for status bias in the review process. In the experiment, every paper that arrived at the American Economic Review over a two-year period was randomly assigned to either a single-anonymized or a double-anonymized condition. The author finds that authors at top-ranked departments and those at colleges and low-ranked universities do not experience significant differences in acceptance decisions based on whether they went through the single-anonymized or double-anonymized reviewing process. However, authors at mid-tier institutions perform better in a single-anonymized setting, as do foreign authors and those from outside academia.

Another interesting experimental study related to our RQ2 is Okike et al. (2016). The authors fabricated an artificial submission to a journal and listed as authors two prominent researchers. The article was then sent to 256 reviewers, with half of the reviewers in a single-anonymized condition and the other half in a double-anonymized condition. Based on the 119 reports they received, the authors find that reviewers were more likely to recommend acceptance when the prestigious authors' names and institutions were visible (single-anonymized review) than when they were redacted (double-anonymized review) and also gave higher ratings for the methods and other categories. This paper shares with our study the property that one and the same manuscript is evaluated by many reviewers. A disadvantage of their design compared to ours is that they have only two conditions to compare for RQ2 (double-anonymized vs single-anonymized with prominent authors) while we have three (double-anonymized vs single-anonymized with prominent author vs single-anonymized with relatively unknown author). A consequence is that part of the effects they are reporting might be due to blinding and not due to the prominence of the authors. With our design we can cleanly control for this shortcoming.

The papers discussed up to now are experimental papers. There are also some retrospective studies addressing issues related to RQ2. An early one is Garfunkel et al. (1994). The authors address the question whether manuscripts from institutions with higher prestige are more likely to be recommended for publication by reviewers and to ultimately be accepted for publication. Their main results are that manuscripts from institutions with higher prestige were no more likely to be recommended or accepted for publication than those from institutions with lower prestige. In contrast, the likelihood of recommendation for acceptance and of selection for publication of brief reports appeared to correlate with the prestige of the institution. Relatedly, Madden and DeWitt (2006) address the question of whether the use of double-anonymized review significantly impacts the rate at which "more senior" researchers' papers get accepted at two database conferences. The authors find that double-anonymized review essentially does not affect the acceptance rates of more senior researchers. A follow-up study by Tung (2006) analyzes the same data and comes to the opposite conclusion, i.e., that double-anonymized review did have an impact in terms of papers accepted from more senior authors at one of the conferences.[1]

[1]There is also a large literature on scientific peer review in general and on single-anonymized vs. double-anonymized review in particular. We have discussed some papers and results that directly relate to our research questions. Cox et al. (1993) and Snodgrass (2006) provide more general summaries of the literature.

Card and DellaVigna (2020) present evidence suggesting the presence of something like a reversed status bias. The authors study editorial decision-making using anonymized submission data for four leading economics journals and match papers to the publication records of authors at the time of submission and to the manuscripts' subsequent Google Scholar citations (as a measure of quality). The authors show that reviewer recommendations are strong predictors of citations, and that editors follow the recommendations quite closely. Regarding the status bias they find that the submissions from more prominent authors receive substantially more citations than those from other authors. From the results of the previous literature, we would expect that this finding is at least in part due to the fact that more prominent researchers get more citations for the same quality of papers than do less prominent researchers. This explanation is dismissed by the authors based on the results of a survey of faculty and PhD students in economics. Based on the results of this survey the authors conclude that the editorial decision process at top economics journals nearly maximizes the expected quality of accepted papers, with the important exception that reviewers and editors impose a higher bar for submissions from more prolific authors.

7. Supplementary tables

Table 1: **Times reviewers took for responses, consent, and questionnaire submission**

	N	mean	std. dev.	minimum	median	maximum
Hours to response						
AL	585	57.45	96.59	0.008056	8.576	948.6
AA	455	68.81	140.3	0.01361	10.78	2144
AH	550	64.92	134	0.005556	10.65	2310
LL	606	60.21	125	0.008333	9.487	2180
HH	410	62.53	111.8	0.01	12.45	1590
Total	2606	62.45	122.1	0.005556	9.771	2310
Hours to consent						
AL	141	55.86	85.33	0.07444	12.28	510.1
AA	147	61.51	94.04	0.07306	14.19	612.9
AH	139	66.99	116.5	0.0475	18.93	786.3
LL	150	64.14	103.6	0.04	14.91	656.1
HH	142	62.35	84.4	0.04306	13.87	367.6
Total	719	62.18	97.28	0.04	14.26	786.3
Minutes on consent page						
AL	140	7.851	57.15	0	0.5583	651.5
AA	146	3.904	13.62	0.01667	1.008	140.9
AH	134	1.683	3.355	0.03333	0.925	32.8
LL	148	10.92	111.5	0	0.8667	1358
HH	139	29.5	204.8	0.06667	0.9167	2094
Total	707	10.77	107.5	0	0.8667	2094
Hours from consent to report submission						
AL	101	596.5	313.8	0.69	667.7	1356
AA	110	617.1	280.8	0.6206	673.4	1441
AH	102	624.3	289	0.1303	668.9	1372
LL	114	601.3	286.2	0.1061	667.2	1336
HH	107	626.1	319.1	0.4133	665.1	1628
Total	534	613	297	0.1061	669.5	1628
Minutes from report submission to questionnaire submission						
AL	86	2.921	5.261	0.6667	2.017	38.87
AA	95	1.856	1.015	0.55	1.7	5.433
AH	93	294.8	2820	0.6333	1.883	27201
LL	97	2.713	1.971	0.6833	2.25	15.28
HH	91	3.279	2.091	0.8667	2.667	14.58
Total	462	61.49	1265	0.55	2.083	27201

Table 2: **Attrition throughout the study.**

	Total		AL		AA		AH		LL		HH	
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Pool	5529											
invited	3299	59.67	739	13.37	576	10.42	696	12.59	781	14.13	507	9.17
Responses	(of 3299)		(of 739)		(of 576)		(of 696)		(of 781)		(of 507)	
accept or decline	2611	79.15	585	79.16	455	78.99	551	79.17	610	78.10	410	80.87
no response	688	20.85	154	20.84	121	21.01	145	20.83	171	21.90	97	19.13
Invitations	(of 2611)		(of 585)		(of 455)		(of 551)		(of 610)		(of 410)	
accepted	821	31.44	163	27.86	161	35.38	165	29.95	174	28.52	158	38.54
accepted, then declined	105	4.02	26	4.44	21	4.62	19	3.45	19	3.11	20	4.88
accepted, never declined	716	27.42	137	23.42	140	30.77	146	26.50	155	25.41	138	33.66
declined	1895	72.58	448	76.58	315	98.46	405	73.50	455	74.59	272	66.34
Consent	(of 821)		(of 163)		(of 161)		(of 165)		(of 174)		(of 158)	
given	671	81.73	132	80.98	141	87.58	129	78.18	135	77.59	134	84.81
given, then withdrawn	4	0.49	1	0.61	1	0.62	1	0.61	0	0.00	1	0.63
given, never withdrawn	667	81.24	131	80.37	140	86.96	128	77.58	135	77.59	133	84.18
not given	154	18.76	32	19.63	21	13.04	37	22.42	39	22.41	25	15.82
Reports	(of 671)		(of 132)		(of 141)		(of 129)		(of 135)		(of 134)	
submitted	534	79.58	101	76.52	110	78.01	102	79.07	114	84.44	107	79.85
not submitted	137	20.42	31	23.48	31	21.99	27	20.93	21	15.56	27	20.15
Questionnaires	(of 534)		(of 101)		(of 110)		(of 102)		(of 114)		(of 107)	
submitted	462	86.52	86	85.15	95	86.36	93	91.18	97	85.09	91	85.05
not submitted	72	13.48	15	14.85	15	13.64	9	8.82	17	14.91	16	14.95

References

Blank, R., 1991. The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review. American Economic Review 81, 1041–1067.

Card, D., DellaVigna, S., 2020. What do editors maximize? Evidence from four economics journals. Review of Economics and Statistics 102, 195–217.

Cohen, J., 1992. A power primer. Psychological Bulletin 112, 155–159.

Cox, D., Gleser, L., Perlman, M., Reid, N., Roeder, K., 1993. Report of the ad hoc committee on double-blind refereeing. Statistical Science 8, 310–317.

Garfunkel, J., Ulshen, M., Hamrick, H., Lawson, E., 1994. Effect of institutional prestige on reviewers' recommendations and editorial decisions. Journal of the American Medical Association 272, 137–138.

Madden, S., DeWitt, D., 2006. Impact of double-blind reviewing on sigmod publication rates 35, 29–32. URL: https://doi.org/10.1145/1147376.1147381, doi:10.1145/1147376.1147381.

Okike, K., Hug, K., Kocher, K., Leopold, S., 2016. Single-blind vs double-blind peer review in the setting of author prestige. Journal of the American Medical Association 316, 1315–1316.

Peters, D., Ceci, S., 1982. Peer-review practices of psychological journals: The fate of published articles, submitted again. Behavioral and Brain Sciences 5, 187–195.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., et al., 2020. A large-scale analysis of racial disparities in police stops across the united states. Nature Human Behaviour 4, 736–745.

Riach, P.A., Rich, J., 2002. Field experiments of discrimination in the market place. Economic Journal 112, F480–F518.

Snodgrass, R., 2006. Single- versus double-blind reviewing: An analysis of the literature. SIGMOD Rec. 35, 8–21. URL: https://doi.org/10.1145/1168092.1168094, doi:10.1145/1168092.1168094.

Tomkins, A., Zhang, M., Heavlin, W., 2017. Reviewer bias in single- versus double-blind peer review. PNAS 114, 12708–12713.

Tung, A.K.H., 2006. Impact of double blind reviewing on sigmod publication: A more detail analysis. SIGMOD Rec. 35, 6–7. URL: https://doi.org/10.1145/1168092.1168093, doi:10.1145/1168092.1168093.