PNAS
www.pnas.org

Supplementary Information for

Nobel and novice: Author prominence affects peer review

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

Juergen Huber and Vernon L. Smith.
E-mails: juergen.huber@uibk.ac.at and vlomaxsmith@gmail.com

This PDF file includes:

Supporting Information Text

Supplementary discussion of materials and methods.

Reviewer selection and disambiguation. We adopted a software-assisted approach to reviewer selection and disambiguation. From all articles published in the selected journals (Table S3) we collected names and email addresses from all authors that provided this information on their publication. Some journals require contact information only for corresponding authors, others require it for all authors. If not available on the article, affiliations were gathered from email domains and Google Scholar (where possible). Affiliations were also matched against a large database of universities and research institutions which also contains the institutions' names in multiple languages. This was important because institutions were the level of randomization and we therefore needed to make sure that reviewer's affiliations were correctly represented in our data set. Our database of institutions is formed from a combination of the university data (identifier Q3918) available on wikidata.org (1) and the domain and name data set by Hipolabs.com (2). Email domains that were not found in the combined data set were researched individually and, if successful, the details of the institution were added to our data set manually. Author disambiguation was done based on names, email addresses, and affiliations. Names and email domains were also used to match Google Scholar profiles. If multiple profiles were potential candidates, we searched for the article within the set of potential profiles to identify the correct one. In addition, authors with identical first and second names were checked for duplicates by hand.

Reviewer characteristics. We collected the following characteristics from our pool of reviewers' Google Scholar profiles: citation count, citation count (5 years), h-index, h-index (5 years), i10-index, i10-index (5 years), and the year of the latest and the year of the earliest publication on record. We used the latter two dates to calculate the years of research activity for each reviewer. Our variable "years active" is defined as the difference between the year of the latest publication and the year of the earliest publication on record on Google Scholar.

As described in our section on data handling and reviewer anonymity, we recoded the characteristics to reduce the re-identification risk for our reviewers. All analyses and statistics that involve reviewer characteristics were based on the resulting categorical data. Table S5 shows the categories for each variable, including the interval boundaries. Table S6 shows reviewer characteristics by treatments. Importantly, the summary statistics are based on the categorical data.

Robustness: Reviewer characteristics. As a first set of robustness checks, we ran additional regressions (reported in Table S8) 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 found that experience decreased the propensity to accept the review invitation (the main effect of the number of citations was negative) and we also found that more experienced authors were 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' was 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 had neither a main effect nor an interaction effect with our experimental conditions and its inclusion left all of our findings qualitatively unchanged. The results regarding reviewer experience also remained qualitatively unchanged if we replaced the number of Google Scholar citation by the h-index, the i10-index or academic age.

Robustness: Different outcome measures. Second, we tested whether the status bias we identified in review recommendations and in the individual questions regarding the quality of the manuscript remained detectable if we created an overall score by averaging the ratings given in the six individual questions. As expected, a Kruskal-Wallis test revealed significant differences in the average scores between treatments. In Table S9 we report summary and test statistics for each question individually as well as for the average of all ratings.

In addition, we analyzed the relationship between our main outcome measure, review recommendations, and the individual quality questions. As a first step, we regressed the recommendation on the responses to the individual items (after standardization). The results are shown in Table S11. All coefficients besides "Subject worthy of investigation" and "Supplement and Figures appropriate" are statistically different from zero (and highly so). Furthermore, the coefficient of "Manuscript organization appropriate" is significantly higher than the coefficient of the next largest ("Manuscript appropriate for journal"; $F(1, 517) = 4.18$, $p = 0.0416$). As a second step, we conducted a principal component analysis. The results are shown in Tables S13 and S14. The analysis revealed that the first component accounts for almost 73% of the total variance and loads approximately equally on all of the individual items.

Robustness: Outliers. Third, we investigated 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, left all of our findings qualitatively unchanged.

Robustness: Timing. Fourth, we 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, (iii) the time between the consent and the report submitted, and (iv) the time it took to submit the post-review questionnaire. Differences in timing between treatment groups might serve as proxies for different levels of dedication, effort, or rigor by reviewers. The only statistically significant difference we found was in the time it took reviewers to complete the post-review questionnaire. However, this was clearly driven by the unequal number of questions that we asked in the different treatments. We report the timings in Table S7 below.

Robustness: Report length. Finally, we checked for differences in the length (number of words) of the reports submitted by reviewers. We did not find any statistical significant differences in the length of the comments to the authors, the confidential comments to the editor, or their combination. We take this as further indication that reviewers' effort and rigor put into the review was not affected by our treatments. Report lengths are reported in Table S10.

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 differed significantly ($z = -2.172, p = 0.0299$), while the assessments in condition AH were statistically indistinguishable from those in HH ($z = 0.774$, $p = 0.4387$). This suggested that selection may have been an issue in the data of the less prominent author while it did not seem to have had a material effect in the data of the more prominent author. Table S2 shows that the main effect of selection was that those reviewers that accepted the invitation to review the paper after having been informed that the corresponding author was the less prominent researcher, were milder in their judgement than 'uninformed' reviewers (that is, reviewers that responded to the anonymous invitation): Informed reviewers recommended rejection in only 51.75 percent of the cases while uninformed reviewers recommended rejecting the manuscript with the less prominent researcher as the corresponding author in 65.35 percent of the cases. Conversely, informed reviewers recommended accept or minor revision in 18.42 percent of the cases while uninformed reviewers did so in only 9.90 percent of the cases. This self-selection effect benefiting the less prominent author (partly) counteracted the negative status bias of the less prominent author's manuscript having been evaluated less favorably. An important question then was whether less prominent authors still benefited from being evaluated anonymously as compared to a single-anonymized procedure. We were able to answer this question by comparing LL to AA. While the data still pointed in the same direction (anonymized review being better for less prominent authors than review with revelation of author names), the differences were no longer significant (for the rejection rate we had LL: 51.8 percent vs. AA: 48.2 percent; $z = 0.695$, $p = 0.4873$; all tests in this paragraph were two-sided Mann-Whitney U tests, the adjusted α-threshold for three tests was 0.0167).

This begged the question of whether our study was underpowered or, more generally, which magnitudes of effect sizes our study had been 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 was not the case. The standardized effect size of the 'reasonable shift' we pre-registered was $r = 0.175$, which is typically considered a small effect (3). While the ex-post effect sizes we found between our treatments AL and AA and between AA and AH as well as between AA and HH were 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 was too small to be detected with reasonable power given our sample size. (Note, however, that we only studied the effect size between treatments LL and AA in a robustness check, since our pre-registered research questions and analysis plan did not rely on, or make statements about, this particular effect size.)

Supplementary discussion of ethical considerations. We are aware of several potential ethical concerns with this project. First, the study involved 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 was 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 considered 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 was that we showed only one of two author names in some of the conditions. Here, we thought this a minor concern that was far outweighed by the potential benefit to the scientific community. Also note that we referred to the shown author as "corresponding author" (which did not preclude there being other authors) and never listed an author who was 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 was 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 was that reviewers might not have been comfortable with the fact that they were being used as participants in an experiment for fear of their integrity as researchers being tested. In this regard, we first highlight that this was not a study on researcher integrity. We believe that the biases we identified in this study were 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 drew were 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 were systematically disadvantaged in the review process indicated 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).

Supplementary discussion of related literature. Here we discuss the relation of our work to Peters and Ceci (4), Blank (5), Fisher et al. (6), Garfunkel et al. (7), Madden and DeWitt (8), Tung (9), Alam et al. (10), Okike et al. (11), Tomkins et al. (12) and Card and DellaVigna (13).

The paper closest to ours in terms of research questions was Tomkins et al. (12). The authors investigated 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 expressed interest in papers to review and second in a reviewing stage in which reviewers gave a recommendation regarding acceptance for presentation at the conference. Four expert committee members reviewed each submission; two of the four received access to author information while the other two do not. For the bidding stage, the authors found 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 was identified in this stage. For the reviewing stage, the authors found that single-anonymized reviewers were 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 was an elegant study that was clearly related to our project because it also investigated two stages of the review process: the bidding stage, which was related to our RQ1, and the review stage, related to our RQ2. An advantage of our design was that (in the single-anonymized condition) we varied 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. (12) the prestige bias could only be inferred indirectly (across many different manuscripts). As a consequence, there was 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 were able to cleanly identify whether there was a bias in the willingness to accept that was purely related to the prominence of the author. Besides the fact that we had more control, there are other subtle differences in the designs of the experiments. An important one was that theirs was an experiment on the performance of a conference review process, while we were 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 was that Tomkins et al. (12) could not control for selection at the first stage (in their study, the pools of reviewers in the two conditions were not identical, as they were the result of a paper allocation mechanism based on the decisions in the bidding stage), while we could (by comparing AA to AL and AH; in those conditions the reviewers all received an anonymized invitation in the first stage).

While we were 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 was some literature addressing RQ2 (effect of author prestige on the recommendation regarding publication). A famous and rather controversial study addressing RQ2 was Peters and Ceci (4). 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 (4) interpreted the change from acceptance to rejection as suggestive of bias based on authors' affiliations. However, there were several details in this study that weakened the interpretation of the results. First, the authors changed author names and institutional names at the same time; second, regarding institutions, they not only changed the relative prestige but also changed from universities to non-academic institutions, from existing institutions to non-existing ones, etc; third, the field may have had moved on since the papers had been originally submitted and papers that had been 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 were unable to 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 have led to some of them being accepted and some rejected – a result that would be indistinguishable from the one the authors reported).

Another prominent study addressing RQ2 was Blank (5). This experimental study had originally been initiated due to concerns about gender bias and in the end found some evidence 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 found that authors at top-ranked departments and those at colleges and low-ranked universities did 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 performed better in a single-anonymized setting, as did foreign authors and those from outside academia.

Similar in design was a study by Fisher et al. (6). In a randomized controlled trial, 57 manuscripts that had been submitted to a pediatrics journal were sent out to four referees each, with two referees receiving an anonymized version of the manuscript and the other two receiving a non-anonymized version. Each reviewer was asked to provide a narrative review and a score ranging from 1 (accept) to 5 (reject). For each manuscript the author with the largest number of publications was designated the senior author of the manuscript. Fisher et al. hypothesized that if a status bias was present in the evaluation of manuscripts, there should be a stronger correlation between non-anonymized scores and the number of previous publications by the senior author than between anonymized scores and the number of previous publications. Contrary to this hypothesis, the study found that reviewers in the anonymized condition favored authors with more publications while reviewers in the non-anonymized condition did not. The authors interpreted this finding as indicating that the former "may have recognized improved quality in the work of those authors with more previous publications. In contrast, reviewers who were aware of author identity did not give better scores to more experienced authors, likely indicating that various types of bias may have entered into their thinking." (p. 146).

One more study similar in design to the two studies discussed in the previous paragraphs was Alam et al. (10). In this research, 40 manuscripts that had been submitted to a dermatology journal were assessed by four reviewers each, two of whom were randomly chosen to receive an anonymized version, while the other two received a non-anonymized version. The primary outcome measure was the initial score (ranging from 1 for accept to 3 for reject) assigned by each reviewer. As a secondary variable of interest the authors also considered the word count of the narrative portion of the reviewer forms (comments to the authors and comments to the editor). The authors found neither a significant difference in the scores nor a difference in word count between anonymized and non-anonymized reviews.

Another interesting experimental study related to our RQ2 was Okike et al. (11). 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 found 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 was evaluated by many reviewers. An advantage of our design compared to theirs was that they had only two conditions to compare for RQ2 (double-anonymized vs. single-anonymized with prominent authors) while we had three (double-anonymized vs. single-anonymized with prominent author vs. single-anonymized with relatively unknown author). A consequence was that part of the effects they reported might have been due to anonymization itself, and not due to the prominence of the authors. With our design we were able to cleanly control for this potential confound.

The papers discussed up to now were experimental papers. There were also some retrospective studies addressing issues related to RQ2. An early one was Garfunkel et al. (7). The authors addressed the question whether manuscripts from institutions with higher prestige were more likely to be recommended for publication by reviewers and to ultimately be accepted for publication. Their main results were 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 (8) addressed 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 found that double-anonymized review essentially did not affect the acceptance rates of more senior researchers. A follow-up study by Tung (9) analyzed the same data and came 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.*

Card and DellaVigna (13) presented evidence suggesting the presence of something like a reversed status bias. The authors studied editorial decision-making using anonymized submission data for four leading economics journals and matched 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 showed that reviewer recommendations were strong predictors of citations, and that editors followed the recommendations quite closely. Regarding the status bias they found that the submissions from more prominent authors received substantially more citations than those from other authors. From the results of the previous literature, we would expect that this finding was 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 was 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 concluded 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 imposed a higher bar for submissions from more prolific authors.

There is also a large literature on scientific peer reviewing in general and on single-anonymized vs. double-anonymized review in particular. Several recent papers (16–18) 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. Surveys on the literature comparing single vs. double anonymization are provided by Cox et al. (14) and Snodgrass (15), and – most recently – by Ucci et al. (19). Based on 11 randomized controlled trails, the latter study finds that the manuscript acceptance rate is significantly lower in the double-anonymized than in the single-anonymized peer review process.

*There is also a large literature on scientific peer review in general and on single-anonymized vs. double-anonymized review in particular. We discussed some papers and results that directly relate to our research questions. Cox et al.(14) and Snodgrass (15) and provided more general summaries of the literature.

Template of 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

---

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. Use the following URL: {{ delete_link }}. Please contact the editor if you have any questions.

Template of 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

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

Supplementary Figures.

Fig. S1. Screenshot of consent page after accepting the review invitation

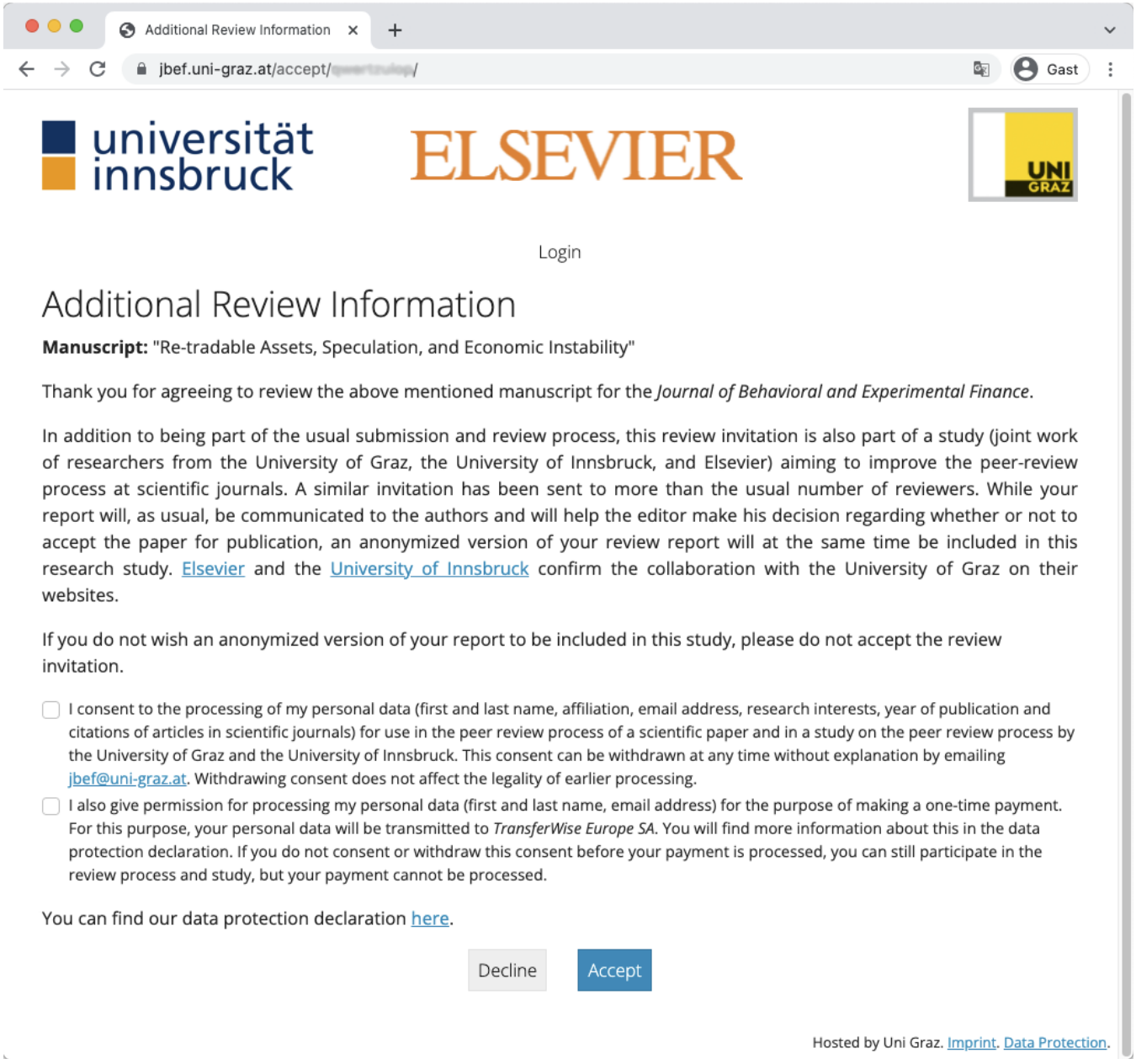


Fig. S2. Screenshot of questionnaire page after declining the review invitation

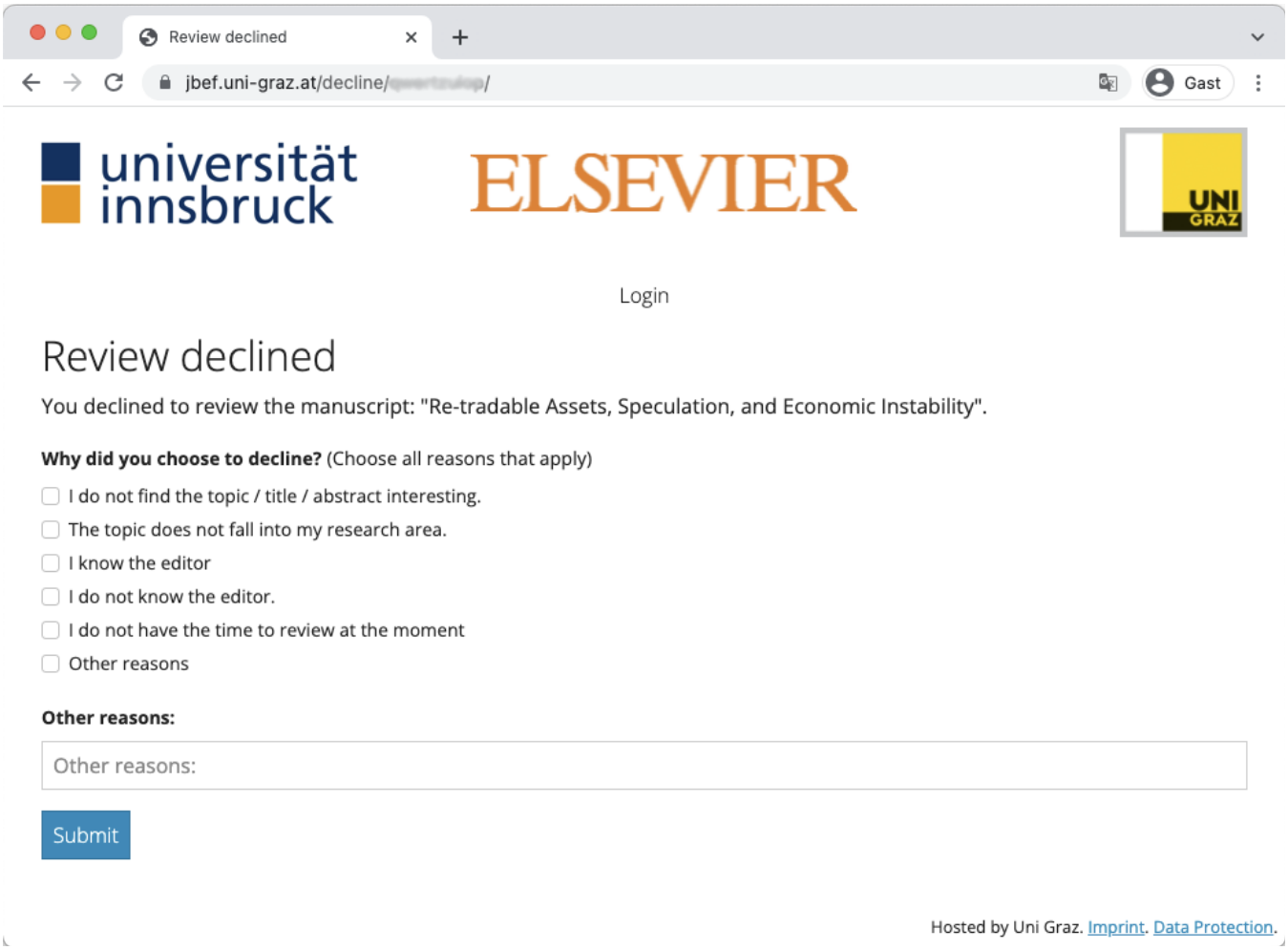
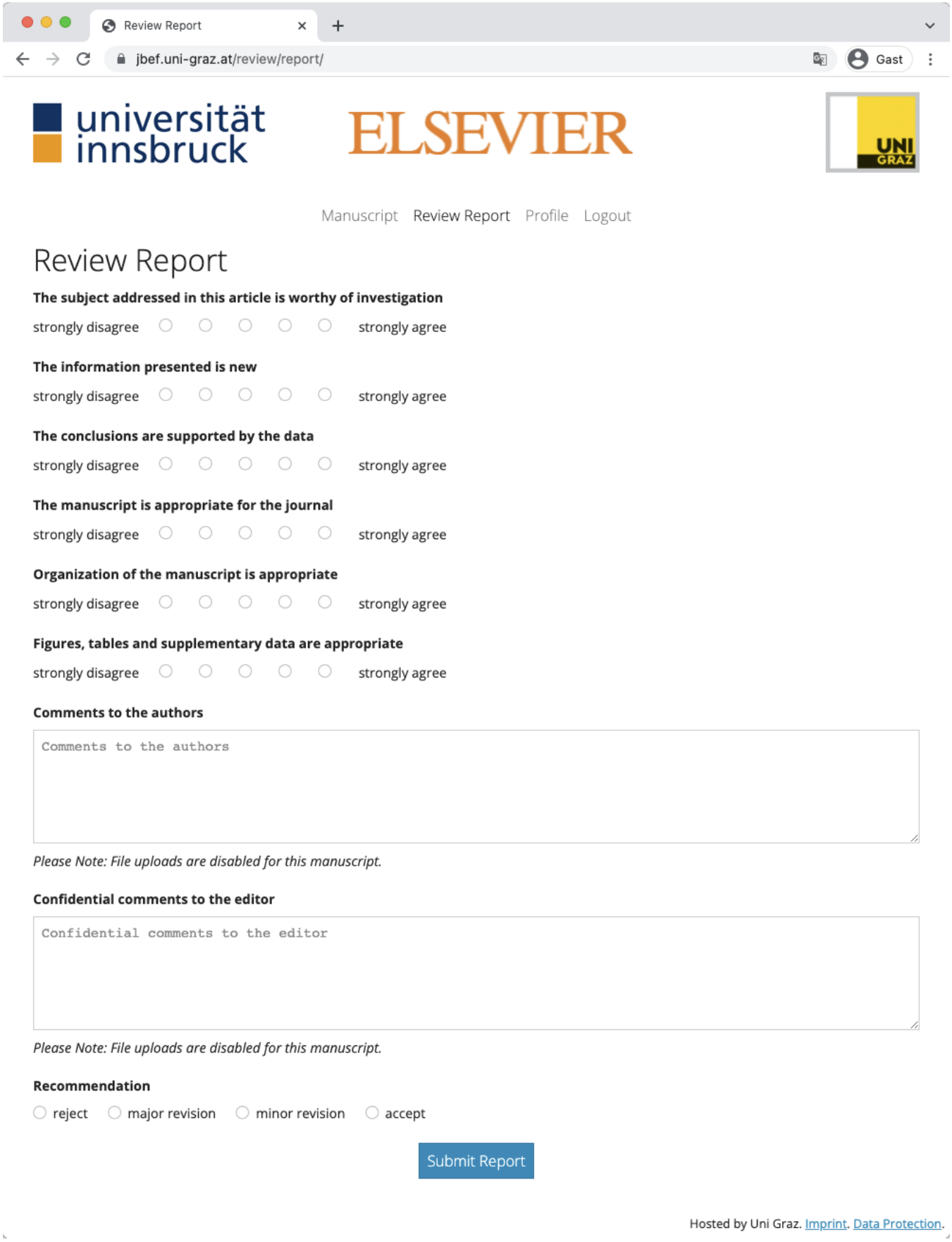


Fig. S3. Screenshot of report submission page



Supplementary Tables.

Table S1. Invitations

	Anonymized			
	AL	AA	AH	(AL, AA, AH)
Invitations sent	739	576	696	2011
Responses received	585	455	551	1591
Invitations accepted	163	161	165	489
Acceptance rate	27.86%	35.38%	29.95%	30.74%

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.

Table S2. Recommendations by condition.

	AL	AA	AH	LL	HH	Total
N	101	110	102	114	107	534
Accept %	1.98	1.82	20.59	3.51	14.95	8.43
Minor %	7.92	21.82	38.24	14.91	36.45	23.78
Major %	24.75	28.18	18.63	29.82	29.91	26.40
Reject %	65.35	48.18	22.55	51.75	18.69	41.39
AA vs. AL	z = 2.798			p = 0.0051		
AA vs. AH	z = -5.373			p < 0.0001		
AL vs. AH	z = -7.350			p < 0.0001		

L=author prominence low, A=anonymized, H=author prominence high, first letter is for the invitation letter, second letter is for the manuscript. We report pairwise, two-sided Mann-Whitney U tests.

Table S3. 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 received its first impact factor of 8.222 in June 2022.

Table S4. Randomization checks: all invited participants.

	Years active	h-index	i10-index	Citations	h-index 5y	i10-index 5y	Citations 5y
AL	0.017	0.191	0.031	0.002	0.275	0.128	0.191
	(0.108)	(0.134)	(0.119)	(0.097)	(0.145)	(0.136)	(0.134)
AH	-0.013	0.111	0.025	0.001	0.228	0.102	0.111
	(0.110)	(0.138)	(0.120)	(0.098)	(0.147)	(0.138)	(0.138)
LL	0.006	-0.011	-0.038	-0.000	-0.004	0.003	-0.011
	(0.107)	(0.138)	(0.120)	(0.094)	(0.150)	(0.139)	(0.138)
HH	0.001	0.073	0.017	-0.001	0.052	0.040	0.073
	(0.119)	(0.149)	(0.130)	(0.105)	(0.163)	(0.152)	(0.149)
N	3209	3284	3284	3284	3284	3284	3284
LR chi2(4)	0.340	9.111	5.284	3.395	11.918	8.662	9.111
Prob. > chi2	0.987	0.058	0.259	0.494	0.018	0.070	0.058

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

Table S5. Recoded reviewer characteristics.

Bin	Minimum	Maximum	Observations	Bin	Minimum	Maximum	Observations	Bin	Minimum	Maximum	Observations
Years active in publishing				Citations							
1	0	14	1687	1	1	21	191	46	1366	1432	25
2	15	29	1139	2	22	42	168	47	1433	1488	24
3	30	44	291	3	43	64	145	48	1494	1554	23
4	45	59	69	4	65	85	120	49	1561	1618	29
5	60	80	23	5	86	106	96	50	1619	1662	25
				6	107	127	106	51	1663	1721	21
h-index				7	128	148	82	52	1725	1796	22
1	1	18	2560	8	148	167	83	53	1803	1894	22
2	19	36	543	9	168	187	66	54	1896	1953	22
3	36	52	125	10	188	207	68	55	1956	2054	23
4	53	71	51	11	208	229	84	56	2057	2137	26
5	72	160	20	12	230	250	60	57	2139	2259	22
				13	251	270	52	58	2274	2384	22
I10-index				14	271	292	59	59	2388	2478	22
1	0	19	2293	15	293	314	53	60	2488	2597	21
2	20	39	536	16	315	337	46	61	2600	2737	20
3	39	57	208	17	338	362	45	62	2740	2898	22
4	58	76	100	18	363	384	38	63	2902	3080	21
5	78	100	53	19	385	407	39	64	3094	3244	20
6	101	122	32	20	408	431	44	65	3247	3419	22
7	125	173	32	21	432	457	42	66	3437	3599	21
8	174	262	23	22	458	480	38	67	3607	3776	21
9	270	1531	22	23	481	516	35	68	3794	4039	20
				24	517	544	43	69	4051	4347	20
h-index (5 years)				25	545	573	36	70	4355	4685	20
1	0	16	2682	26	575	601	34	71	4745	4971	20
2	17	33	485	27	602	627	26	72	4990	5363	22
3	34	50	89	28	628	651	30	73	5373	5888	20
4	51	122	28	29	652	681	30	74	5891	6134	20
				30	682	709	28	75	6147	6487	20
I10-index (5 years)				31	712	740	27	76	6496	7247	21
1	0	20	2580	32	747	789	32	77	7259	8078	20
2	21	41	433	33	790	817	32	78	8119	8810	20
3	42	62	132	34	821	848	32	79	8870	9763	21
4	62	84	52	35	849	876	30	80	9786	11256	20
5	86	112	39	36	877	913	31	81	11382	15356	20
6	113	182	26	37	914	953	33	82	15723	19324	20
7	194	1375	24	38	960	1001	28	83	19550	23776	20
				39	1005	1044	31	84	24206	35120	20
Citations (5 years)				40	1045	1097	28	85	37177	356624	21
1	1	18	2560	41	1098	1143	28				
2	19	36	543	42	1148	1206	30				
3	36	52	125	43	1208	1262	25				
4	53	71	51	44	1264	1324	25				
5	72	160	20	45	1325	1365	26				

To protect the anonymity of our reviewers, reviewer characteristics were recoded into categorical data. The columns Minimum and Maximum report the minimum and maximum values included in the respective category.

Table S6. Reviewer characteristics by treatments

	AL	AA	AH	LL	HH	All
Years active						
Mean	1.62	1.64	1.63	1.64	1.62	1.63
Std. dev.	0.747	0.821	0.813	0.81	0.777	0.794
Median [IQR]	1 [1 - 2]	1 [1 - 2]	1 [1 - 2]	1 [1 - 2]	1 [1 - 2]	1 [1 - 2]
Min., Max.	1, 5	1, 5	1, 5	1, 5	1, 5	1, 5
h-index						
Mean	1.36	1.29	1.31	1.28	1.3	1.31
Std. dev.	0.743	0.665	0.653	0.649	0.648	0.675
Median [IQR]	1 [1 - 2]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]
Min., Max.	1, 5	1, 5	1, 5	1, 5	1, 5	1, 5
i10-index						
Mean	1.72	1.65	1.67	1.57	1.65	1.65
Std. dev.	1.48	1.37	1.36	1.28	1.37	1.37
Median [IQR]	1 [1 - 2]	1 [1 - 2]	1 [1 - 2]	1 [1 - 2]	1 [1 - 2]	1 [1 - 2]
Min., Max.	1, 9	1, 9	1, 9	1, 9	1, 9	1, 9
Citations						
Mean	30.2	27.8	29.1	27.7	27.5	28.5
Std. dev.	26.1	24.3	25.5	24.4	24.6	25
Median [IQR]	21.5 [6.5 - 51]	21.5 [6 - 43]	20 [7 - 50]	19 [7 - 45]	20 [6 - 44]	20 [6 - 47]
Min., Max.	1, 85	1, 85	1, 85	1, 85	1, 85	1, 85
h-index (5 years)						
Mean	1.28	1.2	1.24	1.2	1.21	1.23
Std. dev.	0.603	0.505	0.512	0.511	0.493	0.53
Median [IQR]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]
Min., Max.	1, 4	1, 4	1, 4	1, 4	1, 4	1, 4
i10-index (5 years)						
Mean	1.46	1.36	1.4	1.34	1.38	1.39
Std. dev.	1.07	0.927	0.924	0.912	0.979	0.966
Median [IQR]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]
Min., Max.	1, 7	1, 7	1, 7	1, 7	1, 7	1, 7
Citations (5 years)						
Mean	1.36	1.29	1.31	1.28	1.3	1.31
Std. dev.	0.743	0.665	0.653	0.649	0.648	0.675
Median [IQR]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]	1 [1 - 1]
Min., Max.	1, 5	1, 5	1, 5	1, 5	1, 5	1, 5

The table shows means, standard deviations, medians and interquartile ranges (iqr) of our reviewer characteristics (categorized) by treatment. Example: The median category of the citation count variable across all treatments is 20, indicating that the median reviewer in our sample had between 408 and 431 citations on Google Scholar (see Table S5).

Table S7. Times reviewers took for responses, consent, and questionnaire submission

Treatment	N	Mean	Std. dev.	Median	Interquartile range	Kruskal-Wallis test
Hours to response						
AL	585	57.447	96.586	8.576	1.523 - 65.657	
AA	455	68.812	140.35	10.781	1.928 - 96.096	$chi^2(4) = 2.683$
AH	550	64.915	134.05	10.654	2.223 - 97.67	$p = 0.612$
LL	606	60.212	124.98	9.487	1.917 - 69.719	
HH	410	62.529	111.81	12.446	1.824 - 96.879	
Total	2606	62.45	122.13	9.771	1.845 - 79.056	
Hours to consent						
AL	141	55.862	85.326	12.282	1.953 - 62.849	
AA	147	61.511	94.036	14.19	3.545 - 85.586	$chi^2(4) = 1.335$
AH	139	66.995	116.47	18.926	2.395 - 83.401	$p = 0.856$
LL	150	64.143	103.57	14.914	3.037 - 78.165	
HH	142	62.345	84.403	13.866	2.809 - 121.96	
Total	719	62.177	97.28	14.262	2.644 - 79.094	
Minutes on consent page						
AL	140	7.851	57.149	0.558	0.217 - 1.833	
AA	146	3.904	13.624	1.008	0.367 - 2.15	$chi^2(4) = 8.688$
AH	134	1.683	3.355	0.925	0.283 - 1.75	$p = 0.069$
LL	148	10.922	111.51	0.867	0.367 - 1.983	
HH	139	29.505	204.78	0.917	0.45 - 1.883	
Total	707	10.767	107.54	0.867	0.317 - 1.933	
Hours from consent to report submission						
AL	101	596.48	313.75	667.74	437.73 - 740.7	
AA	110	617.1	280.8	673.35	515.54 - 719.21	$chi^2(4) = 0.699$
AH	102	624.27	288.95	668.94	473.13 - 756.68	$p = 0.951$
LL	114	601.3	286.2	667.23	481.19 - 726.86	
HH	107	626.12	319.15	665.11	434.72 - 750.78	
Total	534	613	296.95	669.46	461.32 - 732.35	
Minutes from report submission to questionnaire submission						
AL	86	2.921	5.261	2.017	1.25 - 2.5	
AA	95	1.856	1.015	1.7	1.15 - 2.25	$chi^2(4) = 46.642$
AH	93	294.84	2820.4	1.883	1.383 - 2.917	$p < 0.001$
LL	97	2.713	1.971	2.25	1.533 - 3.133	
HH	91	3.279	2.092	2.667	1.95 - 4.167	
Total	462	61.491	1265.4	2.083	1.4 - 2.95	

For each treatment, the table reports sample sizes (N), means, standard deviations, medians, and interquartile ranges for the time it took participants to reach various stages of the experiment. The last column reports the test statistics of Kruskal-Wallis tests across all treatment conditions in the respective section. The α-threshold after multiple hypothesis testing correction was 0.0125. As is evident from the means, there were some outliers. These were mainly caused by individuals that left the consent page open for very long times or returned to the questionnaire months after submitting their report.

Table S8. Robustness of main results to citation numbers.

	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 were estimated with Huber and White (robust) standard errors.

Table S9. Responses to the manuscript assessment questionnaire

Treatment	N	Mean	Std. dev.	Median	Interquartile range	Kruskal-Wallis test
The subject addressed in this article is worthy of investigation						
AL	99	3.061	1.114	3	2 - 4	
AA	109	3.165	1.364	4	2 - 4	
AH	99	4	1.195	4	4 - 5	$chi^2(4) = 66.777$
LL	111	3.423	1.133	4	2 - 4	$p < 0.001$
HH	106	4.094	0.971	4	4 - 5	
Total	524	3.546	1.234	4	3 - 5	
The information presented is new						
AL	99	2.455	1.081	2	2 - 3	
AA	109	2.78	1.189	3	2 - 4	
AH	99	3.545	1.248	4	3 - 4	$chi^2(4) = 71.061$
LL	111	2.748	1.124	3	2 - 4	$p < 0.001$
HH	106	3.566	1.078	4	3 - 4	
Total	524	3.015	1.226	3	2 - 4	
The conclusions are supported by the data						
AL	99	2.424	1.107	2	2 - 3	
AA	109	2.862	1.166	3	2 - 4	
AH	99	3.596	1.169	4	3 - 5	$chi^2(4) = 68.599$
LL	111	2.73	1.128	3	2 - 4	$p < 0.001$
HH	106	3.519	1.173	4	3 - 4	
Total	524	3.023	1.23	3	2 - 4	
The manuscript is appropriate for the journal						
AL	99	2.444	1.287	2	1 - 3	
AA	109	2.688	1.386	3	1 - 4	
AH	99	3.687	1.419	4	3 - 5	$chi^2(4) = 68.499$
LL	111	2.865	1.311	3	2 - 4	$p < 0.001$
HH	106	3.689	1.362	4	3 - 5	
Total	524	3.071	1.443	3	2 - 4	
Organization of the manuscript is appropriate						
AL	99	1.717	1.05	1	1 - 2	
AA	109	1.927	1.136	2	1 - 2	
AH	99	3.303	1.305	3	2 - 4	$chi^2(4) = 103.917$
LL	111	2.135	1.14	2	1 - 3	$p < 0.001$
HH	106	3.028	1.457	3	2 - 4	
Total	524	2.414	1.37	2	1 - 4	
Figures, tables and supplementary data are appropriate						
AL	99	2.414	1.161	2	1 - 3	
AA	109	2.881	1.078	3	2 - 4	
AH	99	3.636	1.173	4	3 - 5	$chi^2(4) = 75.114$
LL	111	2.829	1.103	3	2 - 4	$p < 0.001$
HH	106	3.604	1.185	4	3 - 5	
Total	524	3.071	1.228	3	2 - 4	
Average of all ratings						
AL	99	2.419	.8727	2.333	1.667 - 3	
AA	109	2.717	1.009	2.667	1.833 - 3.5	
AH	99	3.628	1.083	4	2.667 - 4.5	$chi^2(4) = 99.517$
LL	111	2.788	.9527	2.833	2 - 3.5	$p < 0.001$
HH	106	3.583	1.021	3.667	2.833 - 4.5	
Total	524	3.023	1.098	3	2 - 4	

For each treatment, the table reports sample sizes (N), means, standard deviations, medians, and interquartile ranges for reviewers' average ratings given in the six questions on the quality of the manuscript. In addition, it shows a combined score of all ratings. The last column reports the test statistics of Kruskal-Wallis tests across all treatment conditions. The α-threshold after multiple hypothesis testing correction was 0.05.

Table S10. Report word counts.

Treatment	N	Mean	Std. dev.	Median	Interquartile range	Kruskal-Wallis test
Comments to the authors						
AL	101	427.05	440.32	303	150 - 548	
AA	110	366.98	407.63	256.5	130 - 458	$chi^2(4) = 7.106$
AH	103	355.06	307.87	310	113 - 496	$p = 0.130$
LL	115	374.93	302.58	308	154 - 511	
HH	107	320.74	329.73	227	111 - 402	
Total	536	368.48	361.06	283.5	130 - 468.5	
Confidential comments to the editor						
AL	101	77.881	123.13	37	1 - 92	
AA	110	71.382	105.16	34.5	1 - 86	$chi^2(4) = 3.125$
AH	103	77.291	100.01	40	8 - 114	$p = 0.537$
LL	115	71.617	136.19	26	1 - 74	
HH	107	94.636	190.33	38	1 - 126	
Total	536	78.435	134.9	35.5	1 - 88	
Combined length of all comments						
AL	101	504.93	477.18	403	181 - 642	
AA	110	438.36	436	283	161 - 556	$chi^2(4) = 5.042$
AH	103	432.35	356.62	349	143 - 596	$p = 0.283$
LL	115	446.55	368.04	333	191 - 620	
HH	107	415.37	461.45	291	150 - 502	
Total	536	446.92	421.28	332.5	168 - 580.5	

For each treatment, the table reports sample sizes (N), means, standard deviations, medians, and interquartile ranges for the length of the reports (number of words) that reviewers submitted. The last column reports the test statistics of Kruskal-Wallis tests across all treatment conditions in the respective section. The α-threshold after multiple hypothesis testing correction was 0.0166.

Table S11. Recommendations regressed on manuscript assessment questions.

	Recommendation
Subject worthy of investigation	0.062
	(0.029)
Information new	0.164
	(0.033)
Conclusions supported by data	0.164
	(0.031)
Manuscript appropriate for journal	0.216
	(0.029)
Manuscript organization appropriate	0.331
	(0.028)
Figures and Supplement appropriate	0.041
	(0.030)
N	524
R^2	0.724

Ordinary least squares regression with standardized variables; Standard errors in parentheses. Because they are standardized, coefficient estimates are indicative of the relative strength of the effects on the dependent variable.

Table S12. 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

Table S13. Correlation of recommendation and quality assessments.

	Recommendation	Subject	Information	Conclusions	Manuscript	Organization	Supplement
Recommendation	1.000						
Subject	0.642	1.000					
Information	0.727	0.725	1.000				
Conclusions	0.719	0.612	0.688	1.000			
Manuscript	0.761	0.694	0.730	0.715	1.000		
Organization	0.770	0.568	0.672	0.666	0.714	1.000	
Supplement	0.676	0.558	0.615	0.691	0.678	0.724	1.000

Pairwise correlations.

Table S14. Loading of principal components on variables and explained variance.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Recommendation	0.393	-0.102	-0.402	-0.286	-0.117	0.335	-0.683
Subject worthy	0.354	0.672	0.136	0.431	-0.120	0.445	0.082
Information new	0.382	0.361	-0.134	-0.138	0.670	-0.485	-0.049
Conclusions supported	0.377	-0.121	0.610	-0.596	0.040	0.241	0.237
Manuscript appropriate	0.393	0.087	-0.079	-0.072	-0.700	-0.564	0.133
Organization appropriate	0.379	-0.392	-0.501	0.146	0.135	0.229	0.599
Supplement appropriate	0.366	-0.481	0.415	0.576	0.113	-0.163	-0.302
Explained variance	0.729	0.077	0.049	0.045	0.037	0.033	0.029

Columns PC1 to PC7 show factor loadings of the principal components. The last row shows the fraction of the total variance that is explained by each component.

References

1. Wikidata, Wikidata category q3918, universities (https://www.wikidata.org/wiki/Q3918) (2022).
2. Hipo, University domains and names data list & api (https://github.com/Hipo/university-domains-list) (2022).
3. J Cohen, A power primer. *Psychol. Bull.* **112**, 155–159 (1992).
4. D Peters, S Ceci, Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behav. Brain Sci.* **5**, 187–195 (1982).
5. 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).
6. L Fisher, SB Strauss, B Strauss, The effect of blinding on acceptance of research papers by peer review. *J. Am. Med. Assoc.* **272**, 143–146 (1994).
7. J Garfunkel, M Ulshen, H Hamrick, E Lawson, Effect of institutional prestige on reviewers' recommendations and editorial decisions. *J. Am. Med. Assoc.* **272**, 137–138 (1994).
8. S Madden, D DeWitt, Impact of double-blind reviewing on sigmod publication rates. *ACM SIGMOD Rec.* **35**, 29–32 (2006).
9. AKH Tung, Impact of double blind reviewing on sigmod publication: A more detail analysis. *SIGMOD Rec.* **35**, 6–7 (2006).
10. 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).
11. K Okike, K Hug, K Kocher, S Leopold, Single-blind vs double-blind peer review in the setting of author prestige. *J. Am. Med. Assoc.* **316**, 1315–1316 (2016).
12. A Tomkins, M Zhang, W Heavlin, Reviewer bias in single- versus double-blind peer review. *PNAS* **114**, 12708–12713 (2017).
13. D Card, S DellaVigna, What do editors maximize? Evidence from four economics journals. *Rev. Econ. Stat.* **102**, 195–217 (2020).
14. 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).
15. R Snodgrass, Single- versus double-blind reviewing: An analysis of the literature. *SIGMOD Rec.* **35**, 8–21 (2006).
16. J Polka, R Kiley, B Konforti, R Vale, Publish peer-reviews. *Nature* **560**, 545–547 (2018).
17. 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. *Nat. Commun.* **10**, 322 (2019).
18. F Squazzoni, et al., Unlock ways to share data on peer-review. *Nature* **578**, 512–514 (2020).
19. MA 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).