

Improving Peer Review in Economics: Stocktaking and Proposals*

Gary Charness, Anna Dreber, Daniel Evans, Adam Gill, Séverine Toussaert

April 2022

*Acknowledgments: We would like to thank all the associations and networks that helped us advertise the study to their members, including CEPR, CESifo and EEA. We also thank the editors at various journals for the statistics provided. Our appreciation goes to Peter Andre and Fulya Ersoy for the statistics and datasets shared with us. This report benefited from the wonderful research assistance of Raffaele Blasone, Ingrid Löfman, Jessica Milligan, Rafael Suchy, Alena Wabitsch, and Le Wu. The survey received ethics approval from the University of Oxford (Approval # ECONCIA21-21-20). Funding: Anna Dreber acknowledges funding support from the Jan Wallander and Tom Hedelius Foundation (Svenska Handelsbankens Forskningsstiftelser), the Knut and Alice Wallenberg Foundation, the Marianne and Marcus Wallenberg Foundation (in the form of a Wallenberg Scholar grant awarded to AD), and the Austrian Science Fund (FWF, SFB F63).

Contents

1	Introduction	1
2	Survey design and sample	2
2.1	Survey content	2
2.2	Our population of respondents	2
2.2.1	Recruitment	2
2.2.2	Characteristics of the sample	3
3	Results	5
3.1	The ecosystem of peer review	5
3.1.1	The demand side of peer review: authors	6
3.1.2	The supply side of peer review: referees	7
3.2	Allocating papers to journals, editors, and referees	9
3.2.1	Mismatch on the quantity dimension	9
3.2.2	Mismatch on the quality dimension	12
3.2.3	The concentration of editorial and refereeing responsibilities	15
3.2.4	Solving the allocation problem	18
3.3	Content of the reports	25
3.3.1	Perceived quality of referee reports	25
3.3.2	When are referee reports most useful (if at all)?	27
3.3.3	Improving the quality of reports	27
3.4	Reviewing process and decision times	33
3.4.1	Delays in peer review	33
3.4.2	Reducing delays	38
3.5	Innovations in peer review	45
3.5.1	The key dimensions of peer review	45
3.5.2	Transparency in peer review	46
3.5.3	Other peer review models	48
4	Taking a step back... but also moving forward	52
4.1	Taking stock of where we are	53
4.1.1	Challenges facing our peer review system	53
4.1.2	Potential solutions and respondent attitudes	53
4.1.3	Gaps in our knowledge	53
4.1.4	Changing both how we produce and evaluate research?	54
4.2	Open questions and proposals	55
Appendix A	Additional tables and data discussions	63
A.1	Full proposals table	63
A.2	Andre and Falk data and discussion (Tables 3 and A.2)	65
A.3	Top 5 table sourcing discussion	67
A.4	Journal population table sourcing discussion	68
A.5	Journal sourcing table	70
Appendix B	Description of variables and statistics	73
B.1	Variable codebook	73
B.2	Descriptive statistics and figures	76
Appendix C	Recruitment	79
Appendix D	Survey questions	83
Appendix E	Additional figures	90

List of Figures

1	The ecosystem of peer review	5
2	Submissions to an economics peer-reviewed journal	6
3	Rank distributions of peer review outcomes	7
4	The peer review efforts of referees	8
5	Time dedicated to peer review	8
6	Evaluation of the role of referee	9
7	Actual vs. reasonable number of referee reports per year	10
8	Referee reports vs. submissions per year (economics journals only)	11
9	Rejections of peer review requests	13
10	Reasons for rejecting or being tempted to reject a request	13
11	Attitudes toward co-authors and friends peer reviewing	14
12	Demographics of editor population	17
13	Views on the use of a centralized platform	20
14	Support for allowing authors to disqualify certain reviewers	23
15	Support for double-blind peer review	24
16	Perceived quality of reports received	25
17	Characteristics of low-quality reports	26
18	Most useful comments given by reviewers	28
19	Views on the provision of guidelines and templates for writing reports	29
20	Views on doctoral training	30
21	Perceived usefulness of grading and rewarding high-quality reports	31
22	Importance of sharing editorial decisions and reports with reviewers	32
23	Attitudes towards a formal appeal procedure	33
24	Delays in returning peer reviews	35
25	Views on the appropriate time length to submit a referee report	35
26	Percentage of late reviews as a function of the number of reports written	36
27	Views on desk rejections	39
28	Finite automaton procedure for deciding on a peer review request (Courtesy of Shengwu Li)	40
29	Incentivizing and rewarding refereeing	41
30	Support for the AER: Insights model	43
31	Dimensions of peer review	45
32	Attitudes towards open peer review (publication + identifiability)	47
33	Views on publishing the peer review history (publication without identifiability)	48
B.1	Age of respondents	76
B.2	Location of respondents	77
B.3	Total number of papers published by respondents	77
B.4	Distribution of research fields of respondents	78
B.5	Types of journals that respondents referee for	78
C.1	Survey response time	81
C.2	Distribution of responses across recruitment channels	81
E.1	Regression of editorial experience on respondent characteristics	90
E.2	Regression of top 5 journal refereeing experience on respondent characteristics	90
E.3	Regression of reports written and submissions made on respondent characteristics	91
E.4	Regressions of rank assigned to stated peer review outcomes on respondent characteristics	91
E.5	Regression of rank assigned to stated benefits of refereeing on respondent characteristics	92
E.6	Regressions of misallocation variables on respondent characteristics	93
E.7	Actual vs. reasonable number of referee reports per year (by position)	94
E.8	Referee reports vs. submissions per year (by position)	95
E.9	Regressions of reasons for (considering) rejecting requests on respondent characteristics	96
E.10	Regression of attitudes towards friends/co-authors refereeing on respondent characteristics	96
E.11	Heterogeneity in rejection rates	97
E.12	Regression of perceived quality of reports received on respondent characteristics	97

E.13	Regressions of characteristics of low-quality reports on respondent characteristics	98
E.14	Regression of rank assigned to proposals for improving refereeing on respondent characteristics	98
E.15	Heterogeneity in percentage of late reports	99
E.16	Regression of reasonable time for reports (in weeks) on respondent characteristics	99
E.17	Average delay and rejection rate of review requests	100
E.18	Regression of referee incentives on respondent characteristics	100
E.19	Perceived usefulness of removing reviewers' anonymity	101
E.20	Correlations between the perceived usefulness (1-5) of various proposals	101
E.21	Word cloud of top 100 most common words	102

List of Tables

1	Summary of important acronyms and abbreviations used in the report	iv
2	Summary of survey content	2
3	Comparison of survey demographics	4
4	Comparison of top 5 journal policies and statistics	16
5	Comparison of incentive treatments	44
A.1	Proposals summary	63
A.2	Comparison of study population to top 5 referees and editors	66
A.3	Journal population	69
A.4	Sources for journal statistics and editorial policies	71
B.1	Descriptive statistics for the main dimensions of heterogeneity	76
C.1	Recruitment strategy summary	80
C.2	Characteristics across recruitment channels ^a	82

Table 1: Summary of important acronyms and abbreviations used in the report

Acronym/ Abbreviation	Meaning
<i>AEA</i>	American Economic Association
<i>AEJ</i>	American Economic Journal
<i>AER</i>	American Economic Review
<i>AERI</i>	American Economic Review: Insights
<i>AF</i>	Andre and Falk (2021)
<i>BMJ</i>	British Medical Journal
<i>CEPR</i>	Centre for Economic Policy Research
<i>COPE</i>	Committee on Publication Ethics
<i>DR</i>	Desk rejection
<i>DOI</i>	Digital Object Identifier
<i>ECMA</i>	Econometrica
<i>EJ</i>	Economic Journal
<i>ESA</i>	Economic Science Association
<i>JDE</i>	Journal of Development Economics
<i>JEEA</i>	Journal of the European Economic Association
<i>JF</i>	Journal of Finance
<i>JFE</i>	Journal of Financial Economics
<i>JPE</i>	Journal of Political Economy
<i>NHB</i>	Nature Human Behaviour
<i>NBER</i>	National Bureau of Economic Research
<i>OSF</i>	Open Science Foundation
<i>ORCID</i>	Open Researcher and Contributor ID
<i>PAP</i>	Pre-analysis plan
<i>PNAS</i>	Proceedings of the National Academic of Sciences
<i>QE</i>	Quantitative Economics
<i>QJE</i>	Quarterly Journal of Economics
<i>REStud</i>	Review of Economic Studies
<i>REStat</i>	Review of Economics and Statistics
<i>RFS</i>	Review of Financial Studies
<i>R&R</i>	Revise and resubmit
<i>TE</i>	Theoretical Economics

1 Introduction

As the gatekeeper of scientific publishing, peer review occupies a central place in the life of a researcher. Despite its important role, it remains a rather slow-moving and opaque institution. While economics has undergone a dramatic transformation in the past century in terms of its scope and methodology, its approach to peer review has remained relatively unchanged. Moreover, much remains unknown about peer review because its inner workings are rarely discussed in the open, partly due to the expectation of confidentiality. This lack of transparency manifests itself in the uneven quality of data provided by journals: while some offer annual reports summarizing submission, review, and editorial activity, others provide almost no information about their operations.¹ Even when available, statistics are often not directly comparable between journals and say nothing about the subjective experiences of authors and reviewers who both support and depend on the system. This scarcity of suitable data complicates the task of evaluating how the peer review system in economics has performed, both over time and in comparison to other disciplines.

To bridge this gap, we surveyed over 1,400 researchers in economics from a variety of backgrounds, fields, and locations between July 2020 and January 2021. The survey is the bedrock of this report, which was written to (i) document the current state of peer review and (ii) investigate concrete steps towards improving it. Making use of the survey's findings, we offer a snapshot of the recent submission and peer review activity of respondents, detail the difficulties they report facing, and measure their attitudes about various issues and proposals to address them. Further information about the report and the survey is available on our website (<https://evalresearch.weebly.com/>). To complement the survey, we also reviewed the existing literature on peer review and researched current journal policies both within and outside of economics. This literature is presented in a thematically-organized bibliography on our website (<https://evalresearch.weebly.com/bibliography.html>). Finally, in search of perspectives and innovative solutions, we engaged in dialogue with various stakeholders to better understand existing constraints and initiatives to improve peer review.

Why is now a good time for this report? The COVID-19 pandemic highlighted the need for (i) a more efficient and dynamic process that can quickly disseminate time-sensitive research and (ii) a system that distributes the burden of peer review more fairly, as its inequity has been exposed and exacerbated by the shift to remote work. Beyond COVID-19, trends in research have moved towards greater transparency and the embrace of open science practices (e.g., data and code sharing and pre-analysis plans), but peer review has shown few signs of rethinking its protocols. Now is the right time to ask defining questions about the role that we want peer review to play in our discipline.

Much like the peer review process can be, this report is certainly too long and does not entirely fulfill its mission. We tried to give a comprehensive overview of the challenges facing our system, but this often came at the expense of concision. In recognition of this, we make abridged summaries of each section available on our website (<https://evalresearch.weebly.com/report.html>). At the same time, our presentation remains incomplete and does not capture all of the nuances and apparent paradoxes of the system, which is characterized by competing objectives and challenging trade-offs. With that in mind, this report is conceived as a living document that will be responsive to new evidence and perspectives as they become available.

Finally, we note that we are not the first ones to comment on the state of peer review or ask researchers about their attitudes towards it. One recent effort from [Tennant et al. \(2017\)](#) offers a fairly exhaustive account of the history and trajectory of peer review from an interdisciplinary perspective. Moreover, a number of surveys on peer review have been conducted with differing scopes and intentions. Within economics, an early effort by [Mason et al. \(1992\)](#) surveyed economists on their experiences with and attitudes towards delays in peer review. An interdisciplinary survey from 2009 asked about the functionality of the current system and proposals to change it, among other things ([Mulligan et al., 2013](#)). [Silbiger and Stubler \(2019\)](#) focused on authors' experiences with unprofessional referee reports. Our survey is both an update on findings from previous surveys and the most comprehensive effort within economics that we are aware of. We hope that this report will provide fertile ground for the development and implementation of practical solutions for improving peer review in economics.

The report is structured as follows. Section 2 introduces the survey design and describes the characteristics of respondents. Section 3 presents our survey results together with some background on various aspects of peer review. Section 4 takes stock of our main findings and presents some questions for further discussion. Additional discussions, tables, and figures are presented in the appendix.

¹See Section A.3 for a summary of the state of data availability at the top 5 journals.

2 Survey design and sample

2.1 Survey content

We conducted an anonymous online survey to gather the views of researchers about the current peer review process and how it could be improved.² Researchers were eligible to participate if, in the preceding two years before taking the survey, they had (i) completed at least one peer review report and (ii) received referee reports on a submission made to a peer-reviewed journal.³ Respondents were asked to provide consent prior to taking part. We estimated that the survey would take 15-20 minutes, and the median response time was 15.5 minutes (see Figure C.1 for the distribution of response times).

The survey was divided into three sections. The first two sections (shown in a randomized⁴ order) asked about peer review from the perspective of (i) the author (demand side) and (ii) the referee (supply side). The author section asked respondents about the general aims of the peer review system, the quality and characteristics of the reports they have received, and various proposals to change the system. The referee section asked about the respondent's current peer review activity (e.g., reviews completed, declined requests, delays in response), the perceived benefits of being a referee, and proposals to improve the peer review experience. The final section collected data on respondents' socio-demographic and professional characteristics. The content of the various sections is summarized in Table 2. The questions took a variety of forms, including drop-down questions, 5-point scales, ranking questions (with the order randomized), multiple selection questions, and fill-in questions. The survey was identical across all respondents except for a few minor changes and the addition of two questions in later survey rounds, the responses to which are only available for 46% of our sample ($N = 684$).⁵ The full questionnaire can be found in Appendix D. After we completed our primary analysis of the survey data, we sent out a short follow-up survey in February 2022 ($N = 117$) to clarify our interpretation of the initial results and gather additional evidence.

Table 2: Summary of survey content

Part	Content
Experience as an author	<ul style="list-style-type: none">• General author characteristics• Evaluation of the current situation• Views about the aims of peer review• Evaluation of various proposals
Experience as a referee	<ul style="list-style-type: none">• General referee characteristics• Current referee load• Delays and requests declined• Benefits of and potential improvements to being a referee
Characteristics of the respondent	<ul style="list-style-type: none">• Field, position, number of publications• Age, gender, location

2.2 Our population of respondents

2.2.1 Recruitment

Data collection took place over a period of seven months (July 2020 – January 2021), with a first wave targeting behavioral and experimental economists and a second wave targeting all fields. The recruitment

²Respondents could leave their email address to participate in a lottery draw and/or receive the report; to maintain confidentiality, emails were collected in a separate link from the survey.

³In some announcements, we further specified “economics journal” instead of simply “peer-reviewed journal.” While the survey itself was ambiguous about this field requirement, it was met by virtually all respondents (only 7 respondents indicated that they had not submitted any paper to an economics journal over the previous two years).

⁴Ultimately, we found no evidence of an order effect.

⁵Before the main collection stage, we sent a pilot survey with more substantive differences to a handful of respondents. To include all feedback received, we use their responses in the analysis for the questions matching those of the main survey.

strategy took a variety of forms in order to maximize the number and diversity of responses.⁶ In Wave 1, most responses came from a mail merge sent to 1,802 contacts, for which we received 655 responses (36.3%). In Wave 2, a mail-merge sent to 3,618 contacts received 269 responses (7.4%). In addition, the survey was advertised to members of various associations, networks, and forums, including the Centre for Economic Policy Research (CEPR), the CESifo network, and the discussion forum of the Economic Science Association (ESA). Personal emails and posts to various social media accounts occurred sporadically across both recruitment waves. A more detailed explanation of the recruitment process, recruitment channels, and response numbers can be found in Appendix C. Across all recruitment methods, we received responses from 1,497 respondents, of which 1,459 (over 97%) were fully complete. To incorporate all the feedback we received, the analysis makes use of responses from partially-completed surveys whenever available.

2.2.2 Characteristics of the sample

To the best of our knowledge, general statistics on the world population of economists are nonexistent. As an informative benchmark, we compare the characteristics of our respondents to the study population, weighted survey sample, and unweighted survey sample of a recent paper, Andre and Falk (2021) (henceforth AF). This study surveyed a large sample of economists on their views about the type of economics research that should be conducted (see 4.1.4 for further discussion). Respondents were drawn from a database of scholars created by compiling publication data from EconLit for the years 2009 to 2019.⁷ The dataset was further restricted to scholars who recently contributed to the field (defined as having (i) at least one publication since 2015 and (ii) at least three articles in their publication sample or at least 50% of their publications in economics journals) and who had a valid email online. In total, the AF study population consists of 53,779 authors. This is the largest and most comprehensive database of economists in academia that we could find and a reasonably good proxy for the overall population of active researchers. The AF survey sample consists of the researchers from the study population who received and completed their survey (7,794 authors or 14.5% of the study population). Statistics from the unweighted sample summarize the raw characteristics of respondents, while the weighted sample adjusts the sample distribution to match the study population characteristics. We make comparisons with these survey samples because they contain additional characteristics, such as age and position, that were not available for the full study population. Furthermore, the self-reported data from the AF survey is more directly comparable with ours.

A summary of these comparisons can be found in Table 3. Most data was taken directly from Andre and Falk (2021). Additional statistics were derived for us by Peter Andre. Comparing the columns, we have good representation on the gender dimension, with female researchers comprising 23.5% of our sample. Our sample is also fairly representative on the age and position dimensions, although we do have an oversampling of younger researchers (42.8%) and assistant professors (26.5%), and an undersampling of older researchers (8.9%) and associate professors (22.0%) (see Appendix Figure B.1 for an age breakdown).⁸ Our survey population deviates more from the AF samples on the location dimension, with an overrepresentation of researchers from Europe (54.2%) and an underrepresentation from Asia and Oceania (8.4%) (see Appendix Figure B.2 for the full map). The researchers in our sample also tend to be more prolific authors, with an average of 25.3 publications (see Appendix Figure B.3). Our survey sample deviates the most along the field dimension (see Appendix Figure B.4 for the full distribution of fields). While we could not compare our proportion of behavioral/experimental economists to that of the AF samples, we used information on the field(s) of CEPR researchers to obtain a benchmark. The proportion of researchers who selected behavioral/experimental economics as a field is about 5 times larger in our sample than in the CEPR group (55.0% compared to 11.5% for CEPR researchers).⁹ This is likely attributable to the familiarity of researchers

⁶Different survey links were generated each time, allowing us to track response numbers by recruitment channel (see Figure C.2). A downside of this approach is that researchers could not be prevented from taking the survey multiple times. However, the evidence we have from respondents who left their email address suggests this concern is likely to be very minor.

⁷Articles published between 2009 and 2019 in the 400 indexed journals with the highest impact factor according to the Scopus 2018 Scimago Journal Ranking in “Economics, Econometrics, and Finance.” This is referred to as the “publication sample.”

⁸Our follow-up survey ($N = 116$ after removing one respondent who selected “Prefer not to say”) sampled relatively fewer full (33.6%) and assistant (19.0%) professors, and relatively more postdoc/PhD candidates (19.0%).

⁹The CEPR figure corresponds to the percentage of CEPR researchers affiliated with the Organizational Economics (OE) programme, which encompasses behavioral, experimental, and some other fields of economics; researchers can belong to several areas. Field statistics were kindly provided to us by the CEPR. CEPR researchers are unlikely to be representative of the entire population of economists, although the CEPR programme areas reflect the diversity of fields in the profession.

in this field with the authors of this report. We also have an undersampling of researchers in some other fields, most notably macroeconomics (13.1%). Finally, it is reasonable to suspect selection on unobservables e.g., researchers who are more likely to respond to surveys and/or are invested in improving peer review might feel differently about the various proposals and be more active and timely reviewers. In consideration of these factors, we do not claim a representative sample of academic economists and caution should be used when extrapolating our survey findings to the entire profession.

Table 3: Comparison of survey demographics

Demographics	Andre & Falk (2021)			
	Peer review survey	Study population	Unweighted sample	Weighted sample
Female	23.5%	26.0%	23.1%	25.8%
<i>Age:</i>				
Under 40	42.8%	-	32.5%	34.8%
40-49	32.4%	-	32.6%	31.6%
50-59	15.9%	-	18.9%	17.7%
60 and over	8.9%	-	16.0%	15.9%
<i>Location:</i>				
US/Canada	35.5%	33.9%	24.2%	33.9%
Europe	54.2%	40.4%	53.6%	40.5%
Asia/Oceania	8.4%	21.4%	17.1%	20.5%
Other regions	1.9%	4.3%	5.1%	5.1%
Field of Research (excl. Behav/Exp)^a				
Microeconomics	27.5%	14.5%	18.2%	15.2%
Macroeconomics	13.1%	31.1%	24.3%	29.4%
Econometrics	8.8%	4.2%	3.2%	3.2%
Development	7.9%	7.3%	7.5%	8.0%
Labor	10.9%	8.6%	12.2%	9.8%
Industrial Organization	6.0%	8.3%	7.4%	8.0%
Public Economics	11.1%	3.6%	4.3%	3.8%
Other fields	14.7%	22.4%	22.9%	22.6%
Position				
Full Professor	38.3%	-	41.1%	37.1%
Associate Professor	22.0%	-	27.3%	28.2%
Assistant Professor	26.5%	-	19.6%	22.0%
Postdoc/PhD Candidate	7.6%	-	9.6%	10%
Other Position	5.6%	-	2.4%	2.6%
Professional Experience				
Average number of publications	25.3	17.1	18.3	16.2
N	1,459 ^b	53,779	7,794	7,794

Notes: The statistics in the last three columns were either directly taken from Andre and Falk (2021) or derived for us by Peter Andre. For a description of the construction of these variables, see Appendix A.2.

^a For the field statistics, we removed the “Behavioral/experimental” selection from our peer review data to improve comparability with the Andre & Falk data for the other fields. As such, it is useful to remember that we report a conditional distribution. For example, the microeconomics category, consisting of respondents who selected either applied microeconomics, decision theory, game theory, or microeconomic theory as a field, accounts for 27.5% of all field selections that were not behavioral or experimental economics ($N = 2,668$ remaining selections). Behavioral/experimental economics accounts for 33.0% of the total number of field selections (out of $N = 3,982$ selections made across all fields).

^b The sample size reported for our peer review survey is the number of completed surveys. However, for our statistics to be comparable with the data in AF, we needed to remove the respondents who selected “Prefer not to say” for Age, Location, and Position. This leads to smaller sample sizes for those variables (with $N = 1,381$, 1,392, & 1,401, respectively). For the Female statistic, responses of “Prefer not to say” were kept.

3 Results

This report follows a roughly chronological order in documenting the stages of the peer review process. Section 3.2 looks at the initial allocation of papers to journals and referees. Section 3.3 turns to the production of reports, with a focus on their content. Section 3.4 studies delays in the delivery of reports and the communication of decisions. Finally, Section 3.5 discusses potential innovations in peer review, including those geared towards increasing transparency in the peer review process. Throughout, we discuss potential problems and solutions, including their limitations. But first, Section 3.1 takes a look at the ecosystem of peer review, with a focus on the main characteristics of the demand-side (authors) and the supply-side (referees), as well as a discussion of the objective function (visualized in Figure 1).

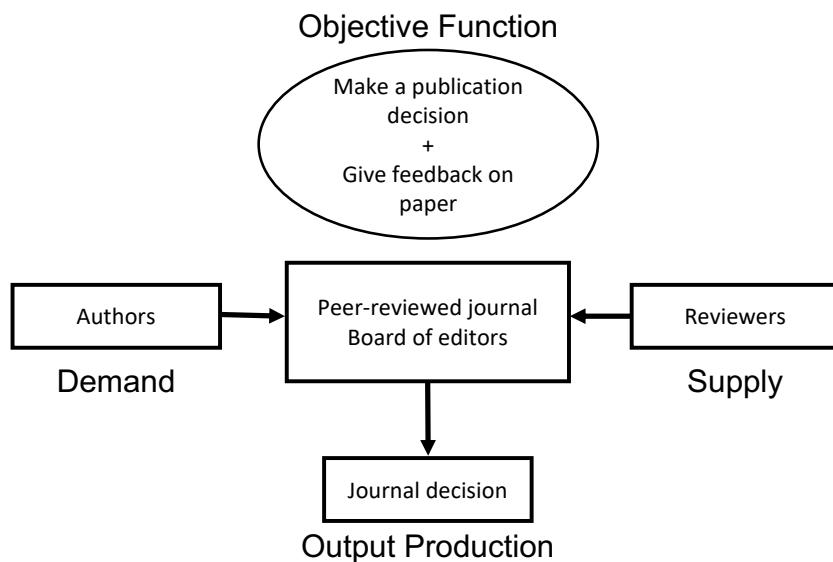
We chose simplicity in the presentation of our results, with an emphasis on descriptive statistics. Several figures present statistics broken down by the following dimensions of respondent heterogeneity (all dichotomous variables): `Behav/Exp` (=1 if in field of behavioral/experimental economics), `Female` (=1 if researcher indicated they were female), `Junior` (=1 if non-tenured), `US/Canada` (=1 if job location in the US or Canada), `Editor` (=1 if they have current or previous editorial experience), `Top 5 reviewer` (=1 if refereed at least once for a top 5 journal over the preceding two years), `Active reviewer` (=1 if wrote an above-median number of reports), and `Active author` (=1 if made an above-median number of submissions). The sample sizes for these variables can be found in Appendix Table B.1. To complement the descriptive statistics, we also make available a (non-exhaustive) collection of figures plotting results from linear regressions; these figures can be found in Appendix E and are referenced in the main text. Further regressions can be run at the reader’s discretion with the dataset available on our OSF page (<https://osf.io/eczkv/>).

The sample sizes for our figures and regressions are based on the number of respondents who answered the relevant questions; this explains why some analyses differ in sample size. Unless otherwise specified, the sample sizes can be calculated as the number of partially completed surveys minus the number of “no response” ($N = 1497 - NAs$). When a figure or analysis utilized the responses of multiple questions, respondents who did not answer all relevant questions were removed.

3.1 The ecosystem of peer review

Peer review is central to the lives of researchers. They are active on both sides of the market, although not necessarily equally so, as we will see. In this section, we document the general activity of authors and reviewers, as well as their expectations about what the peer review process should achieve.

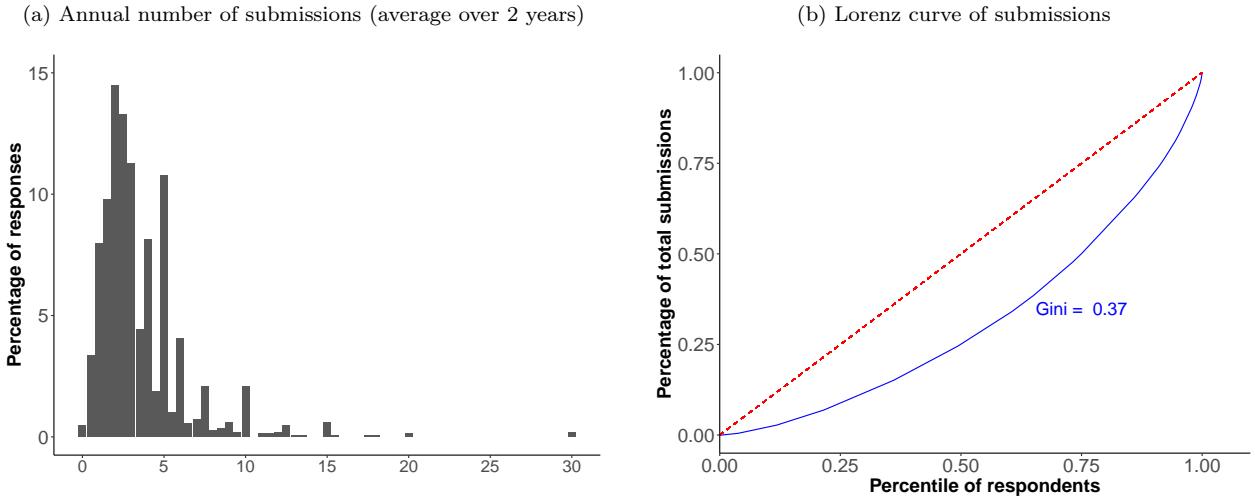
Figure 1: The ecosystem of peer review



3.1.1 The demand side of peer review: authors

In the two years prior to their taking of the survey, our 1,484 respondents made a total of 10,468 new submissions to economics journals. This corresponds to an individual average of 3.5 submissions per year and a median of 3 submissions (where each submission of the same paper to a different journal counts as a separate submission). As shown in Panel (a) of Figure 2, there is some heterogeneity, with a standard deviation of 2.8 and a long tail of very active researchers. The concentration in papers submitted is fairly high, with 50% of submissions being made by only 20% of researchers (Panel (b) of Figure 2). The number of submissions made is significantly higher on average among behavioral/experimental economists ($p < 0.01$) and researchers with more publications ($p < 0.01$), and significantly lower among researchers who referee for top 5 journals ($p < 0.01$) (see Appendix Figure E.3).

Figure 2: Submissions to an economics peer-reviewed journal

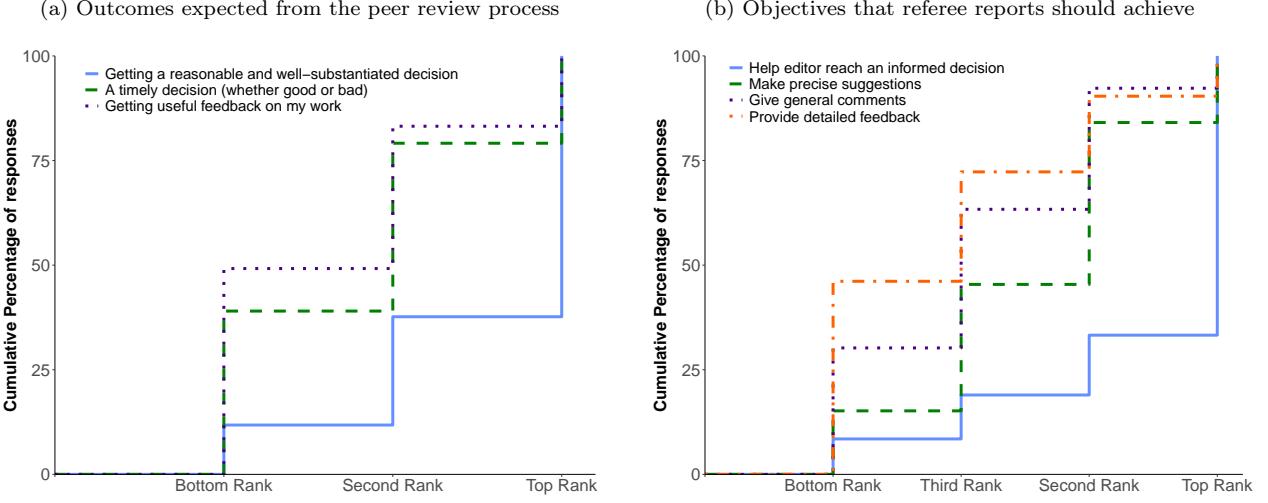


Notes: Both panels are based on responses to question Q1. Panel (b) shows the Lorenz curve of the number of submissions made by our respondents. On the x-axis, respondent percentiles are based on the number of manuscripts they submit. The y-axis gives the cumulative proportion of submissions made by respondents at or below each percentile. For instance, respondents above the 80th percentile are responsible for approximately 50% of all submissions made (blue line). The Gini coefficient given below the line measures the extent of inequality in the distribution (where a Gini of 0 indicates perfect equality - 45° line). N = 1,484 for both panels.

As shown in Panel (a) of Figure 3, what authors expect the most from the peer review process is “*getting a reasonable and well-substantiated decision*.” At the same time, they also care about gathering useful feedback and obtaining a timely decision, with 1 in 5 respondents ranking “*a timely decision (whether good or bad)*” as the primary objective and 1 in 6 prioritizing “*getting useful feedback on my work*.” We note that these objectives might be in tension to some extent, although we will discuss how they could be jointly satisfied. There is also heterogeneity in these selections. For instance, top 5 referees, active reviewers, and active authors give higher priority on average to getting a reasonable response and lower priority to getting useful feedback; meanwhile, behavioral/experimental economists and female researchers put more weight on getting useful feedback and less weight on receiving a timely decision (see Panel (a) of Appendix Figure E.4).

There is also general agreement about what referee reports should achieve (Panel (b) of Figure 3). For about 70% of researchers, the primary objective of referee reports is to “*help the editor reach an informed decision*.” As for the second objective, respondents largely believe that referee reports should “*make precise suggestions*” rather than “*give general comments*” or “*provide detailed feedback*.” Researchers based in the US/Canada and top 5 reviewers on average tend to put relatively more weight on helping the editor decide and relatively less weight on providing detailed feedback, while the opposite pattern holds for those with more total publications (see Panel (b) of Appendix Figure E.4). Later on, we will attempt to evaluate the current situation and analyze various proposals in light of these objectives.

Figure 3: Rank distributions of peer review outcomes



Notes: Panel (a) uses responses to question Q5; Panel (b) uses responses to question Q4. N = 1,476 for both panels.

3.1.2 The supply side of peer review: referees

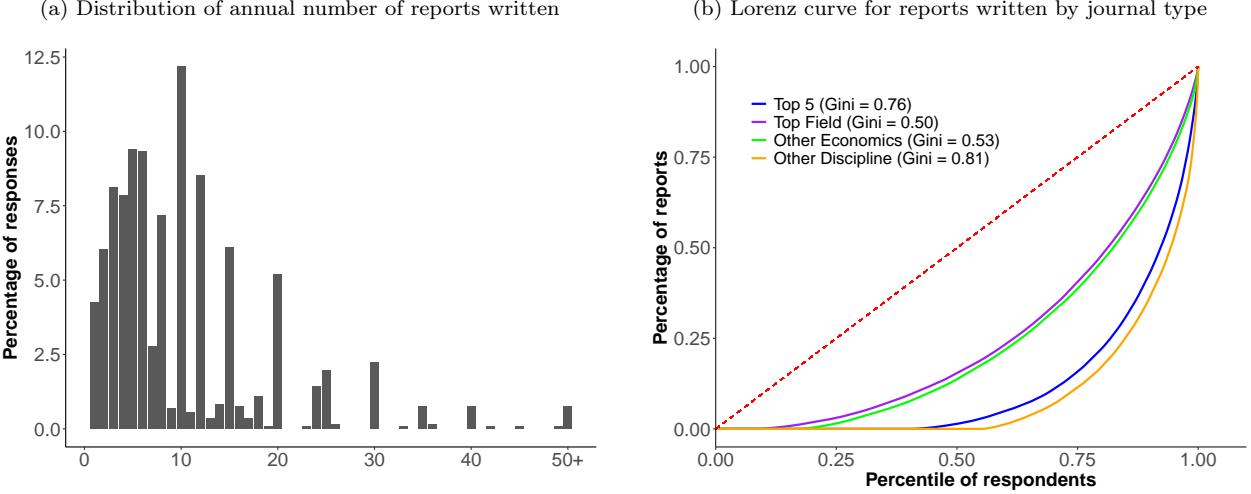
Turning to the supply side of peer review, our data makes clear that writing referee reports constitutes an important professional activity. Our 1,483 respondents wrote a total of 15,060 referee reports per year, corresponding to an annual average of 10.2 reports per reviewer and a median of 8 reports. As was the case for submissions, Panel (a) of Figure 4 shows there is substantial heterogeneity: the distribution of reports written has a standard deviation of 9.1 and exhibits a long right tail, with the top 10% most active reviewers writing at least 20 reports a year. In terms of correlates, we note that the most active reviewers in our sample tend to be more senior researchers, researchers working in the US/Canada, and researchers with editorial experience (see Appendix Figure E.3). Editorial experience is somewhat common in our sample, with over 38% of our respondents reporting that they currently occupy or have previously occupied an editorial position.¹⁰ Unsurprisingly, these respondents also tend to hold more senior job titles (Appendix Figure E.1).

There is also substantial heterogeneity in terms of what journals respondents review for. On average, 90% of reports are written for economics journals, of which 16% are top 5 journals, 41% are top field journals, and 33% are other economics journals. Moreover, about half of respondents write none of their reports for a top 5 journal. Indeed, concentration in peer review efforts is particularly high for top 5 journals, with 80% of submissions to a top 5 journal being reviewed by only 25% of respondents (Panel (b) of Figure 4). Respondents who do review for top 5 journals tend to be full professors, be based in the US/Canada, and have held or currently hold an editorial position. They also tend to make fewer submissions but write more reports (see Appendix Figure E.2).

It is worth noting that, besides writing a fair number of reports, our respondents also devote a non-trivial amount of time per report. About three-quarters of respondents spend at least one day per report, with close to 40% spending two days or more (Panel (a) of Figure 5). Unsurprisingly, early-career researchers tend to spend more time per report, with 51% of PhD students, postdocs, and assistant professors spending two days or more vs. 34% of associate or full professors (p-val < 0.01). This might suggest that experience generates efficiency gains. However, we also note that the Spearman correlation between the number of reports written and the time spent per report is not very strong ($\rho = -0.34$, p-val < 0.01). While very active reviewers are more likely to write “express” reviews (2 hours of work or less), they also represent a non-negligible fraction of those who spend one day or more on a referee report (Panel (b) of Figure 5). As a result, many respondents end up dedicating a considerable amount of their time to peer review. Combining the statistics from Figure 5 and Figure 4(a) and setting “more than 2 days” to 3 days, we can estimate the

¹⁰In our follow-up survey ($N = 117$), 33.3% of respondents reported that they are currently editors.

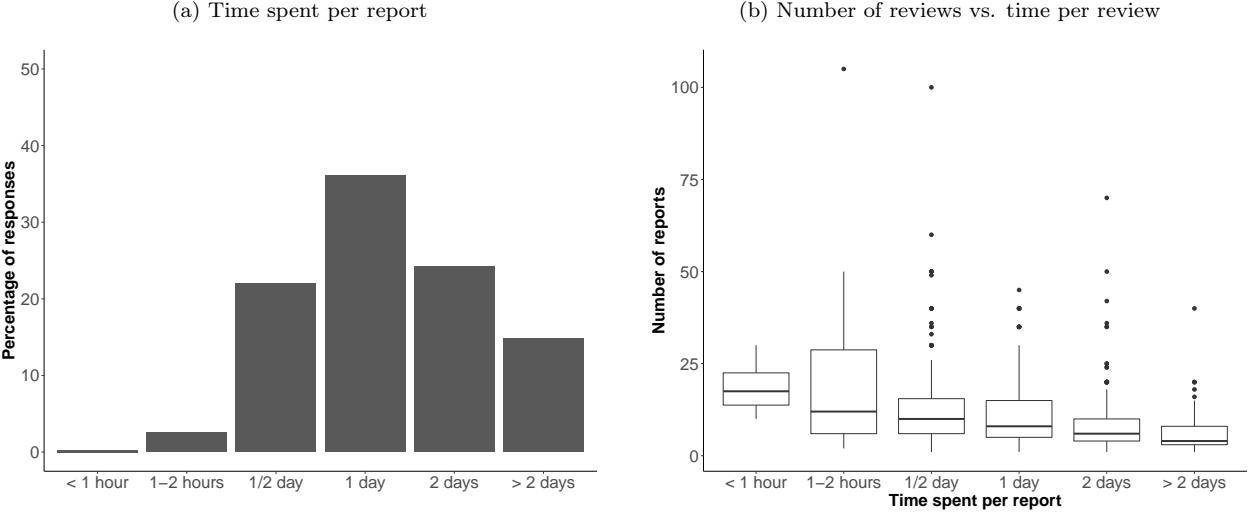
Figure 4: The peer review efforts of referees



Notes: Both panels use responses to question Q22. Panel (b) shows the Lorenz curves of reports written by respondents, separated by journal type. On the x-axis, respondent percentiles are based on the volume of reports they write. The y-axis gives the cumulative proportion of reports written by respondents at or below each percentile. For instance, respondents above the 75th percentile are responsible for 80% of reports written for top 5 journals (blue line). The Gini coefficients, which measure the extent of inequality in the distribution, are stated in parentheses (where a Gini of 0 indicates perfect equality - 45° line). N = 1,483 for both panels.

total amount of time spent by respondents on peer review. If we assume that one working day corresponds to 8 hours, the average (median) number of working days is 12 (9). The top 10% of the distribution dedicates 25 working days or more per year to writing referee reports, which is substantial for an activity that involves little to no financial compensation.

Figure 5: Time dedicated to peer review

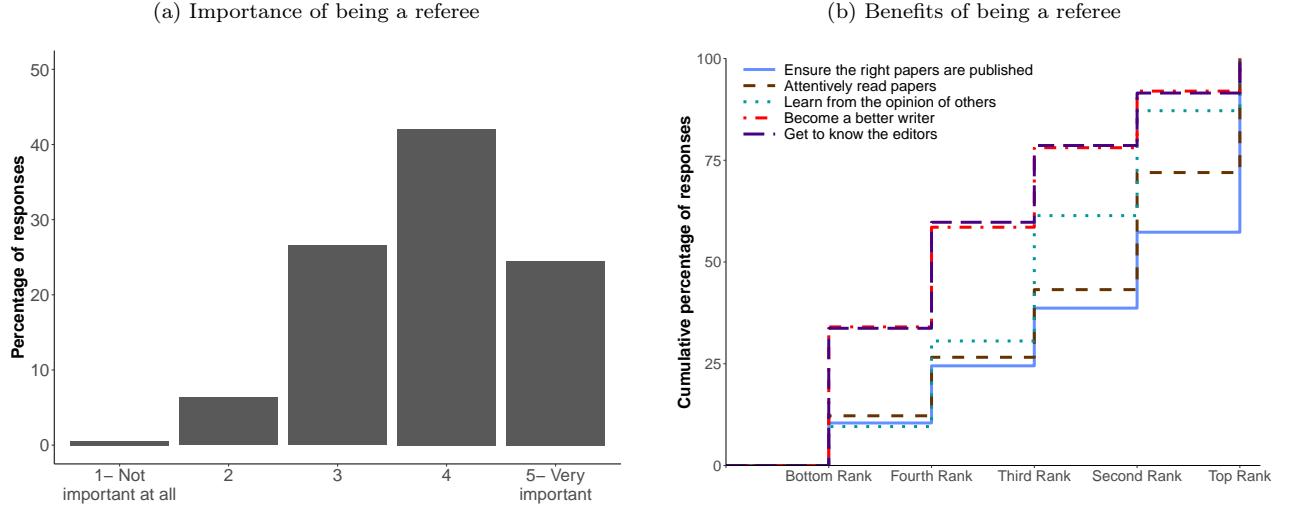


Notes: Panel (a) is based on responses to question Q25; Panel (b) uses both Q25 and Q22. In earlier versions of the survey, the response “1/2 day” was used instead of “half a working day.” N = 1,483 for both panels.

In line with the time and effort dedicated, most respondents consider their role as a referee to be an important part of their job (see Panel (a) of Figure 6). Among the main benefits of performing this role, respondents ranked highly their ability to “ensure the right papers are published or rejected” and “attentively read papers [they] would never read otherwise” (see Panel (b) of Figure 6). In regard to heterogeneity,

researchers who review more for top 5 journals tend to assign more importance to ensuring that the right papers are published and less importance to becoming a better writer or getting to know the editors. On the other hand, junior researchers tend to assign a higher priority to getting to know the editors and learning from the opinions of others, and a lower priority to ensuring the right papers are published (see Appendix Figure E.5).

Figure 6: Evaluation of the role of referee



Notes: Panel (a) is based on responses to question Q4; Panel (b) is based on responses to question Q35. N = 1,474 for both panels.

3.2 Allocating papers to journals, editors, and referees

In this section, we focus on the initial assignment of manuscript submissions to journals, editors, and reviewers. We ask whether the assignment process is satisfactory both from a quantity point of view (distribution of workload) and from a quality point of view (matching on relevant topics, skills, and an absence of conflict of interest). We postpone to Section 3.4 the discussion of the initial step of desk rejections, which will be better examined when we address processing times.

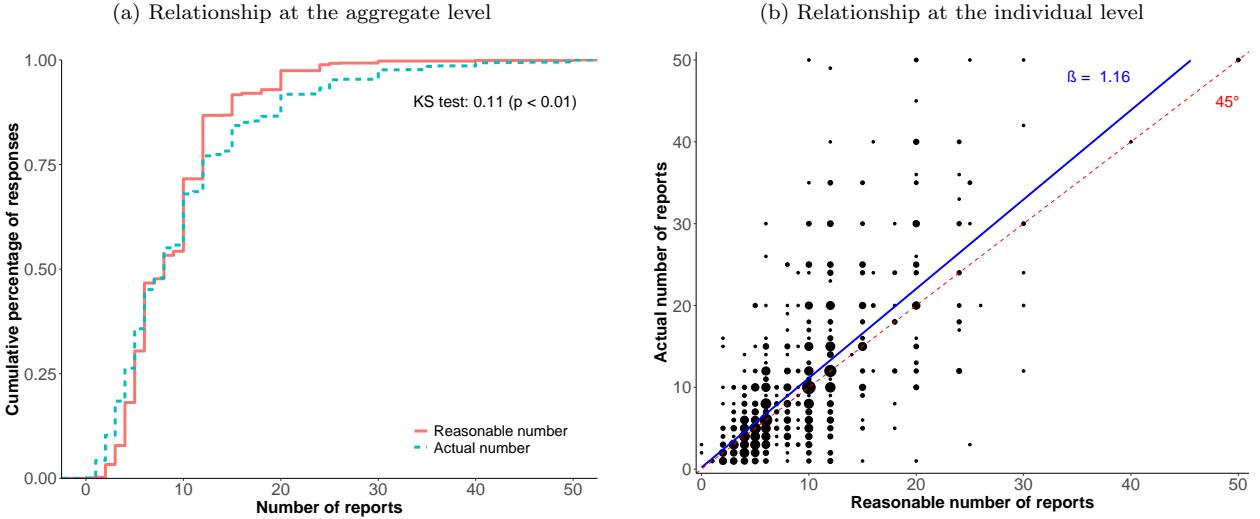
3.2.1 Mismatch on the quantity dimension

To assess the extent of misallocation in the volume of reports assigned to each referee, we proceed in two complementary ways. First, we examine how the number of reports written by respondents relates to how many they think is reasonable to write.¹¹ Second, we relate individual demand for peer reviews to individual supply by examining the link between number of submissions made and number of reports written.

3.2.1.1 Actual vs. reasonable volume of reports At the aggregate level, Panel (a) of Figure 7 shows a substantial amount of overlap between the cumulative distributions of the actual number of reports written by respondents and the number they think is reasonable to write. Only the top quartile of most active reviewers appears to be doing slightly more than deemed reasonable. A linear regression of the actual number on a reasonable number of reports yields a slope coefficient of 1.16 i.e., quite close to 1 (although significantly different from 1, p-val < 0.001). However, Panel (b) of Figure 7 shows that this apparent

¹¹In the survey, Q28 asked respondents “What do you think is a reasonable number of reports to be assigned per year?” This wording might have been ambiguous, if respondents did not know whether to answer for themselves or for a representative referee. To address this concern, we asked respondents to our follow-up survey to consider separately what is reasonable for themselves and what is fair “for someone at [their] career stage and with a similar submission activity.” The Spearman correlation between these answers was very high ($\rho = 0.92$; $p\text{-val} < 0.01$), so we focus on the first question in subsequent discussions.

Figure 7: Actual vs. reasonable number of referee reports per year



Notes: Both panels are based on responses to questions Q22 and Q28. In each one, we removed 4 observations with an actual number of reports > 50 to improve clarity. In Panel (a), KS is the Kolmogorov–Smirnov test; in Panel (b), the blue line is the linear relationship between the actual and reasonable number of reports (outliers included), with β being the regression slope coefficient (intercept = -0.26). N = 1,478 (1,474 without outliers) for both panels.

alignment hides considerable heterogeneity. While the Spearman correlation is strong ($\rho = 0.73$, $p\text{-val} < 0.01$), only about 25% of respondents write the number of reports they deem reasonable (45° diagonal) and the standard deviation of the difference ($\text{actual} - \text{reasonable}$) is 6.6 (mean of 1.2). Interestingly, 36.9% of researchers report writing fewer reviews than they would find reasonable. This is especially true among the least active reviewers who, as we saw in Section 3.1.2, more often tend to be postdoctoral researchers and PhD students. This contrasts with full (associate) professors who on average produce 2.7 (1.9) more reports than they consider reasonable ($p < 0.01$ for both). A regression analysis of the ratio of actual-to-reasonable number of reports on field confirms these patterns, with junior researchers (PhD candidates, postdocs, and assistant professors combined) having on average a lower ratio (see Panel (a) of Appendix Figure E.6). However, this is mostly driven by postdoctoral researchers and PhD candidates who tend to write many fewer reports than they consider reasonable (see Appendix Figure E.7).¹² To clarify our interpretation of these results (see Footnote 11), in our follow-up survey we asked respondents how many reports they think is reasonable for themselves to write. The majority (26/44) of junior researchers indicated that they currently write fewer reports than they think is reasonable for themselves, while most senior researchers (34/67) said that they write more than is reasonable. These findings suggest that some refereeing work could be allocated towards early-career researchers who would benefit from more practice and exposure to emerging research.

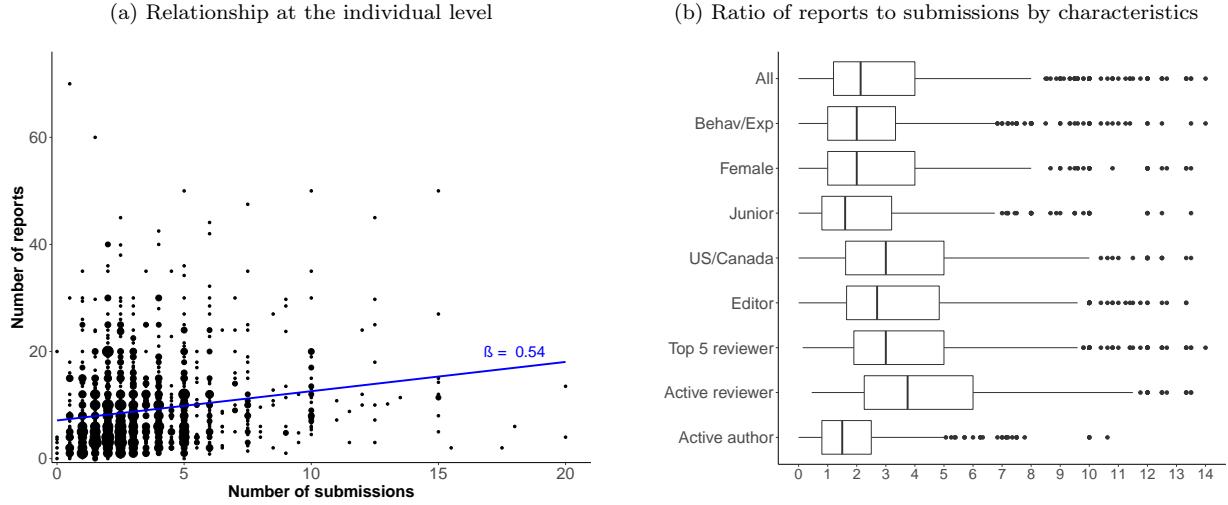
3.2.1.2 Number of submissions vs. number of reports Another way of assessing whether the allocation of reports across researchers is balanced is to contrast how many reports respondents write (their individual supply) with how many reports they generate via their submissions (their individual demand). To make the comparison between annual submissions and reports written as clean as possible, below we focus entirely on economics journals.

Panel (a) of Figure 8 shows that the relationship between the two is fairly weak. A regression of reports on submissions indicates that respondents write on average 0.54 additional reports for each submission they make, with a constant of 7.2. In line with this, the Spearman correlation is low ($\rho = 0.22$; $p\text{-val} < 0.01$). Respondents in the bottom quartile of the distribution of submissions write an average of 8.8 reports a year, not very far away from the average of 12.2 reports among respondents in the top quartile, even if the

¹²We chose to report differences when discussing the general characteristics of the relationship so as to offer a clearer interpretation of the statistics and include respondents who indicated 0 reports as a reasonable number (2 respondents). For the regression analysis, we chose to study the ratio of actual-to-reasonable number of reports so we can look at relative changes. However, a similar picture emerges whether we look at differences or ratios.

difference is statistically significant ($t = -5.08$, $p\text{-val} < 0.01$). While the ratio of average reports written to average submissions is 2.8, there is substantial heterogeneity: about 40% of respondents write at least 3 times as many reports as they make submissions, while about 15% write fewer reports than they make submissions (ratio less than 1). In terms of observable characteristics, we note a discrepancy between researchers based in the US/Canada and the rest of the world (ratio of 3.8 vs. 2.4, $t = 5.54$, $p\text{-val} < 0.01$). In line with the comparison previously made between actual and reasonable number of reports, the ratio of reports to submissions is lower for more junior researchers and higher for editors and referees for top 5 journals (Panel (b) of Figure 8). Looking at the relationship between reports and submissions by position, we observe that it is strongest for full professors and weakest for postdoctoral researchers and PhD candidates (see Appendix Figure E.8).

Figure 8: Referee reports vs. submissions per year (economics journals only)



Notes: Both panels use responses to questions Q22 and Q1. To estimate the number of reports written specifically for economics journals, we used answers to question Q23, and removed the percentage for “journals in other disciplines” from the total number of reports. In Panel (a), 4 outliers were removed from the figure for clarity. The blue line is the linear relationship between submissions and reports written (outliers included), with β being the regression slope coefficient. In Panel (b), the ratio was created by taking the annual reports written divided by the average annual submissions. In order to improve the clarity of the figure, outliers with a ratio > 14 were removed (57 respondents - 3.9%). N = 1,466 for Panel (a) and N = 1,459 (1,402 without outliers) for Panel (b).

To better interpret these ratios, it is useful to ask how many reports need to be written by each author who submitted a paper to a journal for the system to be in balance under various parameter conditions. For this purpose, we created a calculator that considers the following set of parameters: number of co-authors per submission, number of reports needed per paper, desk rejection (DR) rates, revise and resubmit (R&R) rates, and the number of rounds of review needed to make a decision on a manuscript. Throughout, we assume that the supply of referees available to a journal is equal to the pool of authors who made at least one recent submission to the journal. Since results are highly sensitive to the chosen parameter values, we have made the calculator publicly available on our OSF site (<https://osf.io/eczkv/>) for readers to use assumptions they find reasonable or to simulate the impact of proposed changes, such as increasing the desk rejection rate.

To illustrate how the calculator can be used, we contrast two scenarios. In the first scenario, we consider a DR rate of 60% and a R&R rate of 10% of reviewed manuscripts, numbers that are representative of what happens at some top journals. In the second scenario, we consider a DR rate of 30% and an R&R rate of 20%. In both scenarios, we assume that journals contact three referees and have two revision rounds per manuscript. The calculator estimates that a solo-authored submission creates demand for 1.4 referee reports in the first scenario versus 2.9 in the second. This emphasizes the impact that manuscript processing policies, especially DR rates, can have on demand for referee reports.

This modeling exercise gives a way of estimating under different parameters what the relationship between

authors' submission and report activity might be, if the goal was to equalize their supply and demand from the system at any given time point. However, such a static approach might not be appropriate. In particular, one could reasonably argue that it is beneficial for established researchers to write more reports relative to their submissions in order to allow junior researchers to focus on publishing. Thus, this second metric of reports-to-submissions should be carefully evaluated within subgroups (e.g., see Figures E.7 and E.8) and in conjunction with our first metric of actual-to-reasonable volume of reports before drawing any conclusions regarding the misallocation of the refereeing burden. Finally, we note that the statistics presented in this section obviously say nothing about the way the refereeing burden should be optimally allocated across time and various subgroups. At best, our analysis highlights important gaps in our understanding of this complex issue and the need for further theoretical and empirical work to provide greater clarity.

3.2.2 Mismatch on the quality dimension

Along with misallocation in the quantity of refereeing work, there are also questions regarding the quality of the allocation along multiple dimensions: submissions that are unsuitable for the journal considered, manuscripts that do not fit the reviewer's research interests and/or skill set, and assignments that imply a conflict of interest for the reviewer or editor.

3.2.2.1 Submissions unfit for journals Chronologically, the first risk in the allocation process is a mismatch between a manuscript and the journal it is submitted to. Because top 5 publications are a key input into tenure decisions (Heckman and Moktan, 2020), researchers have an incentive to "shop around" between elite general interest journals to find one that will accept their submission, even if their manuscript might fit better in a specialized field journal. This tendency to repeatedly submit to journals with low acceptance rates likely contributes to ballooning submission volumes and causes top journals to demand referee reports for the same manuscripts that are unlikely to be published.¹³ Based on simulated scenarios calibrated with journal acceptance data, Azar (2004) estimated that a paper of average quality¹⁴ is likely to be submitted to journals between 3.2 and 5.5 times before being published, which we suspect has risen further since then, partly due to increases in both manuscript production and journal selectivity. The issue of handling too many papers that are unsuitable for publication also likely concerns lower-ranked journals, especially those with far less stringent desk rejection policies. We will come back to this issue and potential solutions in Section 3.4.

3.2.2.2 Mismatch between submissions and reviewers' skills Even if a manuscript fits a journal, editors might send referee requests to researchers who are ill-equipped to evaluate it for a number of reasons. Panel (a) of Figure 9 shows that most researchers turned down at least one of the peer review requests they received in the two years prior to taking the survey, with over 60% of respondents rejecting more than 1 in 10 requests and nearly 30% rejecting 1 in 4 requests or more. Interestingly, the Spearman correlation between the likelihood of declining a new request and the total number of requests received is relatively low ($\rho = 0.21$, $p\text{-val} < 0.01$), suggesting that the mismatch is not simply in terms of volume.¹⁵ In fact, except for the bottom 25% of least solicited reviewers (who almost never reject a request), rejection rates do not appear to depend on the number of requests received (see Panel (b) of Figure 9).

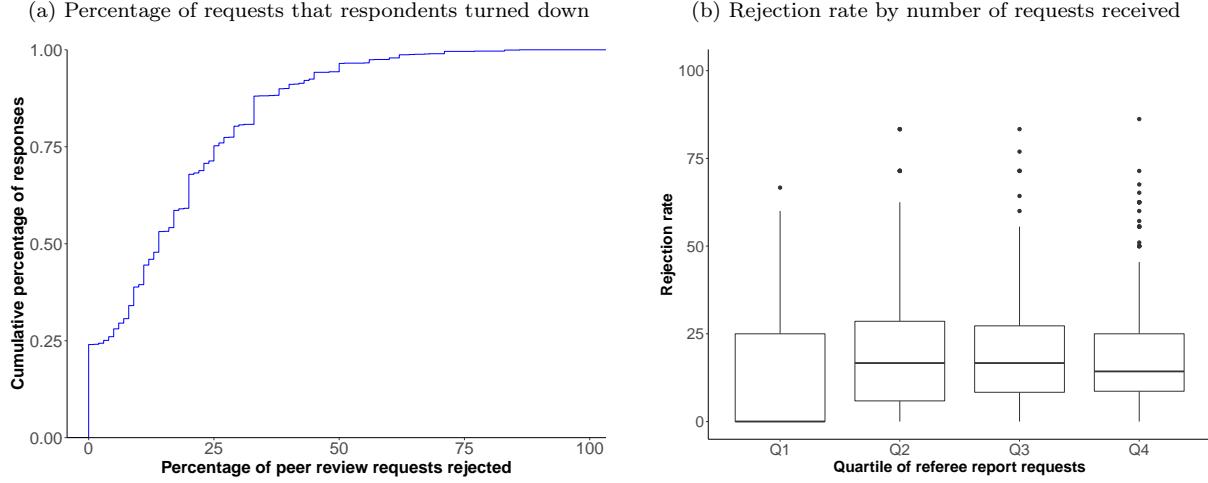
To understand rejection patterns, we asked respondents to provide the main reasons that led them (or tempted them) to reject a request (Figure 10). Unsurprisingly, about two-thirds of respondents selected "lack of time" as one of their main reasons; however, the point-biserial correlation between reporting a lack of time and the number of reports written is only 0.17 ($p\text{-val} < 0.01$). Researchers who have held an editorial position, female researchers, and more active reviewers were more likely to select "lack of time" (see Appendix Figure E.9). Importantly, about 50% of respondents selected "too remote from research field" as one of their reasons and 25% selected "inability to judge the paper." Thus, it appears that many respondents receive

¹³Submission volumes have risen substantially between 2010 and 2020, with *American Economic Review* (AER) submissions growing 37.5% from 1,476 to 2,029 and *Econometrica* submissions growing 61.3% from 747 to 1,205 (AER-1-a, ECMA-1-a, Table 4). This change has been fairly consistent year-to-year, with single-digit positive growth rates in most years.

¹⁴The author assumes that papers are distributed from 1 to 10 on the quality dimension, but does not offer a definition of quality. Manuscript acceptance rates are assumed to be a function of both paper quality and journal ranking.

¹⁵The Spearman correlation between the likelihood of declining a new request and total number of requests accepted is even lower, negative, and insignificant ($\rho = -0.03$, $p\text{-val} = 0.30$).

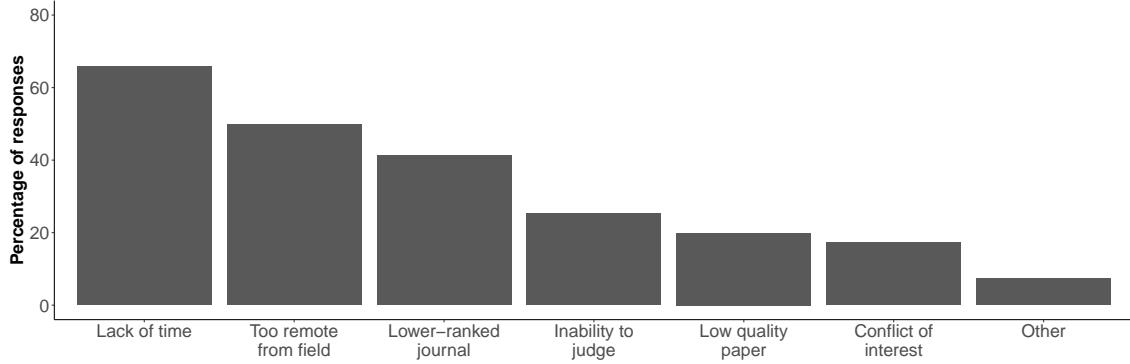
Figure 9: Rejections of peer review requests



Notes: Both panels combine responses to questions Q29 and Q30 to get the number of reports rejected. The percentage of requests rejected is then calculated as $100 \times \text{requests rejected}/(\text{requests rejected} + 2 * \text{reports written})$. “Reports written” is doubled because the number of rejections was given for a two-year timeframe. N = 1,483 for both panels.

requests for papers unrelated to their research or for which they do not have the right skill set to judge. A non-trivial fraction of respondents also mentioned the lower rank of the journal (41% of respondents) and the lower quality of the submission (20%) as reasons for rejection. This finding reinforces the earlier point made regarding the lack of suitability of certain papers submitted for review.

Figure 10: Reasons for rejecting or being tempted to reject a request

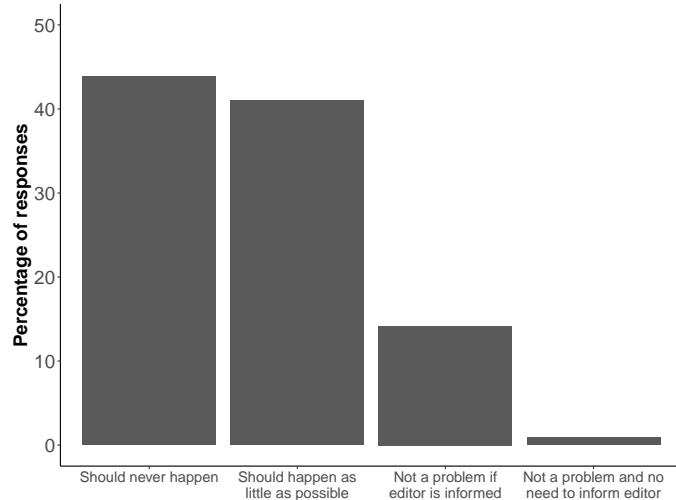


Notes: The figure pools responses to questions Q31 and Q33 as each respondent only answered one of these questions; the percentages do not add up to 100 as respondents could select multiple reasons (2 reasons selected on average). N = 1,383 (as 100 respondents selected no rejections or temptations to reject and no reasons for being tempted).

3.2.2.3 Conflict of interest involving authors and reviewers/editors Although respondents often seem to receive requests for papers that are not aligned with their research interests, they sometimes also receive papers that are *too* aligned with their own interests. Indeed, about 1 in 6 respondents mentioned a conflict of interest as one of the main reasons for rejecting or being tempted to reject a request. While conflicts of interest in peer review can take multiple forms, they generally refer to situations in which the reviewer may have a personal stake in the outcome of the review process, for instance due to close relationships with the author(s) (whether good or bad), competing research, or research that furthers one’s own agenda. Conflicts

of interest can be difficult to handle, because it is often not trivial to distinguish between relationships that imply a strict conflict of interest, relationships that could lead to a biased judgment, and those that do not threaten the impartiality of the review process at all. For instance, feelings about referees evaluating co-workers or former students without active collaborative relationships might be divided. Furthermore, reviewers might underestimate the impact of the specific ties they have if biases in the review process are unconscious. For these reasons, researchers are likely to have diverse views about what constitutes a serious conflict of interest and how it should be handled. As shown in Figure 11, there is general agreement that referees should never or rarely review manuscripts from friends or co-authors, with only a minority finding it acceptable as long as the editor is informed. However, about 40% of respondents selected the answer “*this should happen as little as possible but cannot be avoided sometimes*,” suggesting a possible trade-off between soliciting expert views that might be biased or getting less informative reviews that might be more neutral. This could be especially true in smaller fields, where it might be more difficult to find knowledgeable referees who do not have a conflict of interest. In terms of heterogeneity, behavioral/experimental researchers were less likely to select “*this should never happen*” and more likely to select “*this should happen as little as possible but cannot be avoided sometimes*,” and researchers from the US/Canada were less likely to select “*this should never happen*” and more likely to select “*this is not a problem as long as the editor is aware of the potential conflict of interest*” (see Appendix Figure E.10).

Figure 11: Attitudes toward co-authors and friends peer reviewing



Notes: Figure based on responses to question Q34. N = 1,474.

Research evidence on the matter is mixed, but a recent paper points towards network effects between authors and their reviewers (Carrell et al., 2022). While degrees of separation in co-author networks (beyond direct co-authorship) appear to have little influence on reviewer recommendations, the paper finds evidence of a “club effect” whereby referees view authors from the same or similarly ranked PhD programs more positively.¹⁶ Likewise, these results and results from Brogaard et al. (2014) and Colussi (2018) suggest that editors may give preferential treatment to authors they are connected to (i.e., current/former colleague, same PhD institution, PhD advisor, co-author), while Ersoy and Pate (2021) find a bias in favor of authors at top institutions, though this effect only exists for male authors. Brogaard et al. (2014) note that papers from authors connected to editors actually receive more subsequent citations than other papers, so a reduction in search costs for good papers is an alternative explanation to pure nepotism.

Despite evidence that networks can matter for both reviewers and editors, policies from top economics journals often only address the second group. Most of the top 5 journals have policies addressing editorial conflicts (except for the *Quarterly Journal of Economics* (*QJE*) - see Table 4); links to the journal policies are

¹⁶Specifically, they estimate that reviewers are 6% more likely to provide a positive review to authors who attended the same PhD institution. Reviews also seem to be more positive when both the reviewer and the author(s) have previously published in a top 5 journal.

available in Appendix Section A.3. All of the policies advise against or prohibit editors handling manuscripts submitted by co-authors or advisees, to varying extents.¹⁷ The policies at the *AER*, *Journal of Political Economy (JPE)*, and the *Review of Economic Studies (REStud)* also extend to authors from the same institution and more generally to any editor-author relationship that might create an actual or apparent conflict. But only the *JPE* has an explicit policy about what constitutes a conflict of interest for reviewers.¹⁸ Vagueness about conflicts can create potential ambiguities for referees, which may contribute to quality mismatch. Even when specific guidelines exist, it is usually up to the reviewer to reject a request or notify the editor about potential conflicts in the current status quo, since economics mostly practices single-blind review. Authors have no ability to challenge editorial decisions on the basis of conflicts of interest, even when it is clear that a reviewer showed undue hostility towards the approach or findings of a paper.

3.2.2.4 Conflict of interest involving editors and reviewers To assess whether conflicts of interest between referees and editors could be a concern, we first examined how the rate of rejection of peer review requests relates to the respondent's assessment that they write more reviews than deemed reasonable (ratio of actual-to-reasonable number of reports). We find fairly consistent rejection rates across the quartiles of the distribution (see Panel (a) of Appendix Figure E.11). This would suggest that some researchers, particularly those in the top quartile, might not be rejecting requests as much as they should. There are many reasons this could be the case. One might be the pressure that comes from referees having their own submissions at a journal. In the race to publish, staying on good terms with journal editors could be an asset in advancing the researcher's career. In particular, this pressure might be more substantial when researchers have (more) papers that have been given an R&R at a particular journal. To test this hypothesis, we asked respondents in our follow-up survey whether they “*feel pressured to referee more.... than [they] think [they] should*” under these circumstances. Of those who responded, 69/116 either somewhat (39.7%) or strongly (19.8%) agreed that they feel pressured in this situation. While our survey had a small sample size and is unlikely to be representative of economists in general, this evidence is suggestive that referees' own submissions could cause them to accept more review requests than they otherwise would.

3.2.3 The concentration of editorial and refereeing responsibilities

As we already noted in 3.1.2, participation in the evaluation side of the peer review process is heavily concentrated in the hands of a limited number of academics. This is particularly evident in the assignment of editorial positions and prestigious refereeing work. This could matter for a number of reasons, including if the concentration of editors and referees in particular groups tends to disproportionately benefit authors from the same groups. Below, we investigate the extent to which the current imbalance is associated with researcher characteristics such as co-author networks, geographic location, institutional affiliation, and gender. Before doing so, we would like to stress that these demographic comparisons are merely descriptive statistics, which do not control for other determinants of representation. There are also many other demographic variables, such as ethnicity and age, that are potentially relevant but difficult to study due to a lack of data. Nevertheless, to conduct our investigation, we directly compare the characteristics of the full study population of Andre and Falk (2021) (AF) with those of the subpopulations of (i) editors at the top 50 journals and (ii) referees for the top 5 journals (see Table A.2). Top 50 editors are those in editorial positions at the 50 highest-ranked journals according to the 2018 Scopus Scimago journal ranking. Top 5 referees are those whose names appeared in more than one referee acknowledgment from a top 5 journal between 2015 and 2020. We thank Peter Andre for deriving these additional statistics for us.

¹⁷Editors at the *AER* cannot handle manuscripts submitted by their co-authors from the past two years. This policy extends to the past 10 years at *REStud*. *Econometrica*'s policy only applies to “active co-authors.” At the *JPE*, “editors are generally not assigned manuscripts by [...] recent co-author(s),” suggesting that the rule might be waived in certain circumstances. Meanwhile, editors are not allowed to handle submissions from advisees at *REStud* and *Econometrica*, while the *JPE* again stipulates that it should “generally not” happen. The *AER* allows editors to do so once the advisees become tenured.

¹⁸Top journals in other disciplines often do better. For instance, *Nature*'s guidelines broadly define a competing interest as any situation that could threaten the actual or perceived objectivity of parties involved in the review and publication process, including referees, and give common examples of conflicts. Referees are encouraged to make editors aware of any potential competing interests ([Nature-1](#)).

3.2.3.1 Co-author networks From Table A.2, it is evident that researchers from these subpopulations are connected to extensive and elite co-author networks: top 5 referees and top 50 editors have an average of 9.5 and 11.7 unique co-authors, respectively, as compared to 5.8 for the study population. They also have more co-authors with top 5 publications (4.0 and 4.1 vs. 0.5) and tend to be highly influential in their

Table 4: Comparison of top 5 journal policies and statistics

	<i>AER</i>	<i>Econometrica</i>	<i>JPE</i>	<i>REStud</i>	<i>QJE</i>
Editorial Structure					
Has Lead Editor	Yes	Yes	Yes	No	No
Number of Coeditors (Excl. Lead Editor)	12	7	9	8	5
Number of Associate Editors (Board Members)	78	57	18	76	29
Number of Other Editors	2	1	5	0	1
Submission Policies					
Membership Required to Submit	No	Yes	No	No	No
Annual Membership (Sub.) Fee Amount	\$24-\$44	\$0-\$250	\$50-\$100	\$128	\$49-\$95
Submission Fee Amount (Nonmembers)	\$300	N/A	\$175	\$150	\$0
Submission Fee Amount (Members)	\$200	\$50	\$100	\$150	\$0
Need-based Submission Fee Discount	Yes	Yes	No	Yes	N/A
Submission Fee Refund for Desk Rejections	50%	No	No	No	N/A
Submission Statistics					
Number of New Submissions	2,029	1,205	1,304	1,768	2,181
Number of Resubmissions	67	191	77	205	55
Number of R&Rs Issued	122	183	113	181	27
Desk Rejection Rate (DR)	37%	41%	54%	53%	64%
Report Years for Submission Statistics	2019-2020	2019-2020	2020	2019-2020	2019-2020
Policies and Guidelines					
Conflict of Interest Policy for Editors	Yes	Yes	Yes	Yes	No
Conflict of Interest Policy for Reviewers	No	No	Yes	No	No
Guidelines for Writing Referee Reports	Unofficial	Official	None	None	None
Referee Incentives and Recognition					
Cash Reward	\$100	No	\$100	No	No
Submission Fee Waiver or Discount	No	No	Yes	No	No
Referee Certificate	No	No	No	Yes	No
Total Number of Referees (2020)	2,063	1,475	1,450	1,681	1,607
Number of Distinguished Referees (2020)	324	None	None	42	131
Turnaround Statistics					
Average First Response Time (Excl. DR)	13 weeks	15 weeks	19 weeks	15 weeks	6 weeks
Median First Response Time (Excl. DR)	11 weeks	13 weeks	16 weeks	13 weeks	6 weeks
Requested Referee Turnaround Time	5 weeks	6 weeks	6 weeks	8 weeks	4 weeks
Report Years for Turnaround Statistics	2016-2017	2015-2016	2019-2021	2018	2016-2017
Author Appeal Policies					
Author Appeal of Decisions Mentioned	Yes	Yes	Yes	No	No
Clear Procedures to Appeal	No	No	Yes	N/A	N/A

Notes: For information about the data sources and calculations made to produce this table, see the appendix discussion.

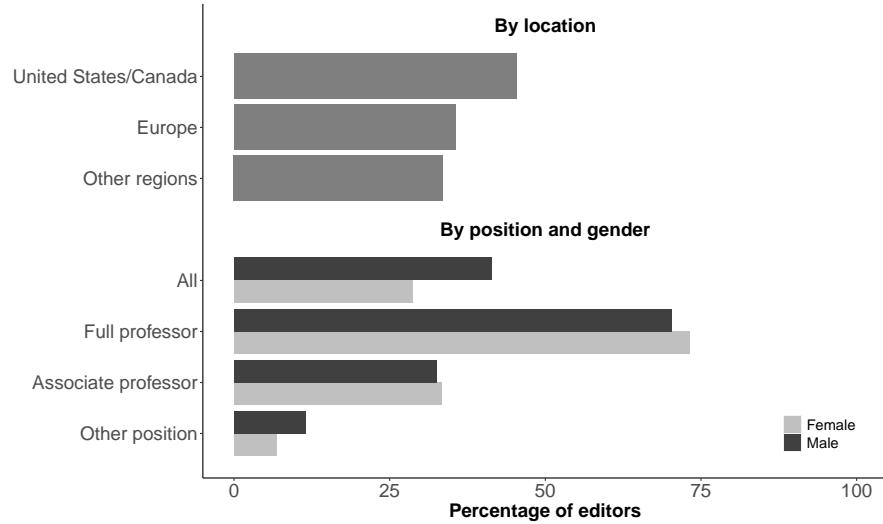
networks, as measured by eigenvector centrality. Finally, they are more successful in terms of their number of publications, the rank of the journals they publish in, and their ability to garner citations (h-index).¹⁹ Given that these researchers are connective nodes to elite networks, it is instructive to see who they are.

3.2.3.2 Geographic and institutional location Top 50 editorial staff overwhelmingly live and work in North America (72%) and Europe (22%), with a much smaller group in Asia and Oceania (6%). Patterns are very similar for the top 5 referee population, with percentages at 60%, 34%, and 5%, respectively. Meanwhile, the corresponding numbers for the AF study population are 34%, 40%, and 21% for the same regions. These differences suggest a substantial over-representation of North American economists in positions of power and prestige. A recent paper from Angus et al. (2021) makes this exact point, based on a dataset of editors as of summer 2020 compiled by the authors (more about this dataset in 3.2.3.4). Researchers from Europe and Asia/Oceania are both underrepresented, but to a much greater extent for Asia/Oceania.

In addition to geography, the institutional affiliation of researchers could be another driver of network effects. We see that the top 50 institutions, as ranked by the 2020 Shanghai Academic Ranking of World Universities, employ 42% of top 5 referees and 47% of the editorial population.²⁰ In contrast, only 12% of the AF study population is employed by a top 50 university, suggesting that editors and referees are highly concentrated at prestigious institutions. There is also some evidence that top journals tend to hire “in-house” i.e., from the same institution that publishes the journal. Among the top 5 journals in Table 4, this practice is evident at two of them: at the time of data collection, all (5/5) of *QJE*’s current coeditors were Harvard-affiliated, while the lead editor and 6/9 of the coeditors at *JPE* were Chicago-affiliated. However, such high levels of in-house concentration are not observable among these journals’ associate editors or at the other top 5 journals.

If geographic and institutional location facilitate the formation of network effects, these imbalances could harm the publication prospects of underrepresented groups. Of course, in addition to network effects creating career success, researcher institution and location could both be endogenous to success if e.g., publishing in prestigious journals leads to better job offers. This could help explain why successful researchers are so concentrated, but it may suggest that already-successful researchers who relocate to more elite networks could benefit from a positive feedback loop.

Figure 12: Demographics of editor population



Notes: The figure is based on responses to question Q24. N = 1,459.

¹⁹See Appendix A.2 for more information about the meaning of the various measures.

²⁰Most of these universities have big departments, which contributes to their large footprint, but they also tend to have extremely high percentages of faculty who serve as editors. For example, 47% of the current economics faculty at Harvard (26/55, at the time of data collection) held an editorial position at one of the top 50 journals at some point from 2015-2020. Many of them have also held positions at these journals outside of that timeframe.

3.2.3.3 Gender We can also examine whether the subpopulations are representative of the wider population based on gender. Although evidence of gender bias in editorial decisions is mixed (as discussed in 3.4.2.3 and 3.2.4.11), male over-representation in positions of power could affect the status and career progression of female researchers in other ways. Cullen and Perez-Truglia (2021) find that male managers and male employees are more likely to form social ties with each other and that these ties are associated with subsequent promotions. A similar effect could be at work in the formation of network effects. Female researchers represent about 20% of the editorial population and 17% of the top 5 referee population, compared to 26% of the full population. These numbers suggest that female researchers are somewhat underrepresented in these roles. We note that this underrepresentation is likely associated with the fact that female researchers are also underrepresented in senior academic positions, such as full professorships. This seems to have some explanatory power, as male and female full professors serve as top 50 editors at approximately the same rate (8% vs. 7%; additional statistics provided by Andre). We find a similar pattern in our survey sample, with nearly identical percentages of male and female full professors who reported having occupied an editorial position (see Figure 12).

3.2.3.4 Editorial tenure and hiring practices In consideration of the findings above, we conclude the section by investigating journal hiring practices that might contribute to the current demographic concentration. Specifically, we look at (i) editorial lengths of tenure and (ii) how common it is for editors to work at multiple journals concurrently. For item (i), we use the example of *Econometrica* (ECMA-6-b). The journal's current editors and coeditors have uniform terms of 4 years; we confirmed with the journal that this is the standard length, noting that it can be extended. There is considerable heterogeneity in the tenures of associate editors: the standard term length is 3 years (currently being served by 60% of associate editors at the time of data collection), with the possibility of renewing. Many have extended their terms to 6 years (28%), 9 years (9%), and even 12 or 15 years in rare cases (2% each). Regarding item (ii), we consulted the dataset of editors that Angus et al. (2021) made publicly available. As we modified the dataset for the purpose of this analysis, we have made it available on our OSF page (<https://osf.io/eczkv/>). The dataset includes individuals in any type of editorial position at the 49 top economics journals (i.e., those journals rated as A* on the 2019 Australian Business Deans Council journal quality list). About 22% of the 1,817 unique individuals listed worked on multiple boards at once (287 at 2 journals, 90 at 3 journals, and 25 at 4 journals). This analysis is restricted to the editors of the 49 journals, implying that many editorships at other journals could be unrecorded. The practices of indefinite editorial term renewals and working concurrently at multiple journals might limit diversity if it shrinks the pool of unique editors.

3.2.4 Solving the allocation problem

In the previous section, we saw that the allocation of reviews across referees is far from optimal. Very active reviewers tend to write more reports than they deem reasonable, while many less active (often more junior) referees might be willing to review a little more. Authors who make very few submissions do not seem to review substantially less than authors who make many submissions. Furthermore, reviewers often report receiving requests that are misaligned with their current research interests or skills. Finally, a non-negligible fraction of respondents report rejecting or considering rejecting requests due to potential conflicts of interest. Many judge that conflicts are hard to avoid, so referees often accept these requests, even when the conflicts are relatively clear-cut. In consideration of these issues, below we explore potential improvements to various dimensions of the allocation process.

3.2.4.1 Towards a centralized platform Many of the allocation problems listed above are arguably caused by the imperfect and asymmetric information generated by a decentralized peer review system. Specifically, the editors at various journals have little information about the current peer review activity of a researcher, which might span across many different journals managed by separate publishers. To assess the level of concentration, we calculated the Herfindahl-Hirschman index (HHI) for the entire population of economics journals listed at <http://citec.repec.org/search.html>, with citation counts as a percentage of all citations used as market shares (available on our OSF page at <https://osf.io/eczkv/>). We exclude journals that do not primarily publish economics output. The results suggest an unconcentrated market (HHI = 109.37) if we assume each journal is its own firm, while they suggest moderate concentration (HHI

= 2,361.20) if we use the publisher as the firm (U.S. DOJ and FTC, 2015). Publishers vary in the level of centralized control they have over individual journals, so the true level of concentration is likely to be somewhere in between. The publisher HHI is almost entirely driven by Elsevier, which publishes about a quarter of the journals listed and accounts for almost half of all citations. Higher levels of concentration facilitate coordination between journals, which might help explain the limited inter-publisher cooperation that currently exists, as well as why Elsevier found it worthwhile to establish an extensive centralized platform for its journals (discussed below).

The low-to-moderate level of concentration implies that editors have limited knowledge about the number of recent submissions made by a researcher.²¹ In addition, they have fairly limited information on the type of research that potential referees might be currently interested in and able to review.²² Finally, they cannot easily screen for conflicts of interest due to insufficient awareness of researchers' networks. The setup of a **centralized platform** for gathering information relevant to the peer review process could help to bridge this informational gap by:

1. *Providing a better tracking of current peer review workloads:* With the help of journals and reviewers, the system could keep a tally of a researcher's currently active and recently completed assignments, as well as recent submissions. This would help editors allocate new assignments more equitably and allow for the possibility of **imposing a global limit** on each reviewer's annual volume of reports.
2. *Giving access to an enlarged and more diverse pool of reviewers:* The platform could broaden access to peer review by giving more visibility to early-career researchers and researchers in less known institutions. This would expand the pool of available reviewers.
3. *Improving the paper-reviewer match on research topics and methods:* Researchers provided with a central peer review account could more easily update their profile to reflect their current research interests and areas of expertise since all the information would be in one place.
4. *Reducing conflict of interest:* Reviewers could reference their PhD institution, current institution, history of mentoring, collaborations, and personal (good or bad) relationships with other researchers that might entail a conflict of interest. One caveat is that many network connections might remain unreported by researchers, either because they hope to receive favorable treatment or because of potential sensitivities around reporting negative relationships.
5. *Helping with the production of a peer-review track record:* The platform could store summary statistics of a researcher's entire peer review history (e.g., number of completed reviews, percentage of requests rejected, delays, etc). Keeping a printable track record could provide higher accountability and better acknowledgment of reviewers' contributions e.g., for tenure decisions.

Although no centralized system is currently in place, some steps towards more integration have been recently taken (i) at the publisher-level and (ii) via partnerships with a commercial platform called Publons.

- (i) Elsevier recently established the Reviewer Hub (<https://www.elsevier.com/connect/reviewers-update/introducing-reviewer-hub>) to facilitate peer review for its journals; we are unaware of similar efforts from other publishers. On this platform, researchers can link their Scopus profile to become discoverable to editors and can volunteer to referee for participating journals. Participation appears high: 19 out of the 20 Elsevier journals in Table A.3 have enabled this feature at the moment, with *The Lancet* being the exception. In addition, user profiles record both reviews in progress and review history. The platform has many benefits but could improve by sharing user workload with editors, mapping reviewer networks, and including non-Elsevier journals.

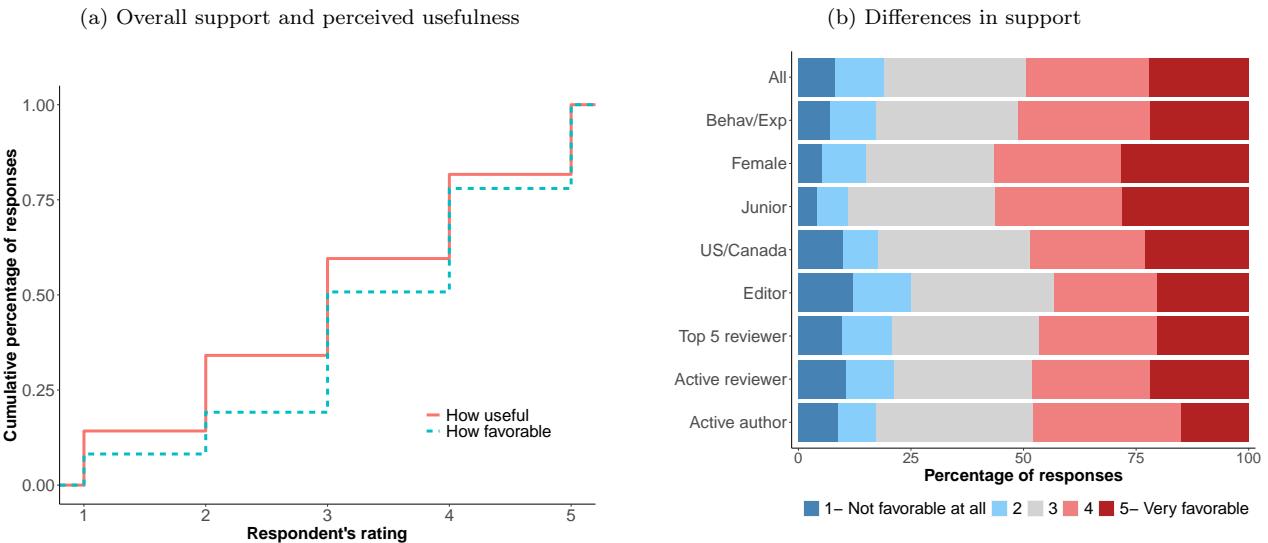
²¹Information about current researcher affiliations and contact details is also decentralized, which complicates keeping track of users. Researchers might end up losing access to their account on a publisher's website after changing institutional email addresses. Actively updating account details requires considerable effort on the part of users who may not see any benefit in doing so.

²²While reviewers are usually invited to reference a set of areas of expertise when they create their account with a journal, those fields are typically not updated after the initial account creation and might quickly become out of date. In practice, editors often select reviewers on the basis of their past submissions or publications on a similar topic, which again might not be reflective of their current research interests.

(ii) Publons (<https://publons.com/about/home/>) is an alternative platform not owned by a publisher.²³ On the platform, users can link their unique ResearcherID or ORCID (more on ORCID in 3.2.4.2) and can signal their interest in refereeing for participating journals such as *Experimental Economics*. Peer review histories can be imported from participating journals or manually added for others and made publicly available, as explained [here](#). We note that Publons' small take-up within economics and limited offerings are likely due to coordination issues and competing interests from publishers.²⁴

To understand support for a centralized system, we asked respondents to rate how favorable they would be to the more widespread use of Publons or a similar type of platform allowing researchers to document their (verified) peer review activity and register their interest in doing peer reviews. We also asked respondents to assess the usefulness for improving peer reviews of “*encouraging the use of a platform that tracks referee activity in a centralized way.*” Panel (a) of Figure 13 shows that about half of respondents would be favorable to such a platform (rating of 4 or 5), with only about 20% reporting unfavorable views (rating of 1 or 2); opinions as to whether this would be useful for improving peer review appear to be more divided. As shown

Figure 13: Views on the use of a centralized platform



Notes: Both panels are based on responses to question Q12. Panel (a) is also based on the answer to question Q6, item vii. N = 1,467 for Panel (a) and N = 1,459 for Panel (b).

in Panel (b), the biggest supporters of such a platform appear to be female and junior researchers, while the opinion of those with editorial experience appears to be more mixed. Although perceived returns to a centralized platform seem somewhat heterogeneous, on balance it appears that respondents would welcome steps towards more centralization.

While we acknowledge that setting up a centralized system for peer review is not a low-hanging fruit, we believe that there are intermediate proposals that could be implemented almost immediately.

3.2.4.2 Referencing peer reviews via ORCID Journals could offer researchers the option to link their peer review account to their **ORCID record**. As a matter of background, ORCID (for “Open Researcher Contributor ID”) provides researchers with a free, unique, and persistent digital identifier (researcher fingerprint), which can be connected to publications, affiliations, grants, peer reviews, and other professional activities. The system presents a high degree of interoperability, with a network of 1,100 member institutions (research organizations, publishers, funders, professional associations, etc.). Some journals such as *PLOS ONE* (PLOS-1) or *Nature Human Behaviour* (NHB) (Nature-2) already allow researchers to sync

²³While not operated by a publisher, it has partnerships with Springer Nature, Wiley, Sage, and others. A caveat is that journals under these publishers, such as *Econometrica*, do not necessarily participate in Publons services.

²⁴Only 8 of the 58 economics journals listed in Table A.3 were official partners as of data collection.

their reviewer account with their ORCID profile so that each completed review is automatically recorded on ORCID (without publishing the contents of the review). This system could be fairly easily generalized to all journals through the manuscript submission systems and publishers they are part of. In this respect, we note that 52 out of the 66 journals listed in Table A.3 already belonged to publishers that are ORCID Member Organizations as of data collection.²⁵

3.2.4.3 Submissions to families of journals and conditional acceptance offers As noted above, peer review in economics is characterized by limited vertical integration. Journal families are somewhat of an exception to this: both the American Economic Association (AEA) family (comprising the *AER*, *AER: Insights (AERI)*, the *American Economic Journal (AEJ)* series, the *Journal of Economic Literature (JEL)*, and the *Journal of Economic Perspectives*) and the Econometric Society family (with *Econometrica*, *Theoretical Economics (TE)*, and *Quantitative Economics (QE)*) engage in some level of internal coordination. But journal families in other disciplines, such as *The Lancet* and *British Medical Journal (BMJ)* families in medicine, do so to a much greater extent. For example, they often allow editors and authors to directly reallocate manuscripts from one journal to another within the family, to prevent the duplication of review efforts. *The Lancet* mentions that editors might **internally redirect** a submitted manuscript to a better-fitting journal (LAN-1),²⁶ while the *BMJ* sometimes makes transfers with conditional acceptance offers attached (BMJ-1). The usage of **conditional acceptance offers within journal families** could be particularly useful when a paper originally submitted to a top journal would be a good fit at an affiliated journal. To prevent authors from shopping around, offers for internal transfers and conditional acceptances could be time-limited.

3.2.4.4 Using reports from past reviews more systematically One form of coordination that some journal families in economics already engage in is the transfer of reports from one submission to the next. With referee permission, the *AEJ* journals allow materials from previous submissions to the *AER* and *AER: Insights* to be transferred with the new submission (AEJ-1). *TE* and *QE* have similar policies regarding materials from past submissions made to *Econometrica* (TE-1, QE-1). This even happens outside of publishing families. For instance, the *Journal of International Economics* allows authors to include past referee reports and decision letters from top 5 journals in their submission (JIE-1). As these procedures are author-initiated, one downside is that only positive reviews might be transferred. To circumvent that, journals could establish a **partnership to share review materials**. An agreement of this type already exists in neuroscience, with over 60 participating journals (<http://nprc.incf.org/>). We note that these practices raise questions about the ownership of review materials, which we return to in section 3.5.3.7. In addition, while these mechanisms could speed up the review process, it is not clear whether they substantially do so in practice. Manuscripts will often be substantially revised after the first submission, which could make the initial feedback obsolete. In this case, the original referees may be requested to give additional feedback on manuscripts, even though they may have forgotten key details. New referees are likely to be contacted as well, implying that the transfer of reports might have little effect on review durations. Finally, the fairness of this practice might also be questioned, since rejected manuscripts are more likely to have received negative reports in the original submission. As such, offering manuscripts a fresh evaluation with each new submission might be preferred. Six survey respondents shared this view, expressing their opposition to journals contacting prior referees who may be anchored to their initial recommendation to reject.

3.2.4.5 Collecting better data on reviewers' interests and availability Journals could improve the way they gather information on the research interests and areas of expertise of potential reviewers, as well as their availability to review. Instead of relying on outdated information provided at the time of account creation, journals could invite reviewers to **update their profile annually**. Besides selecting JEL codes, reviewers could enter specific keywords for topics that are of current interest to them. In addition, researchers

²⁵As of now, over 7,000 journals already collect the ORCID iD from corresponding authors (<https://info.orcid.org/requiring-orcid-in-publications/>). More information about the advantages for publishers of joining ORCID can be found here: <https://info.orcid.org/orcid-for-publishers/>. The full list of ORCID members can be found here: <https://orcid.org/members>.

²⁶Within economics, Elsevier also facilitates internal resubmission with its Transfer Tool and Resubmission Assistant (advertised [here](#)). In contrast with *The Lancet*, these transfers are made only at the discretion of the author.

could indicate how many papers they would be happy to review for the journal during the year and what time of the year they would be most available. Ideally, this information would be shared in a centralized manner, at least within a publishing group. As of now, we note that reviewers can already indicate the dates at which they expect to be unavailable in the Editorial Manager of Elsevier, but we do not know if this feature is available for other systems, such as Editorial Express.²⁷

3.2.4.6 Increasing the specialization of reviewers To improve the match on quality, more division of labor between reviewers could be introduced. For example, instead of requesting a careful examination of the entire manuscript, editors could ask reviewers to **focus on a specific section of the paper**. An early-career researcher could be asked to carefully check a statistical analysis or the logic of a proof, while a more senior reviewer could focus on the importance of the contribution relative to the existing literature. Besides specializing in what aspect(s) of a paper they assess, reviewers could more generally **specialize in what types of papers** they review. Papers cover an increasing variety of topics and methods, so paper-type specialization seems inevitable and already happens to some extent. This could be done in a way that is made common knowledge, with reviewers indicating on their CVs, journal profiles, or other platforms what types of papers they specialize in.

3.2.4.7 Increasing editorial specialization and the division of labor In addition to reviewer specialization, journals could also establish more **specialized roles for editors**. Editorial specialization already happens to some extent, as journals typically have staff members whose responsibility it is to check manuscripts for formatting and grammar. Roles requiring more judgment and expertise exist too: at the AEA and other journals, the Data Editor is tasked with assessing the reproducibility of author submissions (discussed more in Section 3.4.1). For another example, as Editor of *Economic Inquiry*, Preston McAfee had specialized coeditors who handled all manuscripts from specific subfields, such as sports economics, to supplement the more general coeditors (McAfee, 2010). Further division could be introduced, by specializing in tasks such as evaluating the validity of author proofs (as mentioned [here](#)) or verifying pre-analysis plans. Editorial specialization could act either as a complement to or as a substitute for some referee work. Deciding the extent of specialization and its relation to referee work requires some engagement with normative questions about the proper roles of editors and referees in peer review. 36 respondents expressed support for narrowing the scope of referee responsibility, usually either to a strict evaluation of the manuscript's contribution or its overall quality. This proposal could move peer review in that direction by creating a clearer division of labor and responsibilities between editors and referees.

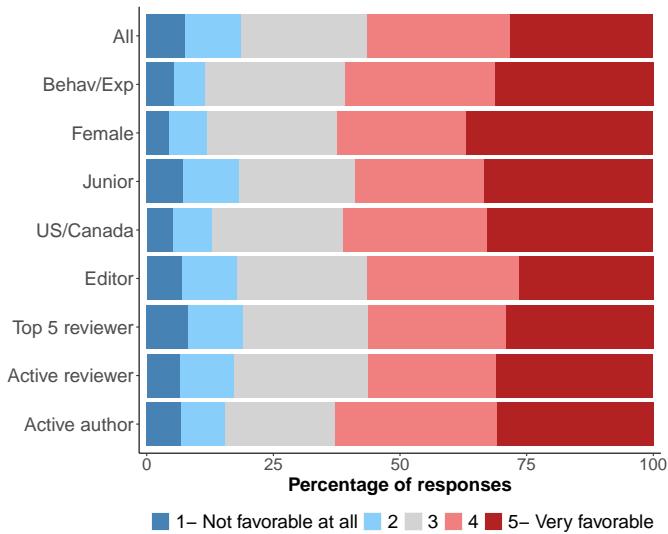
3.2.4.8 Providing clear conflict-of-interest guidelines To limit the chance that researchers accept assignments that they cannot be impartial about, all journals could provide clear and concise **conflict-of-interest guidelines for reviewers**. These guidelines could include concrete examples of situations where a conflict might exist and could be made readily available within the body of the review invitation email e.g., as currently done by the *Economic Journal* (*EJ*) ([EJ-1](#)). We note that the *EJ* guidelines also make reference to the more comprehensive guidelines published by the Committee on Publication Ethics (COPE) at <https://publicationethics.org/guidance>, which all journals could easily point to. Finally, the *Frontiers* journals also offer extensive guidelines to reviewers and editors that cover a wide range of personal, professional, and financial affiliations that could cause conflicts ([Front-2](#)). These guidelines include a list of questions that they should ask themselves about potential conflicts before accepting an assignment. In addition, the journal includes a link to the guidelines in the invitation email and requires reviewers to attest to their lack of conflict in a questionnaire before accepting a review engagement.

3.2.4.9 Allowing authors to suggest referees To facilitate the search for suitable referees, authors could be allowed to **suggest names** to be added to the pool of potential manuscript reviewers. Journals could request a list of potential reviewers from authors at the submission stage, with editors retaining final responsibility for judging referee suitability. Some journals including *Energy Economics* and *Economics*

²⁷For Editorial Manager, this can be done by going to “Update my information” in the menu bar and clicking on “Unavailable dates” at the bottom of the page. Researchers can also enter the name of up to 3 people who can be contacted by the publication in their absence.

Letters already do this ([ENEC-1](#), [ECOLET-1](#)), as do grant organizations such as the National Science Foundation. One concern is that authors will suggest referees who are likely to issue more sympathetic recommendations about their manuscripts. Research suggests that this is indeed the case ([Teixeira da Silva and Alkhatib, 2018](#)), although there are several possible explanations for why that is. For example, suggested reviewers might actually be better matches and care more about the topic. In any case, editors can always discount the recommendations and instead focus on the substantive comments. Since most papers end up being rejected anyway, seeking author input could help ensure that they perceive the process as a “fair trial.”

Figure 14: Support for allowing authors to disqualify certain reviewers



Notes: The figure is based on responses to question [Q20](#). N = 684.

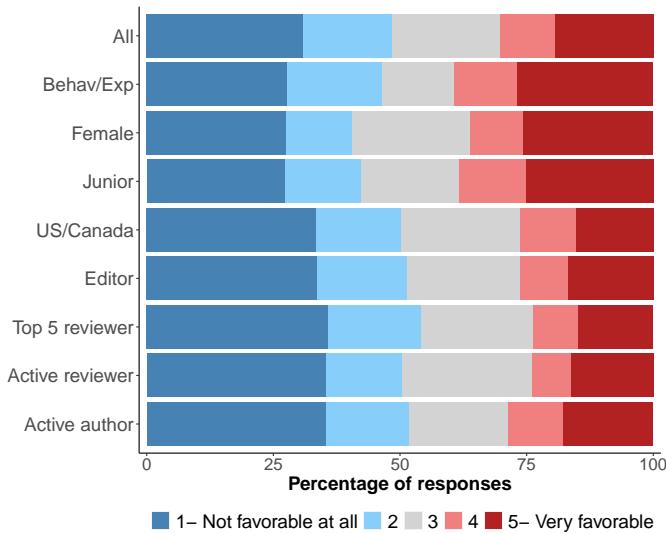
3.2.4.10 Allowing authors to oppose referees Some journals, such as *Science* and *PLOS ONE*, also allow authors to [request that certain reviewers be disqualified](#), again without commitment ([SCI-1](#), [PLOS-2](#)). With this option, authors can preemptively report potential conflicts with reviewers who might not be able to deliver impartial judgments. There is indeed evidence that the reports of author-opposed reviewers are more negative than other reviewers’ assessments ([Moore et al., 2011](#)), though this could be due to legitimate conflicts as well as manipulation. As for the previous proposal, a concern about author manipulation of the review process remains. As a safeguard, journals could follow the example of *Science* by asking authors to justify their exclusion request and introducing a limit on the number of exclusions (5 in this instance). For more discussion about the ethics of this proposal, see the case study published by [COPE Council \(2016\)](#). In a later version of the survey, we asked our respondents to rate how favorable they would be to a disqualification policy. Figure 14 shows that the majority of respondents would be favorable to such a policy; only about 20% would be unfavorable.

3.2.4.11 Making double-blind reviewing optional. There has never been a consensus in economics on the use of single- vs. [double-blind review](#), though attitudes have shifted over time in response to new evidence and technological changes.²⁸ In a randomized experiment, [Blank \(1991\)](#) found that double-blind review resulted in lower acceptance rates than single-blind review, with inconclusive evidence for referee bias. An interdisciplinary meta-analysis found weak evidence that review anonymity affects acceptance rates for female authors, relatively less-known authors, and those who publish many papers, with many contradictory studies ([Snodgrass, 2006](#)). The recent results from [Carrell et al. \(2022\)](#) (among referees) and [Ersoy and Pate \(2021\)](#) (among editors) are suggestive that author institution could be a biasing factor. But even if blinded review can eliminate bias under controlled settings, it may now be infeasible in practice. Attesting to

²⁸The *AER* currently uses single-blind review ([AER-3-b](#)), but it has switched back-and-forth multiple times. *Econometrica* has always used single-blind review ([Hengel, 2021](#)), while double-blind review is used at the *Journal of Financial Economics* (*JFE*), *Management Science*, and *Economic Inquiry* (*JFE-1*, [MS-1](#), [EI-1](#)).

this concern, the editors of the *AER* declared in 2012 their belief that double-blind anonymity had become nearly impossible in the age of search engines (Goldberg, 2012). To assess support among our respondents, we asked them to rate how favorable they are to double-blind reviewing. Figure 15 shows that nearly half of respondents appear unfavorable, while 20% are very favorable. We note that female and junior respondents tend to be more favorable. Given the inconclusive evidence and mixed support, one possibility could be to adopt a hybrid system where **authors choose whether to remain anonymous** or not, which is already an option at the *Nature* journals (*Nature-3-a*).²⁹

Figure 15: Support for double-blind peer review



Notes: The figure is based on responses to question Q19. N = 684.

3.2.4.12 Improving the diversity of editorial and referee pools Eliminating the biases of editors and referees, such as those relating to author institution and network effects, may be difficult in the absence of truly blind review.³⁰ That being said, **improving the diversity of boards** and of **referee pools** could mitigate the disproportionate benefit that well-connected authors may receive or it might help underrepresented groups form their own social ties. Beyond making a conscious effort to recruit candidates from underrepresented groups, proposals that could broaden the pool of editors over time include standardizing or setting **caps on editorial lengths of tenure** and **avoiding hiring those who are already editors elsewhere**. A potential downside is that valuable experience and institutional knowledge held by current editors could be lost with these policies. In terms of proposals to improve reviewer diversity, Elsevier recently published a series of articles surveying current initiatives about referee diversity (Eve et al., 2021c) and offering practical guides for editors (Eve et al., 2021a,b). For general guidance, individuals and organizations can also look to the AEA's list of best practices for improving the diversity and culture of the profession, including their advice about journal practices (<https://www.aeaweb.org/resources/best-practices>).

As a note of caution, aiming for greater diversity on editorial boards and referee pools than the profession has yet achieved could place disproportionate burdens on underrepresented groups. In fact, there are reasons to believe that some female editors in economics are already “overutilized;” since there are fewer female full professors and editors, it might be the case that those who reach those ranks are asked to serve on multiple editorial boards in order to get closer to gender parity. From the statistics discussed in Section 3.2.3 from Peter Andre, we saw that male and female full professors are approximately equally likely to be top 50 editors,

²⁹Of course, the sorting of authors into single- vs. double-blind review tracks could have perverse signalling effects, so offering such an option to authors might not necessarily reduce discriminatory assessments.

³⁰Journals could experiment with the blinding of both editors and referees to author identity i.e., “**triple-blind**” review, but it would considerably complicate the process of choosing a qualified referee who does not have a conflict of interest. The philosophy journal *Ergo* handles this by blinding associate editors and referees to author identity while giving managing editors full oversight and veto power over decisions made in the review process (*Ergo-1*). Authors such as Jung et al. (2017) have proposed other (automated) methods of assigning referees to manuscripts under triple-blind review.

but it is not reported how many boards the typical male and female editor serve on. To try to answer that question, we used the [Angus et al. \(2021\)](#) dataset to create some descriptive statistics. First, we find a difference in the average number of editorships that male and female editors hold (1.28 vs. 1.39, $t = 2.44$, $p\text{-val} < 0.05$). Additionally, a greater proportion of female editors (33/334) than male editors (82/1,482) serve on 3 or more boards (9.88% vs. 5.53%, $z = 2.95$, $p\text{-val} < 0.01$). Finally, although the absolute numbers are small at this level, the gender ratio among editors who serve on 4 boards almost reaches parity, with female editors comprising 11/25 (41%) of the total. Together, these statistics may suggest that accomplished female researchers are asked to perform additional service disproportionately often, in the name of gender parity. This emphasizes that care needs to be taken when implementing measures to improve diversity.

3.2.4.13 Limit of one journal assignment at a time for reviewers Consistent with the themes of widening the reviewer pool and mitigating quantity misallocation, a rule [limiting referees to one manuscript at a time for the same journal](#) could be implemented. Editors should already have information about whether referees are currently reviewing another manuscript at their journal, so implementation should be relatively straightforward. The potential caveats of this proposal are similar to those of the global limit i.e., that it might lead to suboptimal referee-manuscript matches and lengthened turnaround time if editors cannot find a suitable match.

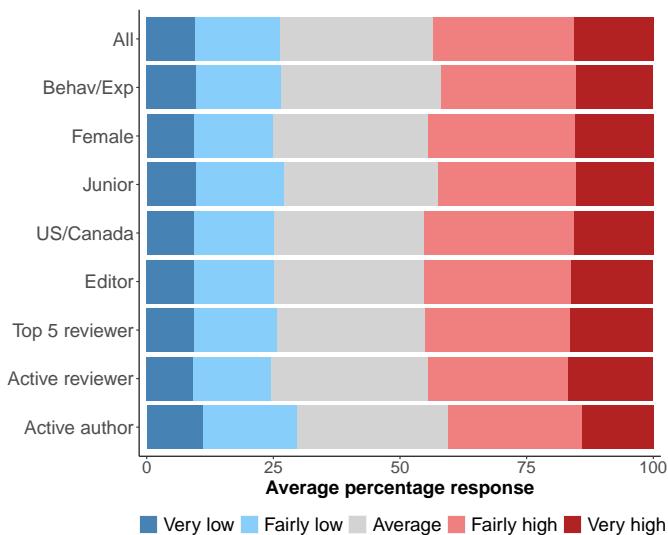
3.3 Content of the reports

3.3.1 Perceived quality of referee reports

As often noted, it is difficult to establish an objective and unambiguous measure of the quality of peer review reports. In evaluating the effect of introducing monetary incentives on referee behavior, [Thompson et al. \(2010\)](#) finds that payments appear to increase the length of reports but is unsure whether this is good, noting that “longer reports are not necessarily better” e.g., if they come with a multitude of unnecessary and inconsistent demands. Short of being able to objectively measure quality, an important subjective measure is whether the authors of the manuscripts being reviewed find the reviews they receive helpful. In the words of [Mavrogenis et al. \(2020\)](#), helpful peer reports “should be formal, constructive, honest, and polite.”

As shown in Figure 16, respondents seem moderately satisfied with the quality of the reports they receive. However, there is considerable variance: while about 43% of reports are judged to be of fairly or very high quality on average, the quality is judged to be fairly or very low in 27% of cases.

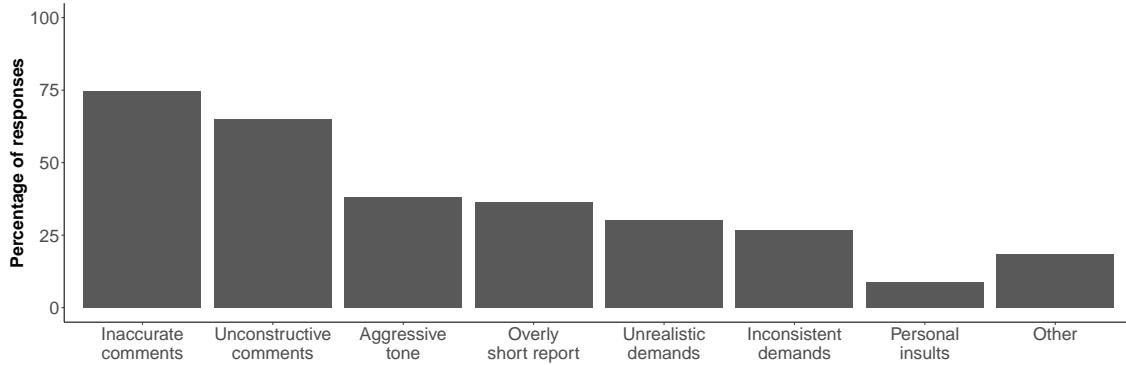
Figure 16: Perceived quality of reports received



Notes: The figure is based on responses to question Q2. N = 1,459.

There is some heterogeneity in views. Authors who make a higher number of submissions per year tend

Figure 17: Characteristics of low-quality reports



Notes: This figure is based on responses to question Q3. The percentages do not sum to 100 as respondents could select multiple reasons. N = 1,484 respondents.

to report a higher proportion of low-quality reports. Regression results also show that reviewers for top 5 journals report a higher proportion of high-quality reports on average (Appendix Figure E.12); to the extent that they are also more likely to submit and publish in top 5 journals, this finding might suggest that the quality of peer reviews tends to be higher at top journals.

We asked respondents to indicate what the main characteristics of the low-quality reports were. As shown in Figure 17, about 75% of respondents mentioned “*inaccurate statements about what the paper does or does not do.*” These inaccuracies in evaluation could be caused by a range of factors, including the limited time spent by the reviewer on the paper or their lack of expertise on the topic.³¹ About two-thirds of respondents also mentioned “*very vague and unconstructive comments*” i.e., comments made without proper justification or real effort to make suggestions for improvement. After these two main reasons, we note that about a third of respondents mentioned reports “*written with an aggressive tone,*” with around 10% citing “*personal insults.*” Among other complaints, respondents raised the issue of reviewers attempting to rewrite the entire paper for the authors (10 comments), as well as making self-interested judgments and/or subjective assessments based on taste rather than objective criteria (26 comments). There is some heterogeneity in the reasons selected for the low-quality reports, with (for instance) junior respondents being more likely to complain about unconstructive comments, female respondents about the aggressive tone of reports, and US/Canada-based authors about personal insults (see Appendix Figure E.13 for more details).

Complaints about the lack of professionalism in certain referee reports are obviously not restricted to the field of economics. In a survey run by Silbiger and Stubler (2019) among members of the academic STEM (Science, Technology, Engineering, and Mathematics) community, 58% of authors reported having received a peer review that they deemed unprofessional. Furthermore, while there was no evidence that the likelihood of receiving an unprofessional comment depended on demographic characteristics, members from minority groups were more likely to report that these types of comments negatively impacted their self-perceived ability, publication output, and career progression. Beyond simply being unhelpful, unprofessional comments may therefore constitute a barrier to the progress of underrepresented groups.

That said, we note that compared to other disciplines, the economics profession is often criticized for having a culture that discourages diversity. Findings from the *AEA Professional Climate Survey: Final Report* suggest that discriminatory behaviors and disrespectful interactions with colleagues likely contribute to a toxic climate in the field, with disparate impacts for underrepresented groups (Allgood et al., 2019). Other empirical research appears to support this conclusion. A text analysis of the Economics Job Market Rumors forum found that conversations about female economists tended to contain 42% less professional and academic terminology and 196% more personal and physical terminology (Wu, 2019). Recent research also suggests that female presenters receive 3.8 more questions during seminars than male presenters on

³¹In the “Other” comments, 42 respondents mentioned that comments about the manuscript were inaccurate, many of whom suggested that the referee seemed not to have read or understood the paper. In addition, 24 respondents argued that the referee was lacking in expertise.

average, and more questions of a patronizing or hostile nature (Dupas et al., 2021).³² Finally, there is experimental evidence suggesting that students might receive differential treatment from professors based on their perceived ethnic background (Milkman et al., 2015).³³ Altogether, the above evidence suggests that academic economics can be an intimidating and discouraging environment for junior researchers and those from underrepresented groups. Referee reports with unprofessional comments contribute to making the environment less productive and less welcoming of diverse perspectives.

3.3.2 When are referee reports most useful (if at all)?

Assessing the effect of referee reports on manuscript quality is difficult without observing the counterfactual of no peer review, but the evidence suggests a positive impact. Laband (1990) investigates the ability of referees to improve “marginally acceptable” submissions by proposing revisions to be made before publication in economics journals. He finds that reviewer suggestions and comments that are assessed to be more extensive in scope and more helpful to the author are positively associated with subsequent citation counts.³⁴ A more robust study in medicine that compares blinded experts’ assessments of manuscript quality before and after revisions concludes that reviewer suggestions are particularly helpful for improving discussions about study validity, conclusions, and confidence intervals (Goodman, 1994). But there may be limits to the positive impact of peer review, at least in terms of its ability to improve manuscripts’ citation counts (Hadavand et al., 2020, 2021).³⁵ Qualitatively, others argue that additional discussions and robustness checks demanded by referees can lengthen papers unnecessarily, obscure their main findings and conclusions, and place undue burden on authors (Berk et al., 2017).

To gain a better understanding of what constitutes a useful report, we asked respondents to indicate the type of feedback from reviewers that they find most helpful and would like to see more of. As Figure 18 shows, the comments found most useful are by far those that help the authors (i) “clarify the contribution of the paper relative to the literature” and (ii) “improve the existing analysis” (both selected by about three-quarters of respondents). Next, about a third of respondents mentioned comments pertaining to “the presentation of the results,” “shortening/restructuring the paper,” “missing previous work and references,” as well as comments that help “put in perspective the assumptions made in the paper.” On the other hand, very few respondents seem to find suggestions about “possible extensions” and “robustness checks” particularly useful. Although not entirely surprising, these points are worth keeping in mind, especially in light of the general trend towards lengthier papers, sometimes in anticipation of, or in response to, requests from reviewers. If anything, authors would welcome more help on the positioning of their paper and the core analyses.

3.3.3 Improving the quality of reports

3.3.3.1 Providing clear guidelines Economics journals often provide little concrete guidance to referees on how to write a report, either on their website or via email invitations sent by editors. In particular, reviewers are rarely provided with a list of evaluation criteria against which to judge a paper. They are also not told how to weigh various criteria when formulating their recommendation. Finally, templates illustrating well-formulated referee reports are almost never made available. In the absence of profession-wide norms about guidelines, some experienced researchers have offered their own perspective on how to write a good report. Examples include *Preparing a Referee Report: Guidelines and Perspectives* in 2015, as well as subsequent contributions *A Checklist for Reviewing a Paper* and *How to Write an Effective Referee Report and Improve the Scientific Review Process* in 2016 and 2017, respectively (Berk et al., 2015, 2016, 2017).

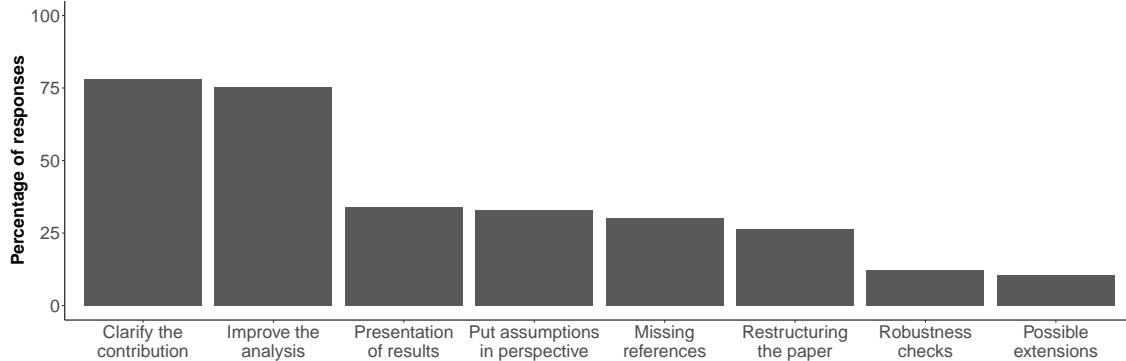
³²However, it is worth noting that some of these effects are only significant at $p < 0.10$ and that effect sizes tend to be small.

³³The authors ran an experiment where professors were sent emails claiming to be from a prospective student who wanted to discuss the professor’s research. Professors from the social sciences (including economics) and business disciplines responded at a lower rate to students with names suggesting they are ethnic minorities than to those with Caucasian-sounding names.

³⁴These findings should be interpreted with caution due to endogeneity concerns i.e., referees are not assigned randomly to manuscripts and the nature of reviewer comments might be a function of manuscript quality rather than vice-versa.

³⁵For articles published in *Economic Inquiry*, Hadavand et al. (2020) estimate the number of productive resubmission rounds to be just one by comparing the citation counts of papers that went through one round vs. more than one round, using a range of specifications. At top 5 journals, Hadavand et al. (2021) found additional improvement in citation counts for rounds past the first one, except for theory papers.

Figure 18: Most useful comments given by reviewers



Notes: This figure is based on responses to question Q7. The percentages do not sum to 100 as respondents were asked to select 3 comments in total. N = 1,463 respondents.

On its website, the *AER* shares a link to the Berk et al. (2015) guidelines, but caveats that they “do not represent official AEA policy” (Table 4). In recognition of the lack of uniform guidance, some journals have taken the initiative to provide their own. Very recently, *Econometrica* became the first top 5 journal to do this; they now ask referees to divide their report into a (i) summary section, (ii) an essential points section, and (iii) a section for additional suggestions (Table 4). In addition, some journals outside of the top 5 provide resources, including the *Review of Financial Studies* (*RFS*), the *Journal of Finance* (*JF*), and the *Journal of Monetary Economics* (*RFS-1, JF-1, JME-1*).

Providing concrete guidance more systematically could help reduce the variance in report quality. At the same time, the imposition of criteria or **templates** might frustrate some referees if they feel constrained in delivering their assessment. Another potential effect of guidance could be to increase the correlation between reviewers’ evaluations of a given manuscript. The evidence shows that inter-referee agreement is currently low or modest: Card et al. (2020a) find rank-order correlations of 0.25 between two referees’ recommendations, with comparable results from Welch (2014). However, inter-reviewer reliability can remain low even in contexts where guidance and templates are provided (Forscher et al., 2019). Additionally, it is not entirely clear whether higher levels of agreement between referees would be a positive development, since a stochastic component to recommendations might allow more bold research to be published (Armstrong, 1997).

To assess support for more guidance, we asked respondents to rate the usefulness of “*providing a set of guidelines for writing referee reports*” for improving peer reviews. Panel (a) of Figure 19 shows that most respondents would find guidelines useful, including 25% who would find them extremely useful. There is some heterogeneity, with female and junior respondents being the most favorable, and editors and active reviewers being the most skeptical. We also asked respondents whether journals or associations should provide templates for referee reports; overall, about 46% answered “yes,” with postdocs/PhD students being nearly two times more likely to give a positive answer than full professors (Panel (b) of Figure 19).

Those in favor of providing guidelines might look to some of the well-developed and instructive guidelines available at general science journals. *Nature*, *Science*, and the *Proceedings of the National Academy of Science* (*PNAS*) all offer bespoke guidance for the benefit of reviewers; they each make specific requests about what aspects of the manuscript should be addressed in referee reports (*Nature-3-b, SCI-2, PNAS-1-a*). Another approach is the one taken by *Frontiers*: they enumerate clear acceptance and rejection criteria on the peer review information page (*Front-1-a*), and also **require reviewers to rate manuscripts** from “low” to “high” on six criteria in the form of a questionnaire.³⁶ Journals taking this approach could easily adapt the questionnaire to match the type of research being assessed (theoretical, empirical, experimental, etc.). As a final note, another possibility is for journals to **provide examples of good reports**; this could be done in the context of a larger effort to practice open peer review (see Section 3.5).

For more examples of referee guidance, we link to various resources posted by individuals, journals,

³⁶This information is according to one of the co-authors of this report who refereed for *Frontiers*. The six criteria are originality, rigor, significance to the field, interest to a general audience, quality of the writing, and overall quality of the study.

publishers, and the Publons platform:

- [Preparing a Referee Report: Guidelines and Perspectives](#)
- [How to Write an Effective Referee Report and Improve the Scientific Review Process](#)
- [A Checklist for Reviewing a Paper](#)
- [Elsevier how to review guidelines](#)
- [Clarivate how-to-write a peer review](#)
- [Wiley step-by-step guide](#)
- [PLOS Peer Review Center](#)

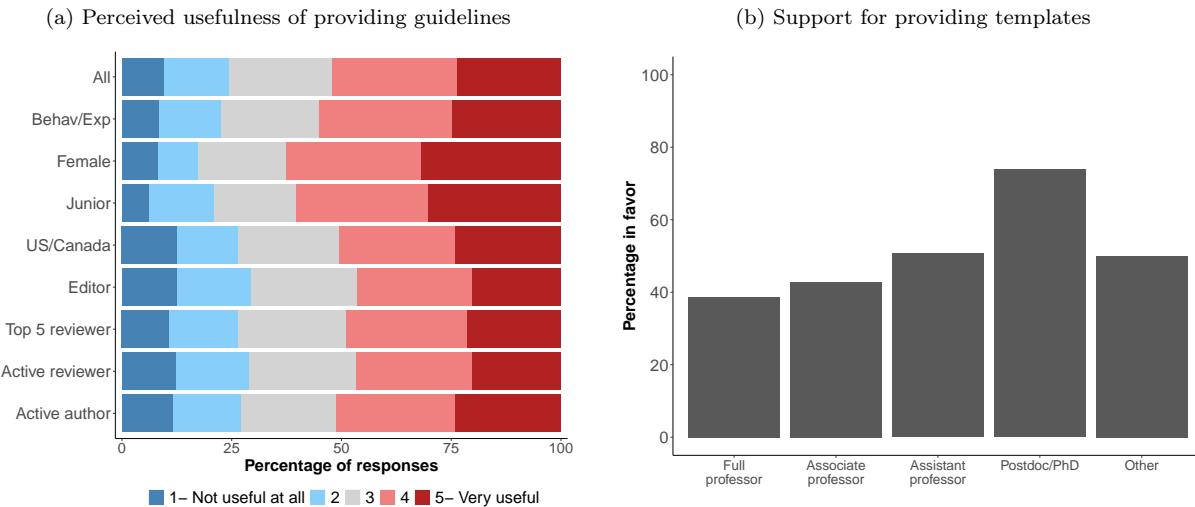
Regardless of the form that guidelines take, journals might want to make them accessible both on their website and in the editor's invitation email. Within the body of the email, editors could also offer advice tailored to specific manuscripts and referees, as necessary. This could be especially useful when editors want reviewers to comment on something based on their particular expertise.

3.3.3.2 Offering doctoral training Although writing referee reports constitutes an important professional activity, researchers receive no training on how to write a good review. As a result, reviewers mostly have to learn on the go. Setting the right standards in the early stages of a research career through appropriate training could have non-trivial consequences on the quality of the peer reviews delivered by a researcher. In this spirit, we asked our respondents to rate the usefulness of providing doctoral training for improving the quality of reviews. As can be seen in Figure 20, the majority of respondents would find it useful to provide doctoral training. Unsurprisingly, support is stronger among early-career researchers.

Although under-utilized, we note that various forms of training already exist. To help novice reviewers, many journals, publishers, and peer review platforms offer online resources, including short training videos and more developed online courses. Below we provide a non-exhaustive list of resources:

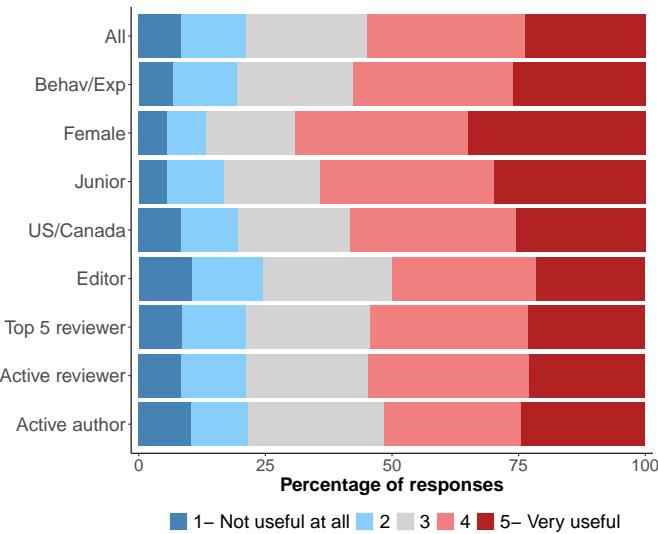
1. [Web of Science Academy course](#) (requires an account): Practical course during which students practice their peer review skills under a designated supervisor (e.g., the PhD advisor).
2. [Elsevier Research Academy course *Fundamentals of Peer Review*](#) (5+ hours)

Figure 19: Views on the provision of guidelines and templates for writing reports



Notes: Panel (a) is based on responses to question Q6, item i; Panel (b) is based on responses to question Q8. N = 1,459 for both panels.

Figure 20: Views on doctoral training



Notes: This figure is based on responses to question Q6, item ii. N = 1,459.

3. *Nature masterclass Focus on Peer Review*: (3-4 hours)
4. *Quality and Peer Review workshop offered by Sense about Science*: Free live workshop designed for STEM and social science early-career researchers.
5. *PLOS “How to review a manuscript” video* (5-min video covering essentials).

We note that although multiple online resources are available, it is unclear whether providing such resources makes a difference in the quality of reviews.

Another way of delivering the necessary training would be to more frequently **use referee reports as a mode of assessment in graduate school**. For instance, students could be tasked with writing a review of a recent working paper as part of their evaluation for a course; student reviews could be discussed in class and possibly transmitted to the authors. This exercise would help students understand what makes a good paper and better grasp what the important criteria for publication are. The **practice of co-reviewing** could also be expanded. PhD students, especially at a more advanced stage of their degree, could learn a lot from co-reviewing papers with their supervisors. Co-reviewing presents the advantage of students getting hands-on and practical experience with the peer review process. But one danger of co-reviewing is that early-career researchers might be deprived of the recognition they deserve without an appropriate system in place. Some platforms such as Publons have a collaborative reviewer feature (see [here](#)). While relatively uncommon in economics, we note that this practice is more frequent in other disciplines. In a survey primarily taken by academics from the life and physical sciences, 73% of respondents reported participating in co-reviewing, while half had also ghostwritten a report ([McDowell et al., 2019](#)). Co-reviewing is explicitly permitted by several scientific journals, such as *PNAS* and *Development*, as long as the work of collaborators is appropriately credited ([PNAS-1-b, DEV-1](#)).³⁷

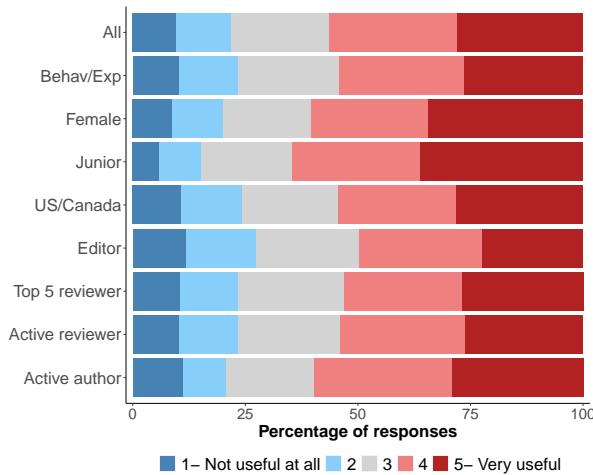
Regardless of the mode of training, it will be important to provide evidence of its effectiveness, as prior research on the topic is somewhat limited. Evidence from randomized control trials assigning interactive workshops and self-guided trainings shows little positive benefits ([Khoo, 2018](#)). [Houry et al. \(2012\)](#) assigned

³⁷ *PLOS ONE* is particularly encouraging of this practice, indicating in its guidelines that “co-reviewing is a great way to gain peer review experience under the mentorship of an experienced reviewer and we encourage this collaboration. If you had help completing the review you must share your collaborator’s name with the journal when you submit the review [...]. Be careful not to include your collaborator’s name in the text of the review itself. Competing interests and confidentiality policies apply to all reviewers” ([PLOS-3](#)).

high-quality reviewers to new journal referees to discuss their first three referee reports for a medical journal by phone or email; no improvement in quality was found for the first three or subsequent reports. While these findings might be discouraging, economics departments could experiment with their own programs. In particular, mentoring might be more effective within the context of the supervisor-student relationship, which is characterized by more frequent contact, more active collaboration, and closer familiarity.

3.3.3.3 Providing feedback to reviewers Reviewers receive virtually no feedback on the quality of their reports. While some journals deliver awards to recognize excellence in refereeing, the criteria are often vague and include other (more easily measurable) dimensions than quality, such as timeliness or number of completed reports. On rare occasions, an editor might directly email a referee to acknowledge a helpful report, but without necessarily explaining what made it helpful. Reviewers learn even less often what they need to improve or could have done differently. Editors could contact referees more frequently to **provide informal feedback**. Alternatively, a more formal evaluation could be put in place, with **editors grading peer reviews** according to a specified list of criteria; perhaps a yearly report could be automatically generated based on these assessments and awards could be delivered to the best reviewers. As shown in Figure 21, respondents would view favorably a grading system combined with awards as a way of improving the quality of peer reviews. We note that the enthusiasm for this proposal diminishes with experience, but remains high overall.

Figure 21: Perceived usefulness of grading and rewarding high-quality reports

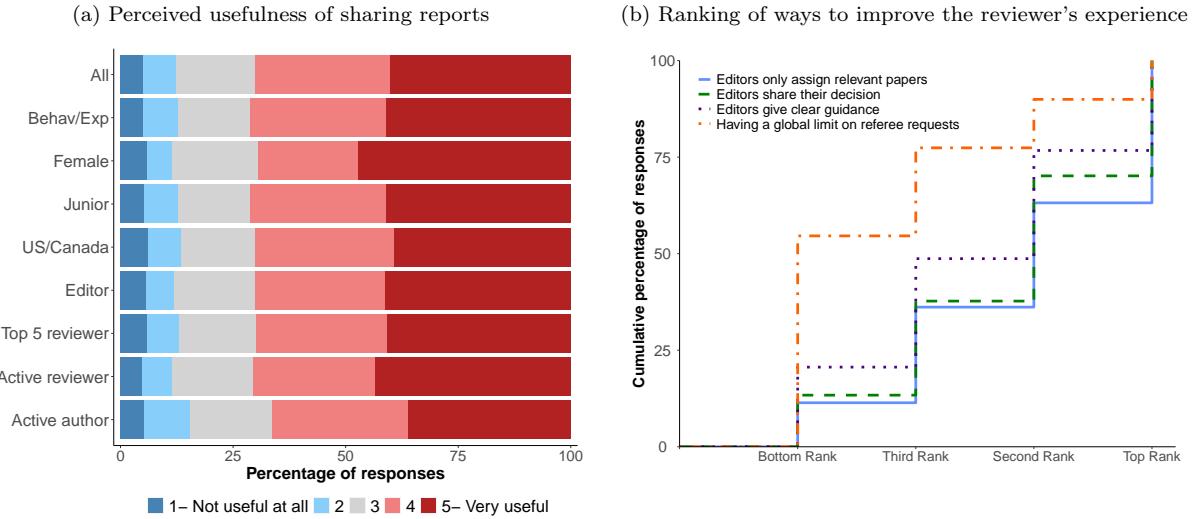


Notes: This figure is based on responses to question Q6, item vi. N = 1,459.

Finally, another useful way of delivering feedback is for editors to **systematically share the decision letter and reports** with all referees. While this is already a common practice, we note that not every editor necessarily does it. Figure 22 shows widespread support for this policy. First, regardless of their level of experience (and any other dimension of heterogeneity considered), respondents believe that sharing reports can improve the quality of peer reviews (Panel (a) of Figure 22). Second, this policy was ranked highly in terms of its potential to improve the experience of reviewers, right after editors assigning only papers relevant to the reviewer's research (Panel (b) of Figure 22). There is also little heterogeneity in support of this policy relative to others (see Appendix Figure E.14), suggesting wide consensus.

3.3.3.4 Asking reviewers to abide by a code of conduct Upon accepting and/or submitting a review, **referees could be asked to consent to an honor code**. In addition to preserving the confidentiality of the review and reporting any conflict of interest, reviewers could be required to abide by a set of principles for ethical peer review. These principles could be reinforced by editors upon sending their review invitation email. Beyond that, editors could play a more pro-active role in challenging, and maybe **sanctioning, inappropriate or inaccurate comments made by a reviewer**. For instance, the editor could step in to ask a reviewer to change the tone of a comment made in the report. In case the report was

Figure 22: Importance of sharing editorial decisions and reports with reviewers



Notes: Panel (a) is based on responses to question Q6, item viii. Panel (b) is based on responses to question Q37. N = 1,459 for Panel (a) and N = 1,474 for Panel (b).

clearly unhelpful, the **editor could decide not to transmit it to the authors**. To limit the arbitrariness of such decisions, **clear policies about editing reviewers' reports** would need to be put in place.³⁸

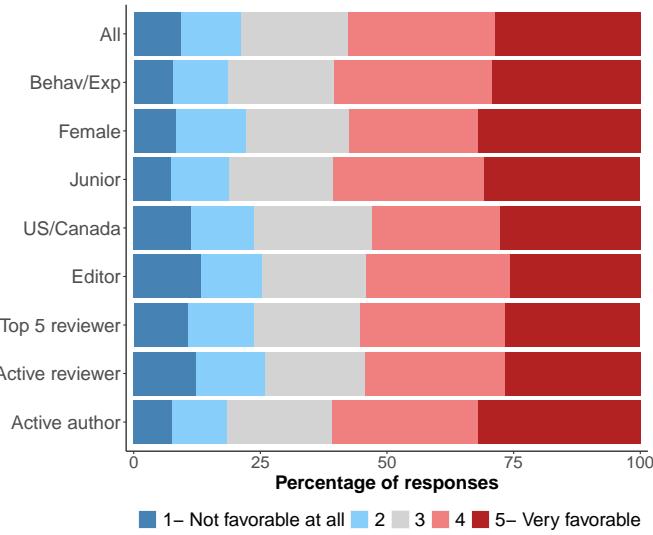
3.3.3.5 Allowing the authors to respond In the event of inaccurate or aggressive comments from reviewers, **authors could be given the right to respond**. At the moment, we note that very few economics journals have a formal appeal process in place: for example, the *AER* and *Econometrica* mention the possibility of appealing but do not give instructions on how to do it and appear to discourage such attempts (Table 4). Journals with formal appeal processes include the *JPE*, *EJ*, and *PLOS ONE*.³⁹

In case of a rejection, the norm is to not challenge the decision made by the editor or the views of the referees, although this is not always followed in practice. Noting this point, we asked respondents how favorable they would be to a policy allowing the authors to submit a single response to the referees and the editor. We specified that “*the referees would be under no obligation to provide additional comments; a ‘cooling period’ could be required before the authors can send their response. There would be no guarantee of the referees taking this rebuttal into account, and the decision would be final after the comment period.*” Figure 23 shows that respondents would be overall supportive of such a policy. Support appears somewhat stronger among active authors, junior researchers, and behavioral/experimental economists. On the other hand, support from editors and active reviewers is only mildly weaker; this is despite the potential time and logistical costs implied by such an appeal procedure, which would especially affect those most involved in the peer review process.

³⁸For instance, the guidelines of *PLOS ONE* indicate that “the editors and PLOS staff will not change any reviewer comments that are intended for authors to read, except with reviewer approval prior to the decision letter being sent. For example, we may request changes if language is deemed inappropriate for professional communication, or if the comments contain information considered confidential, such as competing interest declarations” (*PLOS-4*).

³⁹*EJ* states that “authors have the right to appeal any decision taken on their manuscript at any stage. An appeal letter should be sent to the editorial office (*ej@res.org.uk*) and should explain clearly why the author disagrees with the decision, including detailed responses to any reviewers’ comments” (*EJ-2-a*). *PLOS ONE* has a similar policy: “authors may submit a formal appeal for rejected submissions. Appeal requests must be made in writing to *plosone@plos.org* with the word “appeal” in the subject line. Authors must provide detailed reasons for the appeal and point-by-point responses to the reviewers’ and/or Academic Editor’s comments” (*PLOS-5*).

Figure 23: Attitudes towards a formal appeal procedure



Notes: This figure is based on responses to question Q18. N = 1,459.

3.4 Reviewing process and decision times

3.4.1 Delays in peer review

3.4.1.1 The slow review process in economics The peer review process in economics stands out for its long processing times, both as compared to other disciplines and to its own past performance. According to an interdisciplinary analysis of (self-reported) review times for submitted papers, economics claims both the highest first response times and the highest total review times out of the 10 fields compared (Huisman and Smits, 2017).⁴⁰ Similarly, Hadavand et al. (2021) find that top journals in economics compare poorly to those from the hard sciences and other social sciences. From 1970 to 2000, lengthening review times in economics were the norm (Ellison, 2002a,b). Efforts to reverse this trend have sometimes yielded positive results at top journals,⁴¹ but turnaround times appear to have lengthened further over the 2012–2020 timeframe at the four top 5 journals the authors obtained data for (Hadavand et al., 2021). Furthermore, any actual or future gains might be fragile in the face of competing pressures that have emerged. First, submission volumes to journals appear to have increased sharply (Card and DellaVigna, 2013), and economics papers have become longer and more technical in recent years (Schwert, 2021); we will return to these two points at the end of the report (see Section 4.1.4). Second, efforts at improving the transparency and reproducibility of economics research might start to exert more pressure on review times. For example, the widespread adoption of pre-analysis plans (PAPs), as described by Olken (2015), could increase the number of documents that need to be reviewed by editors and referees.⁴² Over the past few years, the AEA has also implemented a policy requiring their journals to collect author-provided data and code to reproduce the findings of submissions prior to acceptance, with some exceptions made for restricted datasets (AER-10). Facilitating this massive reproduction effort is within the mandate of the AEA Data Editor, who noted that it could easily become another material contributor to review delay without adequate preparation (Vilhuber, 2019). The *EJ* has adopted a similar replication policy (EJ-2-b), and more journals are expected to follow suit.

⁴⁰The fields were: general interest, natural sciences, engineering, mathematics and computer science, medicine, public health, psychology, economics and business, social sciences, and the humanities. We note that the reported review times might not be representative of review times in the general population of submissions e.g., if authors are more likely to report a lengthy review time after a bad experience. However, the data should still be useful as a cross-sectional comparison between fields.

⁴¹At the *AER*, total review duration for accepted papers declined from 62 weeks in 2008 to 36 in 2010, and has remained stable since then (AER-5-a). A similar improvement is observable at *REStud* over this timeframe (RES-1, RES-2, RES-3-a), while *Econometrica*'s performance seems stable (ECMA-1-b, ECMA-2-b). The *QJE* and *JPE* have not published sufficient data to comment on trends over this timeframe.

⁴²Recent evidence suggests that PAPs often contain insufficient information for reproducibility or may not be strictly followed in the final version of the paper (Ofosu and Posner, 2021; Abrams et al., 2020). These limitations emphasize the importance and burden of third party verification, in the absence of proper standardization or improved efforts by authors.

3.4.1.2 Delays in first response times Since most first submissions yield a rejection and first response times are a material contributor to the review duration of accepted papers, the initial submission stage is of primary importance. Ellison (2002b) estimates that increases in first response times were responsible for about 25% of the slowdown in review duration pre-2000. We again note recent improvements at some top journals, but the performance of economics still pales in comparison to that of top science journals. For instance, the median first response time at the *AER* for papers sent to review was about 13 weeks as of 2020, while it was 12-16 weeks at the *EJ* (AER-1-d, EJ-3-a). By comparison, *Nature* and *Science* each report medians of about 6 weeks (Nature-4, SCI-3).

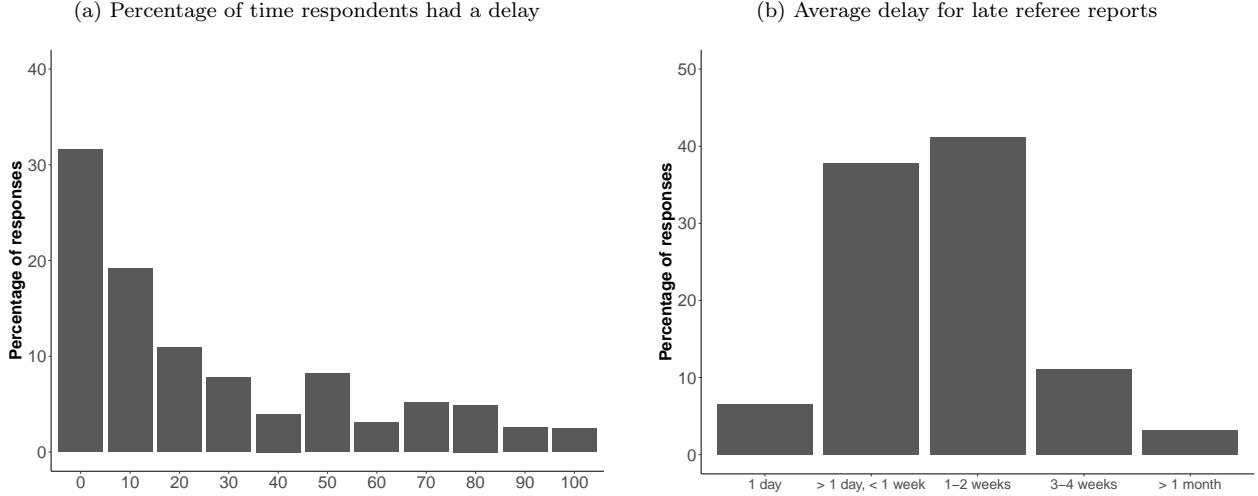
First response times can be decomposed into four stages:

- Stage 1. **Delays in the initial editorial decision to desk reject or not.** In their most recent statistics, the *Journal of the European Economic Association* (JEEA) and JPE each report an average of 9 days for desk rejected manuscripts to receive a first response (JEEA-1-a, JPE-1-a). As such, this step might contribute 1-2 weeks to total first response delay.
- Stage 2. **Delays due to frictions in matching submissions to reviewers.** Acceptance rates for referee invitations were between 61.1% and 72.0% in the Chetty et al. (2014) study (Table 5), and 82.7% in our survey. This implies editors need to send 4-5 invitations to secure 3 referees. An editor at a science journal noted that most invitation acceptances come within 2 days, but rejections or non-responses can take longer (Berquist, 2014). Assuming that editors contact new potential referees up to one week after the first requests, 1-2 weeks may sometimes pass before enough acceptances are obtained.
- Stage 3. **Delays in receiving all completed reports.** The turnaround times requested from referees range between 4 and 8 weeks at the top 5 journals (see Table 4). Chetty et al. (2014) finds that the median unpaid referee submits their report about a week after the requested deadline, while the median paid referee tends to be on time. But since editors typically want to receive all referee reports before making a decision, the worst-performing referee matters more than the median one. To estimate the lateness of the slowest referee, we conduct simulations calibrated to different data sources, as discussed below. We note that an additional source of variability comes from the nature of the submission: more complicated and borderline-quality manuscripts likely take more time to review and are assigned additional referees. For all these reasons, Stage 3 is an essential contributor to first response time, especially once lateness is factored in.
- Stage 4. **Delays in the editor's final decision after all reports are received.** The length of this final stage usually depends on the status of the manuscript: among all papers sent to referees in 2020, the JEEA reported that first response times for R&R decisions were approximately 8 weeks longer than those for rejection decisions (JEEA-1-b), although it is not clear how much of the additional delay is due to Stage 4 rather than Stage 3. Welch (2014) finds that manuscript status is strongly predicted by referee recommendations, with a rejection rate of 97.1% for manuscripts that receive unanimously negative referee reports. In cases of referee disagreement or positive recommendations, decisions are less predictable,⁴³ especially in the absence of a clear decision rule for handling reports. *Energy Economics* is one of the few journals that publicly acknowledges having such a rule, namely that manuscripts “generally need two or more positive review reports” to receive an R&R (ENEC-2). In any case, substantial internal discussion is likely needed before a decision to issue an R&R can be made. Indeed, the editorial policy at the *AER* requires that the Coeditor in charge seeks input from a second Coeditor when considering an R&R or recommending revisions (AER-2-b). As marginally rejected papers might also require additional work, the more complicated cases may end up taking several weeks longer in Stage 4 than the simplest ones.

Using our survey data and an external data source, we investigate below the incidence of referees returning their reports late in Stage 3 and evaluate its overall impact on decision times.

⁴³When referees disagree about a manuscript’s merits, Welch (2014) finds rejection rates in the range of 42.4% to 74.8%; and when both recommendations are positive, rejection rates are between 0.0% and 14.2%. We note that Welch (2014) focuses on reviews with either one or two referees.

Figure 24: Delays in returning peer reviews

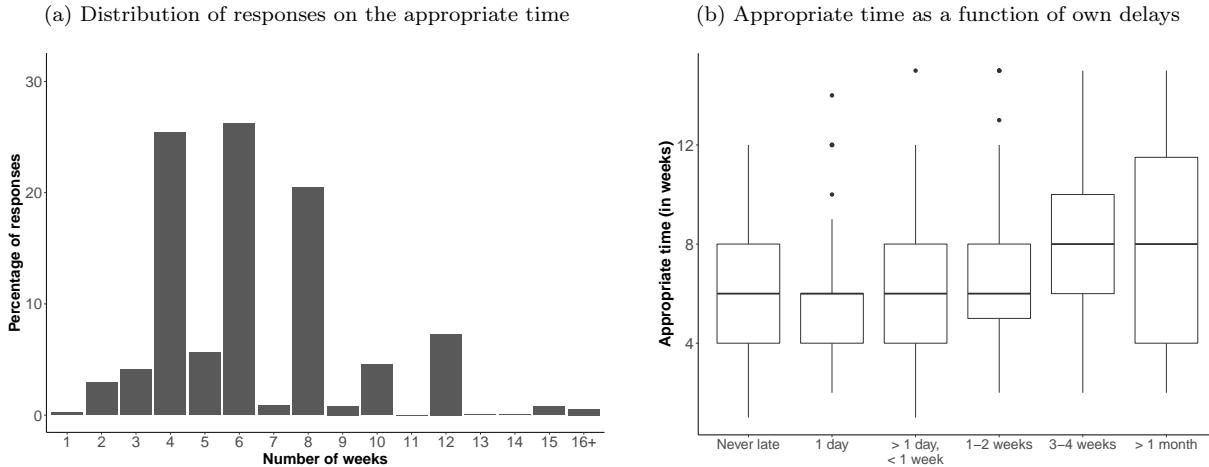


Notes: Panel (a) is based on responses to question Q26. Panel (b) is based on responses to question Q27 (only displays respondents who reported they had been late). N = 1,483 for Panel (a) and N = 1,014 for Panel (b).

3.4.1.3 Delay patterns in the survey To assess the frequency and severity of the Stage 3 delays, we asked respondents to estimate the percentage of the time they are late in returning a report and their average delay when late. Panel (a) of Figure 24 shows that only a third of respondents are always punctual, while about a quarter are late at least 50% of the time. The delays of late referees are also non-trivial. The median delay is 1–2 weeks and the 90th percentile of the distribution is late by 3 weeks or more (Panel (b) of Figure 24). We further note that due to sample selection and potential under-reporting, these numbers likely provide a lower bound for actual delays. Unsurprisingly, the two are strongly linked: reviewers who are late more frequently also take more time to return their reports when late (Panel (a) of Figure E.15).

These delays are particularly suboptimal in light of the time respondents think is appropriate to complete a review. As shown in Panel (a) of Figure 25, over 50% of respondents think that a report should be returned within 6 weeks, and among them, over half would favor a deadline of 4 weeks or less. At the same time, there is heterogeneity in views, with about 20% of respondents who would find 8 weeks appropriate and another 13% who would prefer 10 weeks or more. The turnaround times considered reasonable by junior researchers,

Figure 25: Views on the appropriate time length to submit a referee report

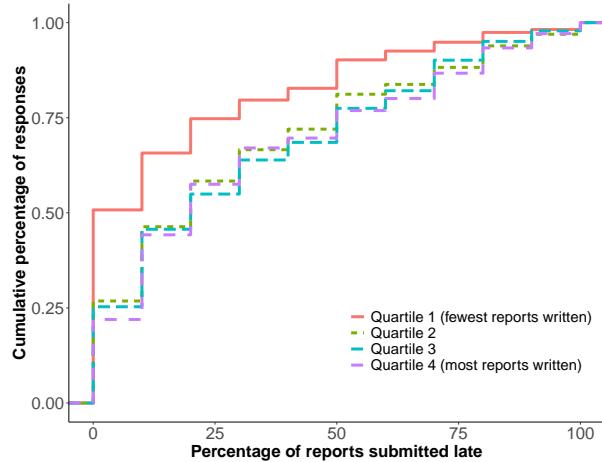


Notes: Both panels are based on responses to question Q15; Panel (b) also uses responses to question Q27. N = 1,467 for Panel (a) and 1,455 for Panel (b).

more active reviewers, and researchers with more publications are on average shorter (see Appendix Figure E.16). A natural question is whether late reviewers have different standards for when reports should be returned. Panel (b) of Figure 25 shows that this is true only to some extent. While reviewers who typically face relatively minor delays (up to 1-2 weeks) have virtually the same views as those who are never late, those with severe delays (at least 3-4 weeks) tend to think that deadlines should be longer (also see Appendix Figure E.16). At the same time, there is considerable diversity in views, with interquartile ranges of 4 weeks or more. Many reviewers are late even relative to the deadline they would find most appropriate.

Besides having different standards, respondents face different constraints on their time (which also likely affect those standards). To better understand the relationship between delays and time constraints, we first examine whether respondents who handle more reviews are more likely to be late. Figure 26 shows that the relationship is actually fairly weak; only respondents who are solicited very infrequently are significantly more on time. On the other hand, regression analyses show that researchers based in the US/Canada and those who review more frequently for top 5 journals are more likely to be late, while junior researchers and authors who make more frequent submissions are less often late (see Panel (b) of Figure E.15). In addition, respondents who write relatively more reports compared to what they deem reasonable face more frequent delays. This suggests that late reviewers are overcommitted. Another sign of this overcommitment is that respondents with the most frequent and largest delays do not turn down review requests at a much higher rate. Although significant, the correlation coefficient between delay frequency and rejection rate of requests is very low (Spearman rho = 0.08, p < 0.01). There is also no clear relationship between length of delay and rejection rate (see Figure E.17). While we cannot fully speak to the nature of the observed delays, these findings reinforce the results of Section 3.2 about the misallocation issues faced by our current system.

Figure 26: Percentage of late reviews as a function of the number of reports written



Notes: This question is based on responses to question Q26. N = 1,483.

3.4.1.4 Estimated impact on total review time Regardless of the precise nature of these delays in returning reports, we note that they are considerable, especially if the editor waits for the latest referee to return their review before deciding. With this in mind, we provide some estimates of the total delay time that is attributable to Stage 3 by estimating the review time of the latest referee on a given assignment. Since our sample is selected and there is likely under-reporting bias, we perform this estimation exercise using various survey subsamples and compare our results with estimates using observational data from *JFE*.

More precisely, we conduct four simulations, each of which is made available on our OSF page (<https://osf.io/eczkv/>). We randomly draw 10,000 groups of 3 referees from our survey respondent list for Simulation 1 (S1)⁴⁴ and from a population of 1,000 referee agents calibrated to *JFE* data for Simulation 2

⁴⁴We collected data on respondents' self-reported number of reports written, their probability of submitting a report late, and their typical delay length. A referee's chance of selection is proportionate to their volume of reports written.

(S2).⁴⁵ Simulations 3 (S3) and 4 (S4) are identical to S1 except for being conducted on subsets of respondents in behavioral/experimental economics and referees for top 5 journals, respectively. In S1, S3, and S4, we also conduct a second draw to assign respondents to be late or on-time according to their reported probability; if late, we report their estimated delay, and if not, we report no delay. This is not necessary in S2, as the referee agents are constructed to be always on-time or late by a fixed amount of time. For each draw, we find the slowest referee by taking the maximum delay time among the 3 referees.

In S1, we find that the median referee is 0 days late in contrast with the results for unpaid referees from Chetty et al. (2014); see Table 5. This could be evidence of self-selection or measurement error in our sample. Among the referees who were the slowest in their draw, the median and 90th percentile performances were 4 days and 24.5 days late, respectively. In S2, the median referee is also on time, but since *JFE* referees are paid, this result is consistent with the Chetty et al. (2014) results for paid referees. The median and 90th percentile performances of the slowest referees increase to 17.5 and 66.5 days late, respectively. The results from S3 and S4 were almost identical to S1, except that the median latest referee took 10.5 days in S4. Based on these results, we expect that the slowest referee could delay Stage 3 by 0.5-2.5 weeks in the median draw, but this could become a month or more in draws at the 90th percentile.⁴⁶ Accounting for the latest referee pushes the estimated interval for Stage 3 from the requested range of 4-8 weeks to 5-12 weeks.

3.4.1.5 Delays from additional review rounds Since most submissions are not accepted, our focus has been on first response times. But before we conclude, we will also address the experience of authors whose manuscripts end up being involved in multiple rounds of review. This stage can contribute materially to total review duration, sometimes in great excess of the initial round.

Stage 5. **Delays due to multiple or prolonged review rounds.** Conditional acceptance offers in the first round are far from the norm in economics (Hadavand et al., 2021), with Huisman and Smits (2017) reporting an average of 2.16 total rounds for accepted papers. Even if the typical paper only needs to be revised once,⁴⁷ this might more than double the total review duration: in 2020, the median decision time for a resubmission at *Econometrica* was up to a month longer than the first response time (ECMA-2-c). Delays in this stage are highly variable but potentially enormous in magnitude. Hadavand et al. (2021) argue that the primary driver of increased turnaround times at the slowest decile of publications is author revision time. In their sample, authors spend 26 months revising their manuscripts in the 90th percentile of cases, as opposed to only 10 months in the median case. At the *AER*, the amount of time authors spend revising their manuscript for resubmissions has skyrocketed in recent years: the number of weeks between submission and acceptance increased from 87 to 104 from 2008 to 2020, even as total time under review fell by 25 weeks over the same timeframe (AER-1-b, AER-5-b).

Frustration and disappointment caused by multiple review rounds was a recurring topic in many of our survey responses. Among these, several suggested that the existence of multiple revision rounds is evidence that editors and referees are not doing their job: in their view, the role of these parties should be to evaluate if a manuscript is essentially publishable in the initial round, with only minor and plausible revisions requested before publication. Instead, the trend appears to be one of referees making increasingly demanding requests for the revision of manuscripts before publication (Berk et al., 2017). A related issue noted by 26 of our respondents is the ambiguity created by a lack of editorial direction about revisions. Since it is often the case that multiple referees offer extensive (and sometimes conflicting) advice, with little regard for feasibility, the absence of clear editorial direction about which of those revisions are actually essential could be a leading contributor to the explosion in revision times. Finally, some also noted that their manuscripts have been rejected even after implementing substantial and time-consuming revisions. One possible explanation for these cases is a lack of initial editorial engagement with the revisions proposed by referees: since it is nearly costless for editors to impose that authors implement any (or all) of the changes requested by referees,

⁴⁵Figure 3b from Schwert (2021) shows the distribution of referees by turnaround time and whether they received the incentive payment. We use this information to construct our referee population of 1,000 agents.

⁴⁶Editors might decide without the last referee in cases of extreme lateness, so a month may be a reasonable cap.

⁴⁷This is usually the case at the *JFE* (Schwert, 2021), but around half of *AER* R&Rs go through 2+ revision rounds before acceptance. With 2012-2013 data from three of the top 5 journals that they were able to obtain data for, Hadavand et al. (2021) find that 51% of papers go through two review rounds, and 22% go through three rounds.

editors may not adequately consider whether the value of proposed revisions justifies the cost. Meanwhile, authors are expected to implement all revisions requested by editors, even at substantial cost and without any guarantee that the R&R will be converted into a publication.

3.4.2 Reducing delays

In this section, we outline proposals that are potentially responsive to the issues identified in each stage. Each proposal is linked to a discrete stage (or discrete stages). But it might be best to evaluate some of them jointly since, for example, implementing certain early-stage proposals could have downstream effects on the type of manuscripts that proceed to subsequent stages. Similarly, authors might change submission behavior if they anticipate a late-stage proposal affecting their publication odds. As such, it is important to keep in mind how each proposal might affect the entire cycle and whether some proposals might best be paired to mitigate the potential for unintended consequences. We proceed in chronological order by stage.

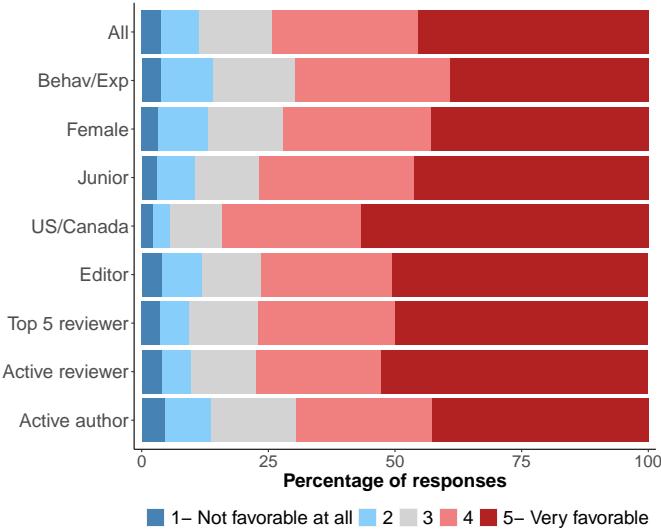
3.4.2.1 Changing incentives around submissions (Stage 0) As mentioned elsewhere, authors face little penalty for “shopping around” between journals and submitting the same paper many times. There is no public record of rejections, and submission fees for top journals sometimes come with need-based discounts and refunds for desk rejections (Table 4). Cotton (2013) suggests that **optimally set submission fees, response times, and other submission costs** could reduce the flow of unfit manuscripts to journals. Indeed, Zheng and Kaiser (2015) estimated the demand elasticity of submissions to economics journals and found that a \$100 increase in submission fees would decrease submissions. This strategy would ultimately favor well-resourced academics, so an alternative would be for journals to share submission records and agree to **impose a cap of (say) 3 submissions** of a manuscript to highly-ranked journals per year, to incentivize authors to choose carefully.⁴⁸

3.4.2.2 Non-exclusive journal submissions (Stage 0) In some disciplines, authors are allowed to make **contemporaneous submissions of the same manuscript to different journals**. For example, the *Harvard Law Review* allows authors to submit their manuscripts to other journals during the review process, although it requests that they wait at least 10 days before doing so and generally discourages the practice (**HLR-1**). Allowing authors to have their submissions considered at multiple journals might incentivize journals to improve their turnaround time performance. Moreover, this policy strongly diminishes the time component of submission costs, which could be useful for researchers racing against the tenure clock. But allowing authors to “shop around” at several journals at once could exacerbate the current trend of growing submission volumes and lead to many superfluous review efforts.

3.4.2.3 Increasing the rate of desk rejection (Stage 1) One way to speed up the review process could be to **increase the desk rejection rate** of papers that have a low chance of obtaining an R&R (Cherkashin et al., 2008). At the top 5 journals, desk rejection rates in 2019-2020 ranged from 37% at the *AER* to 64% at the *QJE* (Table 4). In line with this proposal, the desk rejection rate at *JEEA* climbed from 40% in 2016 to 62% in 2020 (**JEEA-1-c**). In the face of rapidly rising submissions (from around 880 in 2016 to 1,231 in 2020), the *JEEA* editors partially attributed their ability to maintain a reasonable first response time to the higher desk rejection rate. Meanwhile, the *JFE*’s desk rejection rate from 2006 to 2020 was only 14% (Schwert, 2021) and the *Review of Finance*’s desk rejection rate has fallen from 35% in 2017-2018 to 24% in 2019-2020 (**ROF-1**). We note that statistics on desk rejection rates are not available for many journals. To gather some additional data, we asked respondents in our follow-up survey about the number of submissions they had made and the number of desk rejections they had received in 2020-2021. The average desk rejection rate reported by these respondents ($N = 114$) was 27.8%. This rate is unlikely to be representative of the broader population, however.

⁴⁸Managing incentives at Stage 0 might be especially important when implementing policy reforms at later stages. If total turnaround time decreases, this lowers the time cost imposed on the author for each marginal submission. Increases in the volume of submissions after reducing time costs might end up eliminating the progress made on turnaround time, so these proposals could assist with holding submission incentives constant while making changes upstream. See Azar (2005) and Azar (2006) for more discussion and ideas about managing this stage.

Figure 27: Views on desk rejections



Notes: This figure is based on responses to question Q16. N = 1,459.

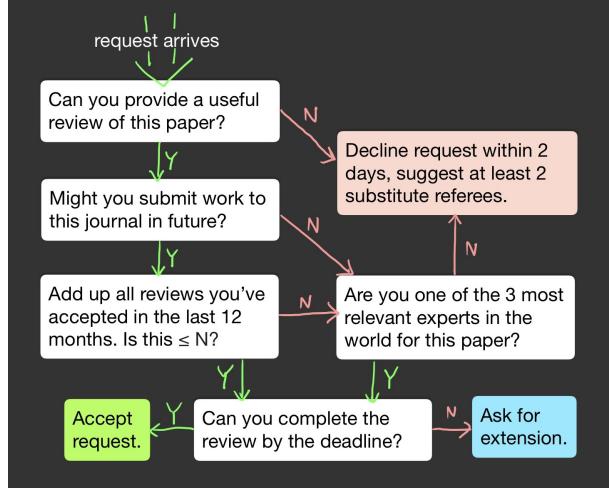
Decisions to desk reject are not always straightforward, and some might be concerned that editors employ unfair heuristics in choosing which papers to desk reject. The evidence about demographic bias in desk rejections is mixed: Clain and Leppel (2017)'s study found that submissions from first-time authors, solo authors, and female authors had higher desk rejection rates, while authors with prior publications at the journal and assistant professors had lower rates.⁴⁹ Results from Schwert (2021) about female authors at the *JFE* support this, while Card et al. (2020b) find that female authors are desk rejected less often than male authors, though at a higher rate than subsequent paper citations would suggest. To increase the desk rejection rate while avoiding actual or perceived incidence of bias, journals could enumerate explicit criteria for these decisions and indicate which ones are applicable to a given manuscript. Journals could produce statistics about the percentage of manuscripts that are desk rejected due to each criterion.

Regardless, we note that our respondents are largely favorable to the practice of journals desk rejecting manuscripts in general, with somewhat more support from respondents based in the US/Canada (Figure 27). However, we did not ask respondents whether current desk rejection rates are too high/low or what the optimal desk rejection rate is. One conjecture is that views might be more heterogeneous regarding the appropriate level of this parameter at different journals. To simulate the impact of desk rejections on referee load, we refer the reader to the calculator we introduced in Section 3.2.

3.4.2.4 Speeding up responses to review invitations (Stage 2) Journals could encourage reviewers to adopt a **clear decision rule for whether to accept or reject a peer review request**, such as the one depicted in Figure 28 (courtesy of Shengwu Li). With a clear decision rule in mind, the decision-making process of reviewers could be greatly facilitated. For instance, it is likely that reviewers would feel less guilty for turning down a request, and thus procrastinate less to make such a decision, if they had a clear procedure for deciding. While the exact decision rule and specific parameters might vary from referee to referee, journals could **provide information about the social norm as a reference point**. For instance, one reasonable target time for a response could be to accept or reject the invitation within 48 hours of receiving it. Editors could keep track of the speed with which reviewers respond to requests (whether positively or negatively). In addition, reviewers could be more systematically encouraged to provide names for alternative reviewers. Finally, anecdotal evidence suggests that review invitations sometimes do not make their way to the intended recipients because the email addresses are no longer in use. One practical suggestion is therefore for reviewers to **regularly update their account information, possibly creating a separate (and non-institutional) email account** for handling requests from journals. However, even

⁴⁹With statistical significance only at the 10% level for some variables in some specifications.

Figure 28: Finite automaton procedure for deciding on a peer review request
 (Courtesy of Shengwu Li)



when the email is valid, a potential reviewer might fail to reply due to being away; in this case, setting an out-of-office reply would greatly facilitate the job of the editor. [Making a note of holiday/sabbatical times](#) in the editorial system could also prevent some requests from being sent.

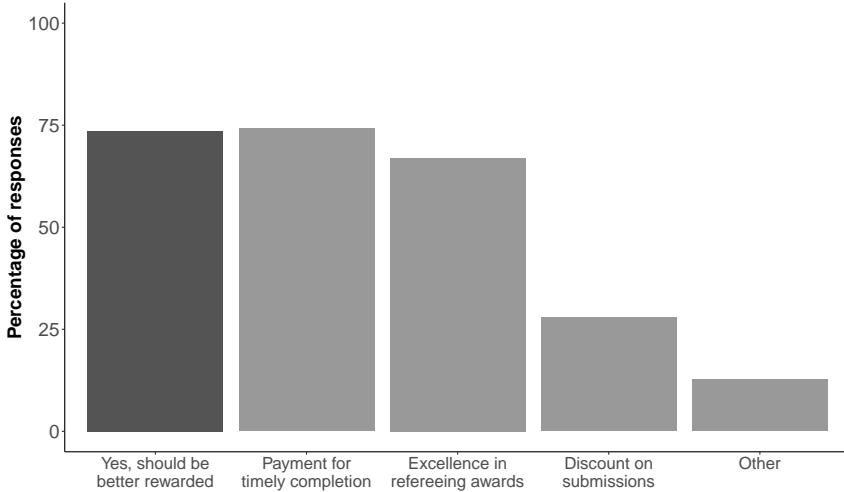
3.4.2.5 Speeding up and improving the quality of referee-submission matches (Stage 2) [Automating some aspects of referee search and assignment](#) could help by quickly identifying qualified and motivated referees. [Mahmud et al. \(2018\)](#) proposed an algorithm for pairing authors and referees based on the manuscript's citations. Some caveats include that a referee with conflicts might be chosen by the algorithm or that authors could strategically choose their citations to get a preferred referee. If data about author and referee networks is obtained as suggested in our centralized platform discussion, the algorithm could be trained to avoid the first issue, although some conflicts are not easily observable. With the latter caveat in mind, these automated tools might be better used for decreasing the search costs of editors than for deterministic decision-making.

3.4.2.6 Soliciting more expedited reviews for clear cases (Stage 2.5) More papers could be [summarily rejected without a full review if one or several reviewers conclude that a paper should have been desk rejected](#). In this case, reviewers could simply submit a paragraph within a few days of accepting the request to explain the rationale. While this option sometimes exists, editors rarely make it explicit in their invitation email. Exceptions include invitations from the *Economic Journal* and the *JEEA* (received by team members), which mention the option of quickly submitting a short referee report or a letter in cases of clear-cut rejection. We note that without clear guidance, reviewers may feel obliged to write an elaborate report detailing their evaluation in order to meet professional standards.

3.4.2.7 Requesting shorter turnaround times (Stage 3) [Chetty et al. \(2014\)](#) find that [changing the requested turnaround time](#) from 6 weeks to 4 weeks reduces referee turnaround time by 12 days (Table 5), even without any change in financial or social incentives. The requested deadlines in Table 4 range from 4-8 weeks, so there appears to be room for experimentation; however, this strategy could have diminishing returns, especially if all journals impose shorter turnaround times.

3.4.2.8 Referee payment or charity donation (Stage 3) Out of the top 5 journals, only the *AER* and *JPE* offer referees a [financial incentive](#) (each \$100, or a two-year journal subscription at the *JPE* - see Table 4). A few journals in Table A.3 also pay, with finance journals like the *RFS* (\$200 - [RFS-2](#)) and the *JFE* (\$500 - [Schwert \(2021\)](#)) appearing to be the most generous, but most do not. The existing evidence

Figure 29: Incentivizing and rewarding refereeing



Notes: The dark gray bar is based on responses to question Q13. The light gray bars are contingent on answering “Yes” to that question and use responses to question Q14. N = 1,467 for the first bar and N = 1,080 for the other bars.

summarized in Table 5 suggests that **offering financial incentives** could be an effective way of reducing turnaround time. However, treatment effects appear to be heterogeneous, with most of the improvement in the Chetty et al. (2014) study coming from some referees now submitting their report immediately before the deadline. The on-time percentages seem similar for cash-incentivized referees in Chetty et al. (2014) (66%) and *JFE* referees (64%) (Schwert, 2021), so whether there is any benefit from paying \$100 vs. \$500 is unclear. In terms of possible downsides, some emphasize that referees are currently driven by a sense of professional obligation (Freeman, 2000), so payments might be unnecessary or counterproductive. Indeed, the lab experiment from Squazzoni et al. (2013) found negative results of payments on measures of quality and efficiency, citing the crowding out of intrinsic motivation as a potential explanation. We note that this contrasts with the finding from Chetty et al. (2014) that payments have no identifiable adverse effect on quality. For the level of payment considered, concerns around the crowding out of motivation might thus be limited.

Our survey results suggest that the practice of offering financial incentives could be popular. Among the 1,080 respondents who thought that referees would do a better job if better rewarded, about three-quarters thought that referees should receive a payment for timely completion (Figure 29). Researchers who make more submissions themselves are more likely to be of this view ($p < 0.01$; see Appendix Figure E.18). Anecdotal evidence suggests that some referees face difficulties claiming rewards via e.g., PayPal or physical check. To address this, journals could avoid direct cash transfers by **offering submission fee waivers or discounts**, as the *JPE* does (Table 4). However, as shown in Figure 29, submission discounts appear less popular and might therefore be less motivating. As an (unexplored) alternative, journals could let referees opt to have the money **directly donated to a charity**, which has been shown to reduce crowding out in other contexts (Johannesson and Mellström, 2008). Journals could also let payments be allocated towards funding research, perhaps in the form of grants to researchers in a particular field or as funding to conduct replications.

3.4.2.9 Non-monetary rewards and recognition for referees (Stage 3) Non-monetary rewards are often more effective at incentivizing recipients than monetary rewards of equal value (Kube et al., 2012). An example of a non-monetary reward is journals **acknowledging the names of all referees** who contributed reports in a given year: 3 of the top 5 journals do this (Table 4), although one study suggests that this could be ineffective or even counterproductive (Zaharie and Seeber, 2018). As an alternative or complement to this, the *AER*, *REStud*, and *QJE* also **acknowledge exceptional referees** who meet some quality or quantity threshold. To assess whether recipients find this recognition worthwhile, we scanned the CVs of referees

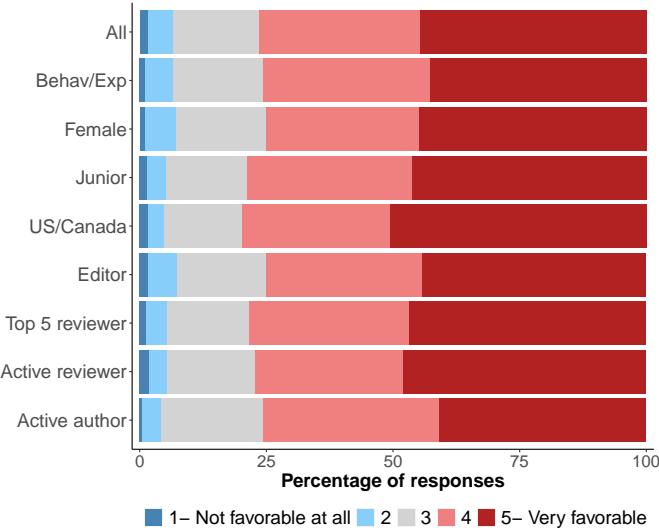
recognized in 2019 (as available) to see whether they mention the acknowledgment. The highest percentage of recipients who report the acknowledgment is for *REStud* (76.9%), while the share is smaller for the *AER* (54.4%) and close to zero for *QJE* (1.0%). One possible reason is that *REStud*'s acknowledgment is more exclusive than the *AER*'s (26 vs. 160 recipients), and the *QJE*'s recognition is not framed as a reward (i.e., it is given to all referees who wrote 4+ reports). These findings might reflect different subjective values for various types of acknowledgments. We also find favorable attitudes towards referee excellence awards within our survey population. As shown in Figure 29, among respondents who thought that more should be done to reward reviewers, two-thirds were in favor of excellence awards; support appears to be stronger among reviewers who write more reports annually ($p < 0.05$; see Appendix Figure E.18). **Certificates for refereeing** are another type of non-monetary award currently offered by multiple journals, including *REStud* and Elsevier journals such as the *Journal of Economic Theory* and *JFE*, through the Reviewer Hub. Kube et al. (2012) suggest that gifts requiring thought and effort induce stronger reactions than ones that do not, so standardized certificates might be of limited effectiveness. Finally, one proposal mentioned by 32 respondents is for **universities to recognize peer review contributions** in tenure, promotion, or hiring decisions. This proposal might have to be implemented in tandem with another one that increases the visibility of refereeing work, such as the creation of a centralized platform or with granting referees more ownership of their work (discussed in 3.5.3.7).

3.4.2.10 Negative incentives for late referees (Stage 3) Referee services are typically free, so withholding payment or imposing a fine is not usually possible. Alternative ways of implementing negative incentives have been proposed, including by **having the submissions of late referees sit on editors' desks for a length of time in excess of the referee's delay** (Hauser and Fehr, 2007). A reader comment on that proposal also suggested that **late referees could have their submissions assigned to other late referees**. These ideas could perversely disincentivize participation in the peer review system and might be ineffective, if tardy referees interpret the punishment as a price worth paying for their delays (Gneezy and Rustichini, 2000). Editors and journals could **consider new ways of leveraging the fact that referees are also authors when designing incentive structures**.

3.4.2.11 Making individual turnaround times public (Stage 3 & 4) Some journals report aggregate turnaround time statistics, but publishing individual-level data is relatively unheard of. At the *Journal of Public Economics*, Chetty et al. (2014) tested a social incentive treatment of **sharing individual referee turnaround times**, for which the main significant effects were reductions in the referee invitation acceptance rate and the word count of reports (Table 5). In 2020, the now-former editor of the *JFE* reported the names of all of the journal's referees, their individual turnaround times, and the editorial acceptance rates of manuscripts they refereed (Goldstein et al., 2021). The referees had not agreed to this information being made public, so it was removed in response to the controversy. Given the sensitive nature of this information, informed consent policies would likely need to be strengthened in order to implement this approach. This treatment could also be applied to editors, either by reporting the total time taken to decide on submitted manuscripts or by reporting stage-specific turnaround times for making a desk rejection decision, choosing referees, or making a final decision once reports are received.

3.4.2.12 AER: Insights model (Stage 5) The adoption of practices similar to those of *AERI* could relieve pressure on the peer review system. *AERI* submission guidelines **limit manuscript length** to 6,000-7,000 words and five exhibits (*AERI-1*). Papers are either **conditionally accepted with implementable changes or rejected**, with **no subsequent rounds of review**. The shorter average and median decision times for manuscripts sent to referees at *AERI* (47 and 44 days) as compared to those at the *AER* (87 and 77 days) over the 2017-2018 timeframe offer suggestive evidence that these policies could be effective, though this is a crude comparison (*AERI-2*, *AER-11*). Certain components of this model have been implemented at other journals. In 2007, the editor of *Economic Inquiry* introduced a “no revisions” track that authors can opt into, arguing that it helps the journal focus on its role of evaluating manuscripts rather than rewriting them (McAfee, 2010). Some concerns offered by commentators include that the policy could increase the risk of both Type I and Type II error with publication decisions or that authors could be deprived of useful feedback, though the editor argues against these claims. The *Review of Economics and Statistics* (*REStat*) created a track for short papers (*REStat-1*) but made no change to the R&R policy. Only certain types of

Figure 30: Support for the AER: Insights model



Notes: This figure is based on responses to question Q17. N = 1,459.

manuscripts are suited for one-round and/or short paper submission tracks, but advocates argue that they allow some authors to make arguments concisely and disseminate publications rapidly. As shown in Figure 30, there is overwhelming support among our respondents for this type of model, and across all population subgroups.

3.4.2.13 Reducing default revision times (Stage 5) Authors who receive R&Rs at top journals are often requested to return the manuscript within a year, as is the case at *Econometrica* and the *JPE*. ([ECMA-3-b](#), [JPE-3-b](#)). With 2+ revision rounds, this can cause multi-year long delays between the initial submission and the official acceptance decision. This could be quite inefficient, as authors, editors, and referees alike might forget key details about the manuscript. To truncate the right tail of delays, one possibility could be to [cap the length of the default revision period](#).

3.4.2.14 Using conditional acceptance offers and avoiding rejects & resubmits (Stage 5) Reducing the ambiguity of editorial decisions could improve the efficiency of peer review and avoid frustration for stakeholders. Issuing conditional acceptances that only include feasible and essential revisions, rather than traditional R&Rs, is one way of achieving this. Reject-and-resubmit decisions also tend to generate uncertainty, so journals might benefit from [minimizing or even eliminating their usage](#), as the *JF* did in 2016 ([Nagel, 2017](#)).

3.4.2.15 Realigning editorial incentives around revisions (Stage 5) A recurring complaint from some respondents was about the asymmetry between editors' ability to request revisions and the cost borne by authors who are expected to implement them. One way of alleviating this imbalance could be to [realign incentives](#) such that journals and editors internalize more of the cost and risk of revisions. For example, requiring that editors explicitly indicate which revisions (if implemented) are sufficient to turn an R&R into a publication could increase editorial engagement with proposed changes and reduce the uncertainty faced by authors. 26 survey respondents mentioned that this or a similar proposal would be a substantial improvement over the editor simply requesting that they address all concerns raised by referees. One respondent even suggested that editors could offer contract-like agreements to authors, stating that their manuscript will be published if certain criteria are met after revision. Finally, another possibility could be for journals to commit to providing funding for additional research or data collection that is necessary to implement the requested revisions. This practice might somewhat internalize the financial and time cost that journals impose on authors when their editors request changes.

Table 5: Comparison of incentive treatments

Study (Year) [Type]	Intervention(s)	Result	Main takeaways
Hamermesh (1994) [Observational Study]	No Incentive (Baseline)	Not given	1. Financial Treatment: reduces turnaround time 2. Mechanism: bunching of submissions before deadline
	\$36 Payment (Financial)	2 week reduction in average days to review	
Thompson et al. (2010) [Natural Experiment]	No Incentive (Baseline)	Average 49.6 days to review Average 1.5 page report length	1. Financial Treatment: reduces turnaround time, increases report length
	\$50 Payment (Financial)	Average 41.6 days to review Average 1.7 page report length	
Squazzoni et al. (2013) [Lab Experiment]	No Incentive (Baseline)	Cooperation Index ^c of 0.63	1. All Treatments: reduce quality & efficiency 2. Mechanism: crowding out of intrinsic motivation
	Fixed Incentive (Fixed Financial)	Cooperation Index of 0.54	
	Incentive A ^a (Var. Financial A)	Cooperation Index of 0.60	
	Incentive B ^b (Var. Financial B)	Cooperation Index of 0.58	
Chetty et al. (2014) [Field Experiment]	6 Week Deadline (Baseline)	67.6% acceptance of invitation Median 47.8 days to review 77.9% editor-referee agreement Median 877 word count	1. Social Treatment: reduces invitation acceptance and word count 2. Deadline Treatment: reduces median review time 3. Deadline + Financial Treatment: reduces median review time and word count, increases invitation acceptance 4. All Treatments: weaker impact on tenured than untenured professors
	6 Week Deadline Publication of Response Time (Social)	61.1% acceptance of invitation Median 45.9 days to review 76.2% editor-referee agreement Median 757 word count	
	4 Week Deadline (Deadline)	64.1% acceptance of invitation Median 35.5 days to review 77.5% editor-referee agreement Median 864 word count	
	4 Week Deadline \$100 Payment (Deadline + Financial)	72.0% acceptance of invitation Median 27.5 days to review 76.2% editor-referee agreement Median 786 word count	

^a *Incentive A* represents a variable incentive designed to align the incentives of referees and editors.

^b *Incentive B* represents a variable incentive designed to align the incentives of referees and authors.

^c The Cooperation Index was an indicator used by the study authors to measure the efficiency, equity, and cooperative nature of the outcome of the peer review game.

3.5 Innovations in peer review

Until now, our focus has been on the diagnosis of discrete problems and the examination of proposals that specifically target those problems. In this section, we explore ideas that are more transformative and controversial in nature, with the aim of promoting innovation and experimentation on the part of journals and other stakeholders.

3.5.1 The key dimensions of peer review

The Committee on Publication Ethics (COPE) identifies six (mostly non-mutually exclusive) dimensions on which models of peer review can vary, which are presented in Figure 31 (COPE Council, 2017a). In general, the right-most column presents the more open/transparent option for how to structure peer review in each dimension. But COPE also notes a tension between openness and confidentiality: even with moves towards openness, the expectation of confidentiality remains in that information should not be shared without explicit agreement from the relevant stakeholders (COPE Council, 2017b). By these metrics, the top 5 journals are unanimous in following what COPE calls the “standard” approach to peer review: pre-publication timing, single-blind identifiability, editor-controlled mediation between reviewers and authors, unpublished referee reports, journal-facilitated review, and reviewer ownership of referee reports. But this apparent consensus belies the heterogeneity that exists elsewhere, especially in other disciplines. Natural science journals have been particularly active in rethinking traditional practices. For many years, the *Nature* journals have run trials to test proposals that increase transparency in peer review, examples of which can be seen [here](#), [here](#), and [here](#). Meanwhile, online-only general science journals like *PLOS ONE* and *F1000Research* embrace models that challenge the most fundamental precepts of traditional peer review. But within economics, comparatively few efforts have been made at pushing the boundaries of the standard model; in consideration of the proposals here and elsewhere, economics could follow suit with trials and innovations of its own. Below we examine a list of proposals, many of which innovate along at least one of the dimensions listed in Figure 31. When this happens, we identify the relevant dimension in parentheses in the paragraph title.

Figure 31: Dimensions of peer review

Timing	Preprints	Pre-publication	Post-publication
Identifiability	Double blind	Single blind	Open
Mediation	Editors mediate all interactions between reviewers and authors	Reviewers interact with one another openly	Reviewers and authors all interact with one another openly
Publication	Peer reviews are not published	Peer reviews are published but not signed	Peer reviews are published and signed
Facilitation	Review facilitated by a journal	Review facilitated by a third party	Review facilitated by authors
Ownership	Review owned by a journal or third party	Review owned by the authors of the reviews	Shared or mixed ownership of reviews

Notes: This figure lists dimensions on which models of peer review can vary (left-most column) and potential policies for each dimension, with the right-most column generally being the most open and/or transparent. *Source:* COPE Ethical guidelines for peer reviewers - English: <https://doi.org/10.24318/cope.2019.1.9>. Version 2: September 2017, COPE Council.

3.5.2 Transparency in peer review

The first path to innovation addresses transparency in the peer review process. In an era of increasing support for open science and transparent research practices for the production of research, a natural question is whether the same degree of transparency should prevail for the evaluation of research, and what transparency means for peer review. At a fundamental level, transparency speaks to two main dimensions of peer review: (i) the identifiability of peer reviewers and (ii) the publication of peer review documents (whether anonymously or not), including referee reports, responses from the authors, and decision letters from the editors. The practice of “open peer review” typically refers to a combination of these two dimensions. Below we analyze proposals that innovate on one of the dimensions separately or on both in combination.

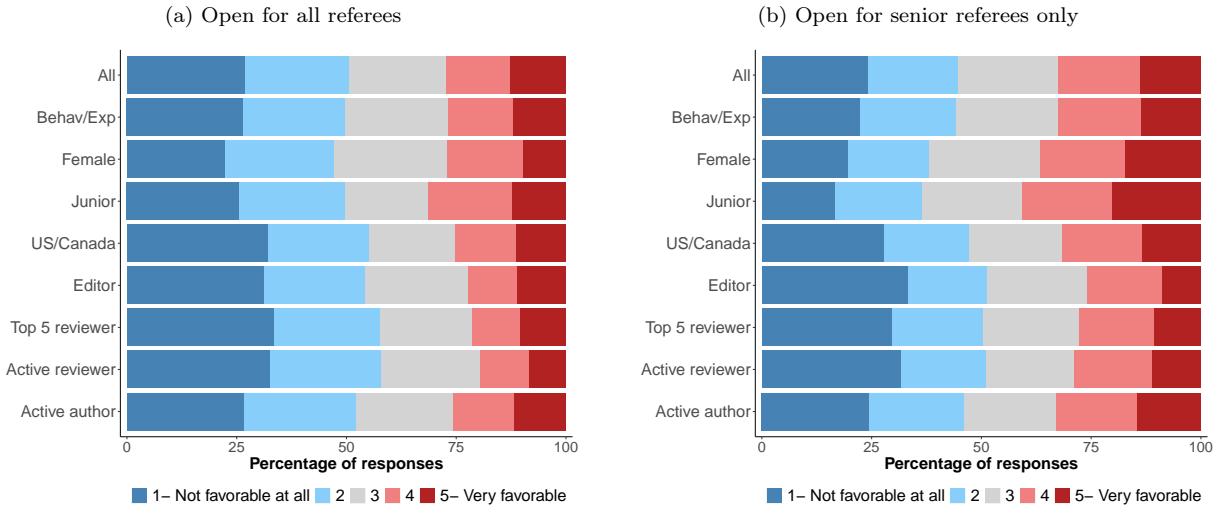
3.5.2.1 Making reports and correspondences publicly available (Publication) Economics journals typically provide no information about the review history of published papers. In contrast, at least nine *Nature* portfolio journals, including *Nature Human Behaviour*, have announced their intention to **publish referee reports and editorial letters alongside articles** as of February 2020 (sometimes still subject to author and referee preferences) (*Nature*, 2020). The status quo in medicine is mixed, with the *British Medical Journal* also publishing reports (BMJ-3) but *The Lancet* treating them as confidential (LAN-2). Publons (since acquired by Clarivate) already offers a **Transparent Peer Review** feature that allows referees to publish their reports with journal permission, although *PLOS ONE* and *F1000Research* are the only journals in Table A.3 with this feature enabled as of data collection. Making this information available could have the benefit of helping readers think critically about the paper, learn from the reviewers’ expertise, and better assess the impact of peer review on research outputs. In addition, publishing reports might motivate referees to write more professional and substantive reviews. As a potential downside, referees who feel uncomfortable with this idea might opt out of peer review or censor some of their constructive comments. There is little evidence of this particular downside from the *Nature Communications* trial (*Nature*, 2020) and from Bravo et al. (2019), but research from medicine finds that the practice is associated with a lower referee invitation acceptance rate and longer report turnaround time (with no adverse effect on report quality) (van Rooyen et al., 2010).

3.5.2.2 Publishing reports of rejected papers (Publication) The previous proposal concerned referee reports for published papers, but *Meta-Psychology* also **publishes them for rejected papers** (MP-1). In fact, they document decisions made for all submissions through their page on the Open Science Framework website. Providing the full population of manuscripts considered could help readers better gauge the requirements for publication at a journal. Furthermore, having the entire submission history of a paper publicly available could provide a more nuanced view of its merits, as compared to it only appearing once in the accepting journal. Despite these potential upsides, such a proposal is likely to face more resistance, particularly from authors who might not want their history of rejections to be made public.

3.5.2.3 Making reviewer and handling editor identity public (Identifiability) Another way of making peer review more open is to **publish the identity of those who reviewed a paper**. This proposal has been implemented on an opt-in basis by some journals, such as *Nature Human Behaviour* and *PLOS ONE* (Nature-3-c, PLOS-6), with varying degrees of information disclosure. One option is for journals to acknowledge contributing referees by name, but without attributing a report to them (the default at *NHB*, although referees can also sign their report). Another option (offered by *PLOS ONE*) is to **link the reviewer name to the report they wrote**, when consent is given. We note that this difference might substantially affect referee and author behavior e.g., in terms of motivating reviewers to do a good job and the potential for retaliation or reciprocity. Revealing identities could help readers understand the reviewer’s expertise and promote dialogue about the article. Others report that open peer review has also helped expose unrecorded reviewer conflicts of interest (Benos et al., 2007). On the negative side, it could also disincentivize participation in the system and lead the two parties to swap favors, especially when it is clear who wrote a report. A fairly comprehensive treatment of the pros and cons of open peer review can be found in Tennant et al. (2017). Another interesting variation on identifiability comes from *Frontiers*, which only publishes the names of reviewers who recommended the manuscript for publication, in order to protect critical referees from retaliation (Front-1-c). A related step that could potentially enhance accountability

for publication decisions would be to generalize the practice of **publishing the name of the editor** who handled a given article. Some top economics journals that routinely note which editor handled each article include *Econometrica*, the *AER*, and *Management Science*, based on recently published articles from these journals. One potential caveat might be a disincentive for editors to accept bold or controversial research for publication.

Figure 32: Attitudes towards open peer review (publication + identifiability)



Notes: Panel (a) is based on responses to question Q9. Panel (b) is based on responses to question Q10. N = 1,459 for both panels.

3.5.2.4 Signing reviews (Identifiability) Some academics already **sign their referee reports**, with or without explicit editorial permission. Many interdisciplinary perspectives about referees' experiences with this practice have been shared.⁵⁰ A common argument in favor of signing reviews is that it can be a useful commitment device for authors to write a quality report. However, if signing reviews is made optional, referees have the potential to abuse this option by selectively signing positive reports with the intention of receiving a favor in return and leaving negative ones anonymous.

3.5.2.5 Respondent attitudes towards transparent peer review practices Since open peer review practices may crowd out contributions to peer review, it is important to evaluate potential support. To this end, we asked respondents to rate how favorable they would be to an open peer review policy where “*referees sign their reports and the entire review history (including responses to referees) is disclosed*.” Respondents were asked to rate the policy, in general, and again if **only applied to senior reviewers**. As shown in Panel (a) of Figure 32, this policy finds little support, with about half of respondents reporting unfavorable views and only a quarter being in favor. Restricting the policy to senior reviewers (Panel (b) of Figure 32) barely increases support and, unsurprisingly, senior reviewers have less favorable views than inexperienced reviewers; support is slightly higher from female and junior respondents. We also note that this lack of support is matched with strong skepticism about its usefulness for improving the quality of peer reviews, although views are more divided when it comes to revealing the identity of associate editors (see Appendix Figure E.19).

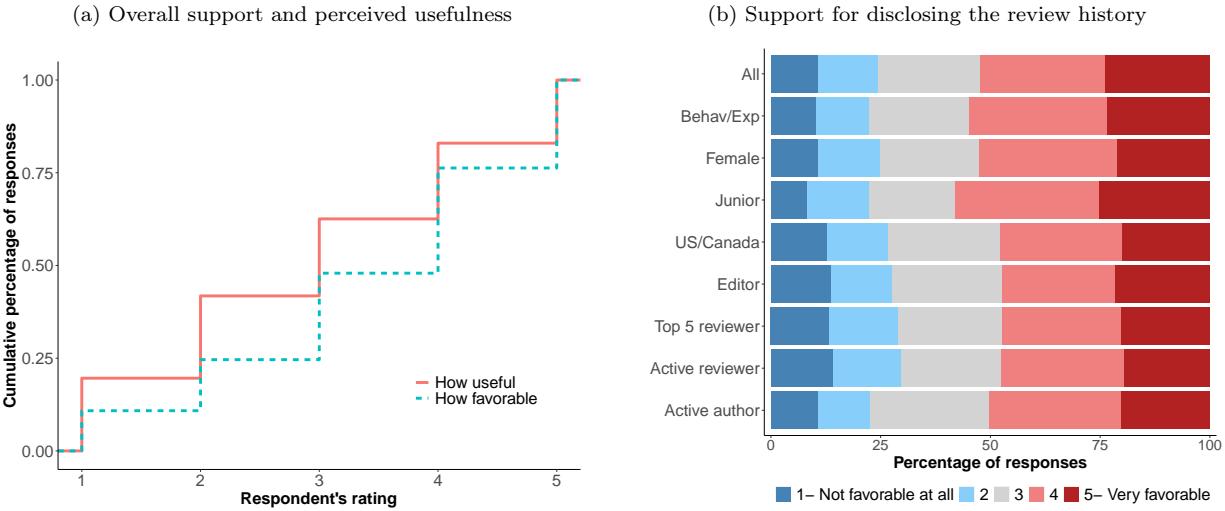
To separate the issue of identifiability from that of publication, we also asked respondents for their views on “*making the history of reports/responses to referees publicly available in an anonymized way unless the reviewers choose to disclose their identity*,” citing the example of *Nature Communications*.⁵¹ As shown in

⁵⁰An economist's account can be found at this [page](#), but perspectives from other disciplines appear more common: see accounts [here](#), [here](#), and [here](#).

⁵¹One potential issue with this approach is that anonymity might be hard to guarantee, especially when it comes to longer and more involved exchanges between referees and authors.

Panel (a) of Figure 33, a slight majority of respondents express favorable views (rating of 4 or 5), with only 25% reporting unfavorable views (rating of 1 or 2). At the same time, respondents appear fairly skeptical about the effectiveness of this policy for improving the quality of reviews, with only about a third of respondents expecting it to be useful (rating of 4 or 5). In terms of heterogeneity, we note again stronger support from junior researchers and, to a lesser extent, active authors and behavioral/experimental economists. In light of these mixed findings, more evidence would need to be gathered to understand the benefits and limitations of disclosing peer review histories.

Figure 33: Views on publishing the peer review history (publication without identifiability)



Notes: Panel (a) is based on responses to questions Q11 and Q6, item *iii*; because of differences in wording between these two questions, caution should be exercised in the comparison of the two ratings. Panel (b) is based on responses to question Q11. N = 1,467 for Panel (a) and N = 1,459 for Panel (b).

3.5.3 Other peer review models

In view of this report's findings, readers might conclude that some of the particularly intractable challenges facing peer review are unsolvable within its current confines. One such challenge worth emphasizing is the ambiguity about the optimal number of referees to assign to a manuscript. In the Welch (2014) study exploring the interplay between referees and editorial decisions, the six economics journals studied⁵² typically assigned two or three referees to each paper, with the averages ranging from 1.6 to 2.6 referees per manuscript, while the finance journals studied tended to only assign one. By soliciting multiple opinions, journals might reduce the risk of Type I or Type II error relative to relying on the potentially erratic recommendation of just one person. But in a study about the assessment of grant proposals, Forscher et al. (2019) found that maximizing the informativeness of refereeing might require many more reviewers than are currently used, to the point of being prohibitive,⁵³ and that adding additional referees seems to have diminishing returns. Furthermore, recent evidence suggests that peer review evaluations are worse predictors of subsequent citations than algorithmic models based on manuscript characteristics (Abramo et al., 2019). While citation counts are just one important metric for journals, the inability of resource-intensive peer review to outperform algorithms on this dimension raises questions about how much value it adds. Other major challenges facing peer review that cast doubt on the traditional model include the potential reproducibility crisis in economics (Maniadis and Tufano, 2017) and the (in)ability of high-impact research (e.g., about COVID-19) to get published in a timely manner while still enforcing a benchmark of quality. Taken together, these challenges may suggest that transformative changes to the peer review process are needed, some of which

⁵²These journals being *Econometrica*, the *International Economic Review*, the *Journal of the European Economic Association*, the *Journal of Economic Theory*, the *Quarterly Journal of Economics*, and the *Rand Journal*.

⁵³The use of three referees results in an inter-reviewer reliability score of 0.2, while 12 brings it up to 0.5.

we will consider below. This subsection is devoted to exploring ideas that fundamentally restructure the process of peer review and the interactions between its stakeholders.

3.5.3.1 Post-publication peer review (Timing) *F1000Research* ([F1000-2](#)) publishes authors' submissions almost immediately, after an initial check for compliance with policies. The journal then conducts an open **post-publication peer review**, after which the manuscript is marked as peer-reviewed and indexed in external research databases if the referees give their approval. The unique appeals of this proposal include its ability to rapidly disseminate research and the minimal amount of subjectivity in the editorial process. There is reason to believe that many authors appreciate this publication-first approach: academics are increasingly sidestepping peer review at top journals and posting their research on their website or on working paper repositories like NBER, IZA, CEPR, RePEc, and SSRN, where they will get quick access to a wide audience ([Ellison, 2010](#)). This option is especially attractive to already-tenured academics who do not face the same pressure to publish in top 5 journals ([Heckman and Moktan, 2020](#)). This practice is not universally appreciated: some emphasize the danger of circulating research before it is properly vetted, arguing that online commentary is an insufficient alternative to peer review and subsequent updates to working papers often go unnoticed, even if conclusions change ([Özler, 2011](#)). Indeed, [Bolboacă et al. \(2019\)](#) show that the impact of the initial versions of retracted papers can continue long past their withdrawal. If post-publication peer review is too risky, then the proliferation of **fast-tracks** for COVID-19 papers (like those at the *Journal of Environmental Economics and Management* and the *Journal of Urban Economics - JEEM-1, JUE-1*) is an encouraging sign that pre-publication review can be done rapidly. We note that pre- and post-publication review are not mutually exclusive: after manuscripts are approved for release in an initial review round, the peer review process could continue in the public sphere. This could take many forms, including the referees and authors continuing their discussion from the initial round, alternative referees being employed for another public round, or peer review happening continuously in the form of comments from the academic community (discussed further below).

3.5.3.2 Peer review of registered reports (Timing) We note the possibility of conducting **peer review of registered reports** and conditionally accepting manuscripts for publication before results are known, although this is not one of COPE's options for the timing dimension. If done correctly, this practice could mitigate the issue of publication bias that contributes to the replication crisis. Another possible benefit of this proposal is that ex-ante peer review may facilitate the pursuit and publication of potentially surprising results more than traditional ex-post peer review does, as argued by [Gross and Bergstrom \(2021\)](#). Some journals already accept registered reports: for example, the *Journal of Development Economics* (*JDE*) started a pre-results review track after successfully conducting a pilot trial ([JDE-1](#)) and *NHB* conditionally accepts manuscripts based on registered reports ([Nature-5](#)). These tracks typically also allow the authors to conduct additional unregistered exploratory analyses, as opportunities arise. This proposal can be combined with other proposals that alter the timing and nature of peer review.

3.5.3.3 Crowdsourced peer review (Mediation, Facilitation) A recent analysis of the book market found that crowdsourced reviews (from Amazon) substantially outperform professional reviews (from the *New York Times*) in terms of their ability to increase consumer welfare by creating a deeper pool of information for consumers ([Reimers and Waldfogel, 2021](#)). At the same time, the rise of crowdsourced reviews has not diminished the impact of expert reviews, suggesting that they are complements, not substitutes. Analogously, **reviews from crowds of academics** might enhance the supply of information and address the issue of low inter-reviewer reliability identified above. This approach could be used either as a replacement for or as a complement to traditional refereeing. Given a platform, crowds could rate a manuscript's attributes, provide feedback, and ultimately come to a consensus about its merits and limitations. An implementation of this idea can be found at CrowdPeer (<https://reimaginereview.asapbio.org/listing/crowdpeer/>); this website automatically indexes preprints and facilitates community discussion with upvotes, downvotes, and comments. Journals or publishers could also set up their own platforms to foster the engagement of their readership and encourage these community reviews. *PLOS ONE* has piloted a program for soliciting community comments on preprints, but it has seen limited engagement; the timing of feedback is also often too delayed to assist editorial decisions ([Porter, 2020](#)). Another attempt at facilitating crowd review came from Academic Karma, but it shut down due to difficulties integrating with journal systems and a lack

of engagement.⁵⁴ Motivating academics to participate in a platform might be challenging if they are not convinced of its value or if it becomes burdensome. Additionally, without appropriate moderation, online communities can tend towards inappropriate discussion (as in Wu (2019)). The feasibility of this proposal could be explored with limited groups of academics before being opened up to the public.

3.5.3.4 Non-academics as referees (Facilitation) Traditionally, economics journals have engaged academics to write referee reports for submitted manuscripts, especially academics who are experts in the relevant subfield. Despite the near-ubiquity of this practice in economics, specialized referees are not necessarily the only stakeholders with an interest in the research and a capacity to provide meaningful feedback. Indeed, two of our survey respondents suggested that **peer review could be opened up to “outsiders,”** such as policymakers who might be able to comment on the applicability of research findings or economists who work outside of academia. As a unique example of this, the *BMJ* solicits feedback from medical patients to gather their thoughts on the significance of submitted manuscripts (**BMJ-2**). Economics could do something similar e.g., by consulting community leaders about the presentation, interpretation, or scalability of research insights from field experiments conducted in development economics.

3.5.3.5 Non-economists as referees (Facilitation) It is common for economics journals to publish articles that engage with other fields, especially other social sciences like psychology or political science. Moreover, most economists appear to support moving in the direction of more interdisciplinary research (Andre and Falk, 2021). When journals receive submissions of a multidisciplinary nature, they might consider **employing referees with credentials in the other relevant field(s)** in addition to referees from economics. This could enhance the credibility of economics journals as outlets for multidisciplinary research and also encourage authors to write their articles to be understood by wider audiences. Lemon et al. (2019) from *PLOS ONE* note that a potential downside of this proposal is the extra burden it places on editors: multidisciplinarity might generate more conflicting views that are harder for editors to reconcile.

3.5.3.6 Prediction markets to predict the replicability of manuscripts (Timing, Facilitation) Another idea that engages the wisdom of crowds is to **use prediction markets or individual betting mechanisms in the evaluation of manuscripts.** This has already been tested by e.g., Dreber et al. (2015) and Nosek et al. (2022), who found that these markets are able to successfully, albeit not perfectly, predict whether specific psychology studies are replicable or not. Given this, using prediction markets or other betting mechanisms to assess manuscripts could be a powerful tool to address the social science reproducibility crisis. These mechanisms could be used at early stages of project development to advise the authors on whether more evidence should be gathered; alternatively, they could also be used post-publication to complement the feedback from the traditional peer review process.

3.5.3.7 Expanding and sharing ownership of reviews with explicit agreement (Ownership) As a matter of copyright law, referee reports are assumed to belong to reviewers unless explicitly agreed otherwise (COPE Council, 2017b). But since peer review is conducted confidentially and reports are unique products created at the request of journals, reviewers are limited in what they can or should do with them, as a practical matter. They might expect backlash if they unilaterally post their referee reports elsewhere, for example. Furthermore, it is ambiguous whether the expectation of confidentiality remains once the peer review process is complete. One possibility could be to grant referees explicit permission to **include referee reports in their regular academic portfolio** (after a decision is made), which could increase the visibility of refereeing and incentivize better performance. This is sometimes already possible on Publons, but journal policies rarely permit it: most of the noteworthy journals that allow this are open access journals belonging to the publisher Multidisciplinary Digital Publishing Institute (https://publons.com/journal/?permission_content=1&order_by=reviews) or to other publishers, like *BMJ Open* or *SAGE Open*. A caveat is that referees might only share reports that cast themselves in a positive light. Authors may also want to share the reports they received from a prior submission when submitting their work to a different journal. Earlier, we also discussed open peer review and the expansion of inter-publisher sharing of reports, both of which imply an increased level of journal ownership over reports. When

⁵⁴See announcement here: <https://academickarma.wordpress.com/2019/04/24/academic-karma-is-closing-down/>.

considering implementing any of these proposals, the COPE Council (2017b) advises that explicit agreement be obtained from all stakeholders about the appropriate usage of referee reports so as to avoid violations of confidentiality.

3.5.3.8 Professionalizing the role of referee Refereeing as currently conceived is essentially a form of ad-hoc volunteering done by academics with little formal recognition. An alternative model would be to **contract academics to be part-time or full-time referees**, in the same way that they can become editors or coeditors. In addition to the editorial board, there could be a board of referees who agree to write a certain number of reviews for a journal each year. If made public, such information could greatly facilitate the allocation process, increase the visibility of reviewing, and reduce the possibility that a researcher is asked to work for multiple journals at the same time. This specialization already occurs organically to some extent: among our respondents, 27% write at least 75% of their reports for a specific category of journals (top 5, top field, other economics, or other field). The top 5 journals vary considerably in the extent and structure of their editorial staff (Table 4), so there exists room for experimentation with different roles. These professional referees could be offered compensation, but universities could also take the opportunity to start **considering peer review contributions more systematically when making promotion or tenure decisions**. One potential downside might be that professional referees lack independence from the editors who oversee them, although this is already somewhat of an issue with traditional peer review.

3.5.3.9 Platform for authors to solicit feedback without journal submission When authors hope to receive feedback on their manuscript before submission, they are often limited to asking their immediate network for help. Researchers with more extensive and elite networks (as discussed in 3.2.3) might be able to solicit such feedback easily, which may not be true for less well-connected authors. Those without adequate networks might instead submit manuscripts to journals before they are ready and/or hope for good feedback from referee reports, which could contribute to the submission misallocation documented in 3.2.2.1. One way to address this could be to facilitate **manuscript feedback outside of journal peer review**, as is done by <https://prereview.org/>. The platform catalogs preprints that have DOIs and allows any user to request or provide reviews for them. If this or a similar platform gained momentum, it could become a channel for authors to receive useful feedback before making any journal submissions.

3.5.3.10 Providing channels for a more interactive review process Current peer review procedures do not give much room for interaction between the parties involved in the review process. More channels for communication could be opened to make the review process smoother. An ambitious version of this idea was implemented by *Frontiers*: after one round of traditional peer review, a second round of “interactive review” between authors, referees, and the editor is facilitated (**Front-1-b**). In this phase, the parties communicate directly with each other about the manuscript via comments in an online forum. This practice is a departure from the traditional round-by-round process of revision and review, but is comparable to how authors often cooperate to write their manuscripts initially i.e., interactively with comments on documents and live discussions. The potential advantage of this approach is efficiency: the active back-and-forth allows authors to ask questions and receive immediate clarification, which could help to prevent misunderstandings that might extend the process. As a downside, participating in this phase might require more attention from reviewers at specific times. In addition, more frequent contact between parties might increase the likelihood of unconstructive comments in the absence of any moderator. If a fully interactive system is too demanding, intermediate **steps towards making communication more interactive** between authors, editors, and referees could also be implemented. In particular, 15 respondents indicated their support for enhanced communication channels between authors and referees (and sometimes editors), often noting that the ability to make instant clarifications might clear up some ambiguity in the review and revision process. Communication between referees could also be enhanced: 5 respondents suggested allowing for **discussion and/or evaluation between referees** who are reviewing the same manuscript. For example, referees could read each other’s reports to correct inaccurate statements and possibly harmonize their assessments and suggestions. This proposal could increase the level of agreement between referees from its current low level, although that development might not be strictly positive (see discussion in 3.3.3.1).

3.5.3.11 Author- and third party-facilitated peer review (Facilitation) The current paradigm of peer review in economics is journal-facilitated peer review. That is, journals act as mediators and match the authors who demand peer review to the referees who supply it. With the internet democratizing access to communication and information, it might now be possible for authors to **facilitate their own peer review** in a credible way with the help of third-party platforms. This could either be done by enabling authors to contract their own referees or by hosting a public commentary platform separate from a journal. One implementation of the latter variety comes from the [Researchers.One](#) platform. It allows authors to upload their manuscripts online, which become immediately available to the public for commentary, and easily upload revisions. Authors might find the disintermediation appealing for the level of control it grants them and the immediacy of the submission and publication process. But since any paper can be uploaded onto the platform and they are not taken down if they receive negative feedback, some may be concerned about the quality of submissions and the lack of gatekeeping. Moreover, the public commentary aspect of this implementation shares the same downsides as crowdsourced peer review, including the possibility of toxic discussion and limited engagement.

3.5.3.12 Centralized manuscript processing and allocation (Mediation, Facilitation) The centralized platform discussion in Paragraph 3.2.4.1 was primarily framed around centralizing information and communication. One particularly ambitious extension of this idea could be to **centralize processing** as well: instead of authors making discrete submissions to individual journals, submission to the platform could facilitate consideration at all (participating) journals. Many variations on this idea are possible. The platform could conduct peer review itself (e.g., by sending manuscripts to reviewers or using a more innovative method) or act as a repository for editors to compare available manuscripts. Centrally conducted peer review could help to prevent redundant review rounds and increase the manageability of referee workloads. Even simply acting as a repository for editors to compare manuscripts might induce journals to engage with normative questions about the type and mix of research they want to publish. As the next step, the platform's allocation mechanism could determine where the manuscript gets a publication offer, if anywhere. [Prufer and Zetland \(2007\)](#) study a model where journals can bid to obtain the exclusive right to publish a paper, while still being able to reject it after a full review process. An appropriately designed mechanism could mitigate the issues of submission misallocation and authors "shopping around." The benefits of the platform would likely increase with participation rates, but inducing widespread coordination would not be straightforward, if only because the various stakeholders might have very different views on what constitutes the best allocation mechanism. Nevertheless, we note that economics has previously succeeded in centralizing key functions, in particular the job market. The AEA's annual Allied Social Sciences Associations (ASSA) meeting facilitates the matching of job candidates to employers by serving as a centralized venue for conducting interviews ([Coles et al., 2010](#)). Just as job candidates put great effort into their job market papers in the hope of securing scarce interview slots, authors in a similar setup might feel pressure to maximize the quality of their manuscripts prior to submission (as opposed to submitting underdeveloped manuscripts for feedback). To assess the feasibility and efficiency of such a solution, perhaps a centralized submission system could be set up for a specific subfield (e.g., experimental economics) as a trial.

4 Taking a step back... but also moving forward

The great engagement we received from our survey respondents attests to the importance of our peer review system and the urgency of making changes. Over 1,400 economists made the effort to respond to yet another not-so-short survey during a difficult time, presumably while juggling countless other personal and professional responsibilities. We are especially grateful for the thoughtful and often highly-detailed comments that respondents left, describing their experiences and ideas. Having now incorporated these insights into the report, we invite respondents and other community members to continue their involvement in this ongoing project with further discussions and initiatives. We will briefly summarize our key takeaways, make note of remaining unanswered questions, and outline action-oriented steps that authors, referees, editors, and other stakeholders can take to induce change.

4.1 Taking stock of where we are

4.1.1 Challenges facing our peer review system

A number of significant concerns about the state of peer review emerged in the course of writing this report. First, we noted that the distribution of refereeing work within the discipline appears to be both inequitable and inefficient. In addition, manuscripts might be allocated to journals and reviewers who are inappropriate fits, either in terms of submission quality, manuscript topic, and/or potential conflict. Next, we found that lower-quality reports and review delays are common complaints about the peer review process. This could be for a number of reasons, such as the insufficient training and guidance given to reviewers or the perception of many referees that their work is not adequately acknowledged or rewarded. Finally, the status quo in peer review is increasingly being called into question due to concerns about a lack of transparency in decision-making and the publication of papers that are not reproducible.

4.1.2 Potential solutions and respondent attitudes

With input from respondents and a review of the existing literature, we presented a diverse set of potential solutions to the issues identified. The proposals we considered included doctoral training programs, enhanced reviewer guidance, modified default deadlines, redesigned incentive structures, information-sharing partnerships, and many other tools. Since we could not address every issue and perspective, we also list proposals submitted by respondents for the reader's consideration in Table A.1. Despite heterogeneity in views, we found that some proposals are relatively uncontroversial and are possible to implement immediately, such as systematically sharing editorial decisions and reports with reviewers. Other proposals only command support from slim majorities or pluralities, such as disclosing the review history of manuscripts. In these cases, conducting a trial at a journal would be a reasonable step forward. Finally, other proposals were perceived as less plausible or possibly counterproductive. Although a large number of respondents would support a centralized platform, it is unclear whether there is enough enthusiasm to overcome the collective action problem. More immediate steps could be taken in this direction, such as linking reviewers to their ORCID iDs. Meanwhile, most respondents appear skeptical about open peer review practices, especially the disclosure of reviewer identities. In this area and in some others, a major shift in policy will likely require stronger support from the profession, possibly after successful trials conducted at journals. Interestingly, our analysis suggests fairly little correlation in attitudes towards any two proposals, suggesting that they could, and maybe should, be treated independently (see Appendix Figure E.20).

4.1.3 Gaps in our knowledge

Our study presents several limitations that we would like to acknowledge. First, our respondent pool is not entirely representative of the profession, even based on observables; as a result, care should be taken in extrapolating our findings to the entire population. Second, we had to make difficult choices in the survey design to minimize response time and could not investigate all dimensions of the peer review process. Third, self-reported data is typically subject to measurement error, and our survey efforts are not immune to this issue. In addition, we cannot guarantee that respondents interpreted our question wording uniformly, potentially leaving some interpretational gaps. Importantly, the data produced in this study is merely correlational and provides limited information about the factors that may drive the patterns we observe. For all these reasons, it will be important to conduct further survey efforts and randomized trials that can provide causal evidence. Economists from various backgrounds may use their specialized skill set to contribute in more specific ways, as we note later (see Section 4.2). Journals like the *JDE* and *Nature* already run trials, but more coordinated efforts would be worthwhile. Systematic initiatives have precedence in similar contexts, such as with a recent project to improve grant peer review (NBER, 2021a). We hope that our report will provide a springboard for further investigations. For perspective on what has been done and is possible, it is worth mentioning the ReimagineReview registry (<https://reimaginereview.asapbio.org/>), which keeps a comprehensive list of ongoing and completed projects to improve scientific peer review.

We note that there are many positive and normative questions pertaining to peer review that our report left largely untouched. First, while we gathered information on submission volumes, we did not ask about authors' submission strategies or the peer review history of their most recent articles. A deeper analysis of

the demand side of peer review would provide us with a better understanding of the impact of our current system on publication trajectories. We also only partially explored the supply side of peer review. For instance, it would be useful to better understand the specific journals researchers referee for, the arrival process of refereeing requests, and the drivers of decisions to accept a review. On the normative side, it will be important to clarify what the profession thinks should be the involvement of each type of researcher who interacts with the peer review process. For instance, should the burden of peer review be allocated evenly across researchers or should it fall more heavily on tenured faculty? How should responsibilities be distributed? One unaddressed question that seemed at the top of our respondents' minds was about the role of editors. In fact, the word "editor" appeared almost as frequently as the word "referee" in the comments they left (141 vs. 144 instances; see word cloud presented in Appendix Figure E.21). Respondents frequently argued that editors are unrepresentative of the profession and take an excessively hands-off approach to peer review. Editors seek referee input, but is it their job to dutifully abide by referee recommendations or to exercise their own independent judgment? In other words, are editors delegates or trustees? In light of the comments we received, further elucidation and discussion on this topic would likely be welcome.

4.1.4 Changing both how we produce and evaluate research?

There are also normative questions about the type of research the profession should be encouraging (Akerlof, 2020). Peer review is not a neutral agent here: there is a clear interdependence between the production and the evaluation of research. Evaluation can affect production, like when we noted that ex-ante peer review might encourage more daring research than ex-post review (Gross and Bergstrom, 2021). Inversely, many adverse trends affecting peer review might originate from changing patterns of production, such as:

- Submissions appear to be getting longer and more technical (Schwert, 2021). Card and DellaVigna (2013) found that articles written in 2011-2012 were three times longer than those written in the early 1970s, even without considering online appendices. The trend appears to have continued at the *AER* since then, with average page length hitting its all-time high of 41 pages in 2018 (AER-4-a), up from 29 pages in 2010 (AER-5-c). Berk et al. (2017) argue that authors are increasingly adding robustness checks in anticipation of (or in response to) heightened journal acceptance standards and requests from reviewers. Despite this, they express skepticism that longer papers are accompanied by commensurate improvements in quality. Instead, long papers could increase the severity of review delays and the chance of the reviewer misunderstanding their content, due to constraints on referee time and energy. If that is true, authors who write simpler and sharper papers that focus on a key message may be rewarded with a smoother review process. Editors who appreciate these type of submissions might choose to explicitly solicit them or adopt practices similar to those of *AERI*.
- Meanwhile, the number of submissions has dramatically increased. In particular, submissions to the top 5 journals have increased by nearly 100% from 1990 to 2012. As noted earlier, the *AER* and *Econometrica* have also seen double-digit percentage growth from 2010 to 2020. While this trend may be exogenous to peer review in some ways e.g., as the result of improved computer processing power and an increasing number of co-authors per paper (Card and DellaVigna, 2013), enhanced pressure on researchers to publish in the top 5 journals is another potential explanation (Heckman and Moktan, 2020). Presumably, fewer papers being written would relieve pressure on peer review by lowering submission volume,⁵⁵ but we also note concerns that over-production might be adversely affecting research quality. The massive production of COVID-19 papers makes this point salient: NBER alone hosts more than 440 COVID-19 related working papers (NBER, 2021b), despite only 1.5 - 2 years passing since the pandemic began in earnest. In the hard sciences, where the volume of COVID-19 papers has been much more extreme, there are signs that the value and rigor of research output was compromised by this rushed approach (Redden, 2020; Silberner, 2021). Similar concerns could extend to other publication races, such as ones engaged in by competing researchers hoping to be the first to publish on a given topic. While researchers should be free to explore new ideas and contribute to pressing issues, there seems to be a broad consensus in the profession that quality should prevail over

⁵⁵One potential caveat is that producing fewer papers might lead researchers to invest more in each paper, which could exacerbate the complexity issue and possibly lead them to make more submissions per paper. However, we doubt that this would completely mitigate the impact of reduced paper production.

quantity (Andre and Falk, 2021). Some tangible consequences of over-production and over-publishing have been identified: Chu and Evans (2021) argue that innovative and disruptive scholarship has a harder time gaining momentum in over-saturated fields. Individual journals and authors might be negatively impacted as well. The model proposed by Atal (2010) suggests that the failure of journals to cooperate leads to suboptimally high acceptance rates, while they would collectively benefit from an agreement to publish fewer papers. Meanwhile, an experiment from Powdthavee et al. (2018) found that CVs that only list a small number of publications, all in higher-ranked journals, are perceived better than CVs with the same publications plus additional ones in lower-ranked journals. Thus, authors should ask whether each project is worthwhile by considering its marginal value to themselves, to the research community, and to society. In turn, this requires engaging with deeper questions about what is worth knowing and publishing (Frankel and Kasy, 2021; Andre and Falk, 2021).

Determining what researchers should be producing is outside the scope of this report, but we certainly believe these questions are worth discussing,⁵⁶ if only because they help us evaluate and improve peer review. Previously, we highlighted the lack of practical guidance given to reviewers on how to assess a paper, including how to structure the report, what to write about, and what the assessment should focus on. Providing these resources will require implicit engagement with questions about what deserves to be known and published, as journals will have to specify (i) the different publication criteria and (ii) how they should be weighted in the recommendation. Due to both horizontal and vertical differentiation, the precise publication criteria may reasonably vary between journals, reflecting different viewpoints. Clarifying these criteria could help reviewers formulate their recommendations and authors understand the basis of editorial decisions. It could also alleviate some of the misallocation issues by better communicating to authors how to focus their production and submission efforts.

4.2 Open questions and proposals

Ultimately, the issues we have discussed in this report cannot be addressed unless the community steps up to the plate. As such, we take this opportunity to issue a call to action. As a first step towards change, we developed a list of over 160 proposals that are potentially responsive to the problems identified (see Table A.1). Some proposals come directly from our report, but we did our best to make our presentation more comprehensive and reflective of the diverse array of perspectives in our profession. In particular, we read all comments from respondents and incorporated their ideas into the list. Respondents often indicated support for several proposals in a single comment; in these cases, we broke the comment into multiple rows such that each line represented a discrete idea. We then mapped these ideas to proposals that were either (i) already in our report or (ii) newly created to represent respondents' comments. This required judgment about which comments were sufficiently similar as to be grouped together. As some readers might disagree with our coding and interpretation, we provide the entire sheet of comments and their mapping on our OSF page (<https://osf.io/eczkv/>). This classification exercise was undertaken to organize proposals and facilitate discussion, not to advocate for specific solutions. In other words, we take no official stance on whether particular proposals will solve the problems they are intended to, or whether the cost of implementing them is worth the benefit. While we did our best to summarize the state of the evidence and offer a balanced perspective, we hope that our commentary is only the beginning of a longer and more fruitful dialogue. As we would like to spur on this discourse, we invite the community to use the forum on our website as a [discussion platform](#). To do our part in facilitating this discussion, we drafted a list of 15 open questions about peer review. They are presented here and embedded with links to the corresponding discussion threads:

- QUESTION 1.** Should we aim to build a centralized platform for gathering information about peer review?
- QUESTION 2.** Should peer review workloads be allocated on the basis of criteria like seniority, submission activity, and diversity?
- QUESTION 3.** Should we strengthen conflict-of-interest policies for reviewers and editors?

⁵⁶We use the word “discussing” rather than “knowing,” because as Andre and Falk (2021) emphasize, there are no objective answers to these questions and no consensus in the profession.

- QUESTION 4.** Should authors receive more direction about where to submit their papers?
- QUESTION 5.** Should the use of pre-submission feedback and evaluation be expanded?
- QUESTION 6.** Should editors and journals provide more information about publication prospects?
- QUESTION 7.** Would you find larger investments in training and guidance for reviewers worthwhile?
- QUESTION 8.** Should journals increase the use of desk rejections and expedited reviews?
- QUESTION 9.** Should more journals adopt short paper/fast tracks?
- QUESTION 10.** Do you think that referees need to be rewarded more for their contributions?
- QUESTION 11.** Should peer review contributions be explicitly considered in career progression?
- QUESTION 12.** Should more communication channels between authors, referees, and editors be opened?
- QUESTION 13.** Should evaluation continue after publication?
- QUESTION 14.** Should journals strengthen their transparency practices by releasing more data on peer review?
- QUESTION 15.** Should economics move away from its current model of peer review?

Users eager to convert these discussions into action are welcome to use the messaging feature of our forum for the coordination of working groups and other initiatives. However, we note that there is some ambiguity about which stakeholders would need to cooperate to implement each proposal. This may require further engagement with normative questions about the proper roles of certain actors in peer review e.g., editors and referees. In the last column of Table A.1, we make a first suggestion about who the primary action takers for each proposal might be, but these indications are in no way dispositive.

These issues are not straightforward, but the economics community seems uniquely well-equipped to think about them in productive and structured ways. These efforts would greatly benefit from broad-based participation by economists from many subfields. Economists who work on matching theory, social choice theory, and information aggregation might be particularly helpful in designing procedures to pair referees with manuscripts and deal with conflicting referee opinions. Econometricians with machine learning credentials might develop algorithms to quickly analyze manuscripts for their fit at a particular journal. Behavioral economists who study what makes incentives effective could advise journals on how to improve the speed and quality of reviews. More generally, researchers who work in metascience or at the intersection of several disciplines might know about the best practices from other fields. Finally, researchers with rich experience as journal editors or reviewers might have unique insights into which solutions are likely to be more feasible and effective. We conceive of this report not only as an invitation to share perspectives and engage in debate, but also as a challenge to act and contribute to improving our peer review system. We hope to see the development of working groups devoted to tackling specific issues and thinking about the practical implementation of potential solutions. We will do our best to assist in any way we can and look forward to seeing many of our fellow economists on the frontlines pushing for change.

References

- Abramo, Giovanni, Ciriaco Andrea D'Angelo, and Emanuela Reale**, “Peer review versus bibliometrics: Which method better predicts the scholarly impact of publications?,” *Scientometrics*, 2019, 121 (1), 537–554.
- Abrams, Eliot, Jonathan Libgober, and John A List**, “Research Registries: Facts, Myths, and Possible Improvements,” Working Paper 27250, National Bureau of Economic Research May 2020.
- Akerlof, George A.**, “Sins of Omission and the Practice of Economics,” *Journal of Economic Literature*, 2020, 58 (2), 405–418.
- Allgood, Sam, Lisa D. Cook, Nick Bloom, Sandra E. Black, Marianne Bertrand, Amanda Bayer, and Lee Badgett**, “AEA Professional Climate Survey: Final Report,” 2019.
- Andre, Peter and Armin Falk**, “What’s Worth Knowing? Economists’ Opinions about Economics,” IZA Discussion Paper No. 14527, 2021.
- Angus, Simon D., Kadir Atalay, Jonathan Newton, and David Ubilava**, “Geographic diversity in economic publishing,” *Journal of Economic Behavior & Organization*, 2021, 190, 255–262.
- Armstrong, J. Scott**, “Peer review for journals: Evidence on quality control, fairness, and innovation,” *Science and Engineering Ethics*, 1997, 3 (1), 63–84.
- Atal, Vidya**, “Do journals accept too many papers?,” *Economics Letters*, 2010, 107 (2), 229–232.
- Azar, Ofer H.**, “Rejections and the importance of first response times,” *International Journal of Social Economics*, 2004, 31 (3), 259–274.
- , “The Review Process in Economics: Is It Too Fast?,” *Southern Economic Journal*, 2005, 72 (2), 482.
- , “The Academic Review Process: How Can We Make it More Efficient?,” *The American Economist*, 2006, 50 (1), 37–50.
- Benos, Dale J., Edlira Bashari, Jose M. Chaves, Amit Gaggar, Niren Kapoor, Martin LaFrance, Robert Mans, David Mayhew, Sara McGowan, Abigail Polter, and et al.**, “The ups and downs of peer review,” *Advances in Physiology Education*, 2007, 31 (2), 145–152.
- Berk, Jonathan, Campbell Harvey, and David Hirshleifer**, “Preparing a Referee Report: Guidelines and Perspectives,” Working Paper, 2015.
- , —, and —, “A Checklist for Reviewing a Paper,” Duke I&E Research Paper No. 2017-03, 2016.
- , —, and —, “How to Write an Effective Referee Report and Improve the Scientific Review Process,” *Journal of Economic Perspectives*, 2017, 31, 231–244.
- Berquist, Thomas H.**, “Reviewer Assistance: How Important Is the Invitation Letter?,” *American Journal of Roentgenology*, 2014, 203 (3), 467–467.
- Blank, Rebecca M.**, “The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The American Economic Review,” *American Economic Review*, 1991, 81 (5), 1041–1067.
- Bolboacă, Sorana D., Diana-Victoria Buhai, Maria Aluaş, and Adriana E. Bulboacă**, “Post retraction citations among manuscripts reporting a radiology-imaging diagnostic method,” *PLOS ONE*, 2019, 14 (6), 1–14.
- Bravo, Giangiacomo, Francisco Grimaldo, Emilia López-Iñesta, Bahar Mehmani, and Flaminio Squazzoni**, “The effect of publishing peer review reports on Referee behavior in five scholarly journals,” *Nature Communications*, 2019, 10 (1).

- Brogaard, Jonathan, Joseph Engelberg, and Christopher A. Parsons**, "Networks and productivity: Causal evidence from editor rotations," *Journal of Financial Economics*, 2014, 111 (1), 251–270.
- Card, David and Stefano DellaVigna**, "Nine Facts About Top Journals in Economics," *Journal of Economic Literature*, 2013, 51.
- , — , and **Funk**, "What Do Editors Maximize? Evidence from Four Economics Journals," *The Review of Economics and Statistics*, 2020, 102, 195–217.
- , — , **Patricia Funk, and Nagore Iribarri**, "Are Referees and Editors in Economics Gender Neutral?," *The Quarterly Journal of Economics*, 2020, 135 (1), 269–327.
- Carrell, Scott E, David N Figlio, and Lester R Lusher**, "Clubs and Networks in Economics Reviewing," Working Paper 29631, National Bureau of Economic Research January 2022.
- Cherkashin, Ivan, Svetlana Demidova, Susumu Imai, and Kala Krishna**, "The Inside Scoop: Acceptance and Rejection at the Journal of International Economics," *Journal of International Economics*, 2008, 77, 120–132.
- Chetty, Raj, Emmanuel Saez, and Laszlo Sandor**, "What Policies Increase Prosocial Behavior? An Experiment with Referees at the Journal of Public Economics," *Journal of Economic Perspectives*, 2014, 28 (3), 169–88.
- Chu, Johan S. G. and James A. Evans**, "Slowed canonical progress in large fields of science," *Proceedings of the National Academy of Sciences*, 2021, 118 (41).
- Clain, Suzanne and Karen Leppel**, "Patterns in Economics Journal Acceptances and Rejections," *The American Economist*, 2017, 63 (1), 94–109.
- Coles, Peter, John Cawley, Phillip B. Levine, Muriel Niederle, Alvin E. Roth, and John J. Siegfried**, "The Job Market for New Economists: A Market Design Perspective," *Journal of Economic Perspectives*, December 2010, 24 (4), 187–206.
- Colussi, Tommaso**, "Social Ties in Academia: A Friend Is a Treasure," *The Review of Economics and Statistics*, 2018, 100 (1), 45–50.
- COPE Council**, "Author requests for certain experts not to be included in the editorial process (Case number 16-08)," *Committee on Publication Ethics*, 2016.
- , "Ethical guidelines for peer reviewers," *Committee on Publication Ethics*, 2017.
- , "Who 'owns' peer reviews?," *Committee on Publication Ethics*, 2017.
- Cotton, Christopher**, "Submission Fees and Response Times in Academic Publishing," *American Economic Review*, 2013, 103, 501–509.
- Cullen, Zoë B and Ricardo Perez-Truglia**, "The Old Boys' Club: Schmoozing and the Gender Gap," National Bureau of Economic Research Working Paper 26530, 2021.
- da Silva, Jaime Teixeira and Aceil Alkhatib**, "Should Authors be Requested to Suggest Peer Reviewers?," *Science and Engineering Ethics*, 2018, 24, 275–285.
- Dreber, Anna, Thomas Pfeiffer, Johan Almenberg, Siri Isaksson, Brad Wilson, Yiling Chen, Brian A. Nosek, and Magnus Johannesson**, "Using prediction markets to estimate the reproducibility of scientific research," *Proceedings of the National Academy of Sciences*, 2015, 112 (50), 15343–15347.
- Dupas, Pascaline, Alicia Sasser Modestino, Muriel Niederle, Justin Wolfers, and The Seminar Collective**, "Gender and the Dynamics of Economics Seminars," National Bureau of Economic Research Working Paper 28494, 2021.

- Ellison, Glenn**, “Evolving standards for academic publishing: A q-r theory,” *Journal of Political Economy*, 2002, 110, 994–1034.
- , “The slowdown of the economics publishing process,” *Journal of Political Economy*, 2002, 110, 947–993.
- , “Is peer review in decline?,” *Economic Inquiry*, 2010, 49 (3), 635–657.
- Ersoy, Fulya and Jennifer Pate**, “Invisible Hurdles: Gender and Institutional Bias in the Publication Process in Economics,” Working Paper, 2021.
- Eve, Katie, Bahar Mehmani, and Kate Wilson**, “How to improve reviewer diversity: a toolkit for editors (part I),” 2021. <https://www.elsevier.com/connect/reviewers-update/how-to-improve-reviewer-diversity-a-toolkit-for-editors-part-i>.
- , — , and — , “How to improve reviewer diversity: a toolkit for editors (part II),” 2021. <https://www.elsevier.com/connect/reviewers-update/how-to-improve-reviewer-diversity-a-toolkit-for-editors-part-ii>.
- , — , and — , “Peer review diversity in action,” 2021. <https://www.elsevier.com/connect/reviewers-update/peer-review-diversity-in-action>.
- Forscher, Patrick S., Markus Brauer, William Taylor Laimaka Cox, and Patricia G. Devine**, “How many reviewers are required to obtain reliable evaluations of NIH R01 grant proposals?,” Preprint, Center for Open Science April 2019.
- Frankel, Alexander and Maximilian Kasy**, “Which Findings Should Be Published?,” *American Economic Journal: Microeconomics*, 2021, Forthcoming.
- Freeman, Craig**, “Do Economic Journals Obey Economic Prescriptions?,” *Review of Industrial Organization*, 2000, 17, 371–384.
- Gneezy, Uri and Aldo Rustichini**, “A Fine is a Price,” *The Journal of Legal Studies*, 2000, 29 (1), 1–17.
- Goldberg, Pinelopi Koujianou**, “Report of the Editor: American Economic Review,” *American Economic Review*, 2012, 102 (3), 653–65.
- Goldstein, Itay, Stefan Nagel, and Toni Whited**, “Joint Statement by the Editors of the JF, RFS, and the Incoming Editor of the JFE Concerning Referee Confidentiality,” 2021. <https://www.journals.elsevier.com/journal-of-financial-economics/announcements/joint-statement-by-the-editors-of-the-jf-rfs-and-the-incomin>.
- Goodman, Steven N.**, “Manuscript Quality before and after Peer Review and Editing at Annals of Internal Medicine,” *Annals of Internal Medicine*, 1994, 121 (1), 11.
- Gross, Kevin and Carl T. Bergstrom**, “Why ex post peer review encourages high-risk research while ex ante review discourages it,” *Proceedings of the National Academy of Sciences*, 2021, 118 (51).
- Hadavand, Aboozar, Daniel Hamermesh, and Wesley Wilson**, “Is Scholarly Refereeing Productive (at the Margin)?,” National Bureau of Economic Research Working Paper 26614, 2020.
- , — , and — , “Publishing Economics: How Slow? Why Slow? Is Slow Productive? Fixing Slow?,” National Bureau of Economic Research Working Paper 29147, 2021.
- Hauser, Marc and Ernst Fehr**, “An Incentive Solution to the Peer Review Problem,” *PLoS biology*, 2007, 5 (4), e107.
- Heckman, James J. and Sidharth Moktan**, “Publishing and Promotion in Economics: The Tyranny of the Top Five,” *Journal of Economic Literature*, 2020, 58 (2), 419–470.
- Hengel, Erin**, “Publishing while female: Are women held to higher standards? Evidence from peer review,” Working Paper, 2021.

- Houry, Debra, Steven Green, and Michael Callaham**, "Does mentoring new peer reviewers improve review quality? A randomized trial," *BMC Medical Education*, 2012, 12 (1).
- Huisman, Janine and Jeroen Smits**, "Duration and quality of the peer review process: the author's perspective," *Scientometrics*, 2017, 113, 633–650.
- Johannesson, Magnus and Carl Mellström**, "Crowding Out in Blood Donation: Was Titmuss Right?," *Journal of the European Economic Association*, 2008, 6, 845–863.
- Jung, Jisoo, Joo-Im Kim, and Ji Won Yoon**, "A Practical Approach to Constructing Triple-Blind Review Process with Maximal Anonymity and Fairness," *Information Security Applications*, 2017, p. 198–209.
- Khoo, Shaun**, "There is little evidence to suggest peer reviewer training programmes improve the quality of reviews," *Impact of Social Sciences Blog*, 2018.
- Kube, Sebastian, Michel André Maréchal, and Clemens Puppe**, "The Currency of Reciprocity: Gift Exchange in the Workplace," *American Economic Review*, 2012, 102 (4), 1644–1662.
- Laband, David N.**, "Is There Value-Added from the Review Process in Economics?: Preliminary Evidence from Authors," *The Quarterly Journal of Economics*, 1990, 105 (2), 341.
- Lemon, Kaelyn, Erika Mann, and Anna Simonin**, "Multidisciplinary Peer Review at PLOS ONE – improving the quality of peer review for interdisciplinary research," *EveryONE*, 2019.
- Mahmud, Tamim, B Hossain, and Dilshad Ara**, "Automatic Reviewers Assignment to a Research Paper Based on Allied References and Publications Weight," in "2018 4th International Conference on Computing Communication and Automation (ICCCA)," 2018, pp. 1–5.
- Maniadis, Zacharias and Fabio Tufano**, "The Research Reproducibility Crisis and Economics of Science," *The Economic Journal*, 2017, 127 (605), F200–F208.
- Mason, Paul M., Jeffrey W. Steagall, and Michael M. Fabritius**, "Publication delays in articles in economics: what to do about them," *Applied Economics*, 1992, 24, 859–874.
- Mavrogenis, Andreas, Quaile Andrew, and Marius Scarlat**, "The good, the bad and the rude peer-review," *International Orthopaedics*, 2020, 44.
- McAfee, R.**, "Edifying Editing," *American Economist*, 03 2010, 55.
- McDowell, Gary S, John D Knutsen, June M Graham, Sarah K Oelker, and Rebeccah S Lijek**, "Co-reviewing and ghostwriting by early-career researchers in the peer review of manuscripts," *eLife*, 2019, 8.
- Milkman, Katherine L., Modupe Akinola, and Dolly Chugh**, "What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations.," *Journal of Applied Psychology*, 2015, 100 (6), 1678–1712.
- Moore, Jessica L, E. Neilson, and Vivian Siegel**, "Effect of recommendations from reviewers suggested or excluded by authors.," *Journal of the American Society of Nephrology : JASN*, 2011, 22 9, 1598–602.
- Mulligan, Adrian, Louise Hall, and Ellen Raphael**, "Peer review in a changing world: An international study measuring the attitudes of researchers," *Journal of the American Society for Information Science and Technology*, 2013, 64 (1), 132–161.
- Nagel, Stefan**, "Report of the Editor of the "Journal of Finance" for the Year 2016," *The Journal of Finance*, 2017, 72 (4), 1859–1874.
- Nature**, "Nature will publish peer review reports as a trial," *Nature*, 2020, 578 (7793), 8–8.

NBER, “Improving grant peer review within a science of science funding,” 2021. <https://www.nber.org/programs-projects/projects-and-centers/8089-improving-grant-peer-review-within-science-science-funding>.

— , “NBER studies related To Covid-19 PANDEMIC by topic area,” 2021. <https://www.nber.org/nber-studies-related-covid-19-pandemic-topic-area>.

Nosek, Brian A., Tom E. Hardwicke, Hannah Moshontz, Aurélien Allard, Katherine S. Corker, Anna Dreber, Fiona Fidler, Joe Hilgard, Melissa Kline Struhl, Michèle B. Nuijten, Julia M. Rohrer, Felipe Romero, Anne M. Scheel, Laura D. Scherer, Felix D. Schönbrodt, and Simine Vazire, “Replicability, Robustness, and Reproducibility in Psychological Science,” *Annual Review of Psychology*, 2022, 73 (1), 719–748. PMID: 34665669.

Ofosu, George K. and Daniel N. Posner, “Pre-Analysis Plans: An Early Stocktaking,” *Perspectives on Politics*, 2021, p. 1–17.

Olken, Benjamin A., “Promises and Perils of Pre-Analysis Plans,” *Journal of Economic Perspectives*, 2015, 29 (3), 61–80.

Özler, Berk, “Working Papers are NOT Working..,” *World Bank Blogs*, 2011.

Porter, Harry, “Community comments continued: An update on the preprint commenting pilot at PLOS,” *The Official PLOS Blog*, 2020.

Powdthavee, Nattavudh, Yohanes E. Riyanto, and Jack L. Knetsch, “Lower-rated publications do lower academics’ judgments of publication lists: Evidence from a survey experiment of economists,” *Journal of Economic Psychology*, 2018, 66, 33–44.

Prufer, J. and D. Zetland, “An Auction Market for Journal Articles,” *Public Choice*, 02 2007, 145.

Redden, Elizabeth, “Fast pace of scientific publishing on Covid comes with problems,” *Inside Higher Ed*, 2020.

Reimers, Imke and Joel Waldfogel, “Digitization and Pre-purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings,” *American Economic Review*, 2021, 111 (6), 1944–71.

Schwert, G. William, “The remarkable growth in financial economics, 1974–2020,” *Journal of Financial Economics*, 2021, 140 (3), 1008–1046.

Silberner, Joanne, “Scientists say the rush to do covid research led to a whole lot of waste,” *National Public Radio*, 2021.

Silbiger, Nyssa and Amber Stubler, “Unprofessional peer reviews disproportionately harm underrepresented groups in STEM,” *PeerJ*, 2019, 7.

Snodgrass, Richard, “Single- Versus Double-Blind Reviewing: An Analysis of the Literature,” *SIGMOD Record*, 2006, 35, 8–21.

Squazzoni, Flaminio, Giangiacomo Bravo, and Karoly Takacs, “Does Incentive Provision Increase the Quality of Peer Review? An Experimental Study,” *Research Policy*, 2013, 42, 287–294.

Tennant, Jonathan, Jonathan Dugan, Daniel Graziotin, Damien Jacques, Francois Waldner, Daniel Mietchen, Yehia Elkhatib, Lauren Collister, Christina Pikas, Tom Crick, Paola Masuzzo, Anthony Caravaggi, Devin Berg, Kyle Niemeyer, Tony Ross-Hellauer, Sara Mannheimer, Lillian Rigling, Daniel S. Katz, Bastian Tzovaras, and Julien Colomb, “A multidisciplinary perspective on emergent and future innovations in peer review,” *F1000Research*, 11 2017, 6, 1151.

Thompson, Gary D., Satheesh V. Aradhyula, George Frisvold, and Russell Tronstad, “Does Paying Referees Expedite Reviews? Results of a Natural Experiment,” *Southern Economic Journal*, 2010, 76 (3), 678–692.

U.S. DOJ and FTC, “Horizontal Merger Guidelines (08/19/2010),” 2015.
<https://www.justice.gov/atr/horizontal-merger-guidelines-08192010>.

van Rooyen, S., T. Delamothe, and S. J. Evans, “Effect on peer review of telling reviewers that their signed reviews might be posted on the web: randomised controlled trial,” *BMJ*, 2010, 341 (nov16 2), c5729–c5729.

Vilhuber, Lars, “Report by the AEA Data Editor,” *AEA Papers and Proceedings*, 2019, 109, 718–729.

Welch, Ivo, “Referee Recommendations,” *The Review of Financial Studies*, 2014, 27 (9), 2773–2804.

Wu, Alice, “Gender Bias in Rumors among Professionals: An Identity-Based Interpretation,” *The Review of Economics and Statistics*, 10 2019, 102, 1–40.

Zaharie, Monica Aniela and Marco Seeber, “Are non-monetary rewards effective in attracting peer reviewers? A natural experiment,” *Scientometrics*, 2018, 117 (3), 1587–1609.

Zheng, Yuqing and Harry Kaiser, “Submission demand in core economics journals: A panel study,” *Economic Inquiry*, 2015, 54.

Appendix A Additional tables and data discussions

This section of the appendix is devoted to presenting additional tables of importance and explaining the construction of certain tables in the report in greater detail.

A.1 Full proposals table

Table A.1: Proposals summary

ID	Proposal Description	Proposal Location	% In Favor or Useful ⁺	N ^a	Action-taker(s) ^b
Section 3.2: Allocating papers to journals, referees, and editors					
1	Centralized platform	3.2.4.1	49.2%	14	All
2	Link referees to ORCID	3.2.4.2		0	J, R
3	Internal transfers within journal families	3.2.4.3		2	J, E, R
4	Conditional acceptance offers within journal families	3.2.4.3		1	J, E
5	Intra-publisher journal partnership to share reports	3.2.4.4		5	J
6	Global limit on refereeing requests	3.2.4.1		5	All
7	Referees can only review one paper at a time for a given journal	3.2.4.13		1	J, E, R
8	Self-imposed limits or minimum contribution for refereeing	Comment		6	R
9	Invite referees to update profile annually	3.2.4.5		2	J, R
10	Avoid sending reminders during time off	Comment		1	J, E
11	Allow authors to suggest referees or editors	3.2.4.9		6	J, A
12	Allow authors to oppose referees or editors	3.2.4.10	56.4%	7	J, A
13	Double-blind reviewing as standard	3.2.4.11	30.1%	17	J
14	Double-blind reviewing as option	3.2.4.11		0	J, A
15	Triple-blind reviewing as standard	Footnote 30		3	J
16	Enact measures to make submissions truly double-blind	Comment		6	E
17	Discourage authors from guessing referee identity	Comment		2	J, E
18	System for making anonymous references to referee's own work	Comment		1	J, E, R
19	End preferential treatment for connected authors	Comment		23	E, R
20	Recruit editorial members from underrepresented groups	3.2.4.12		9	J
21	Standardized or capped editorial terms	3.2.4.12		4	J
22	Do not hire current editors from other journals	3.2.4.12		1	J
23	Disclose all editorial communications and conflicts to referees	Comment		2	E
24	Editors explain to referees how reports affected decision	Comment		10	E
25	Editors make independent manuscript decisions more often	Comment		27	E
26	Editors follow referee suggestions more often	Comment		8	E
27	Editors put more weight on high-quality reports over others	Comment		13	E
28	Editors explain rationale for not desk rejecting and referee selection	Comment		6	E
29	Clear editorial decision rules to accept/reject based on reports	Comment		3	E
30	More informal editor/referee communication channels	Comment		2	J, E, R
31	Choose diverse pool of referees	3.2.4.12		6	E
32	Clear conflict-of-interest guidelines for referees	3.2.4.8		8	J
33	Employ non-expert academics as general interest referees	Comment		4	E
34	Only employ expert referees from within same field	Comment		15	E
35	Have referees focus on specific aspect of manuscript	3.2.4.6		12	E, R
36	Have referees specialize in specific types of papers	3.2.4.6		0	E, R
37	More editorial specialization and division of labor	3.2.4.7		0	J, E
38	Assign junior researchers more responsibility in peer review	Comment		5	E, R
39	Put less pressure on juniors to publish and review	Comment		6	E, U
40	Match seniority of referee to author	Comment		1	E, R, A
41	Option to indicate time-sensitive submissions for tenure track	Comment		4	J, E, A
Section 3.3: Content of the reports					
42	Provide report guidelines and evaluation criteria	3.3.3.1	52.1% ⁺	39	J, E
43	Publish papers with grade to indicate quality	Comment		2	J, E, R
44	Provide referee report template	3.3.3.1	46.3%	5	J, E
45	Provide examples of high-quality referee reports	3.3.3.1		0	J, E
46	Allow or require shorter referee reports	Comment		17	J, E, R
47	Allow referees flexibility in report format and substance	Comment		3	J, E, R
48	Have referees rate manuscripts on specific evaluation criteria	3.3.3.1		4	J, E, R
49	Narrow referee focus to manuscript's substance and/or contribution	Comment		36	J, E, R
50	Request that referees focus on comments essential for publication	Comment		30	J, E, R
51	Have referees offer constructive feedback for improving manuscripts	Comment		6	J, E, R
52	Referee training in doctoral programs	3.3.3.2	54.9% ⁺	7	A
53	Co-reviewing between junior and senior researchers	3.3.3.2		0	J, R
54	Frequent informal feedback from editors to referees	3.3.3.3		26	E
55	Formal editorial evaluation or grading system for referees	3.3.3.3		14	J, E
56	Feedback from authors to referees about report quality	Comment		12	J, E, A

Continued on next page

57	Share decision letters and referee reports with referees	3.3.3.3	13	E
58	Make referee reports available to all journal's editors	Comment	1	J, E, R
59	Share referee's letter to the editor with authors	Comment	2	J, E, R
60	Ask referees (and others) to abide by code of conduct	3.3.3.4	2	E, R
61	Allow editing of inappropriate referee comments	3.3.3.3	7	J, E
62	Do not transfer low-quality reports to authors	3.3.3.3	7	J, E
63	Punish referees for reports with inappropriate comments	3.3.3.3	11	J, E, R
64	Avoid using referees who reviewed the same manuscript elsewhere	Comment	6	J, E, R
65	Have referees state their relevant expertise in referee report	Comment	2	J, E, R
66	Do not send revisions to referees who recommended rejection	Comment	7	J, E
67	Do not allow referees or editors to suggest irrelevant self-citations	Comment	4	J, E, R
68	Allow author response to decisions and reports	3.3.3.5	57.5%	22 J, E, R, A
69	Remove anonymity during appeal process to deter low-quality reports	Comment	1	J, E, A
70	Editors request additional reports when others are inadequate	Comment	7	J, E, R, A
71	Invite authors to give a revision proposal before issuing an R&R	Comment	1	J, E, A

Section 3.4: Reviewing process and decision times

72	Calibrate submission costs to screen out unsuitable manuscripts	3.4.2.1	6	J
73	Cap total submissions per manuscript and/or author	3.4.2.1	1	J
74	Reduce submission fees	Comment	1	J
75	Partial or full submission fee refunds for desk rejections	Comment	3	J
76	Allow authors to simultaneously submit to several journals	3.4.2.2	2	J, A
77	Refund authors and let them resubmit elsewhere after excessive delay	Comment	1	J, E, A
78	Have referees and editors suggest alternative journals after rejection	Comment	1	E, R, A
79	Impose deadline on editors to decide after receiving reports	Comment	1	J, E
80	Increase the desk rejection rate	3.4.2.3	31	J, E
81	Decrease the desk rejection rate	Comment	9	J, E
82	Committee to evaluate submission abstracts for desk rejections	Comment	1	J, E
83	Automated search tools to find referees	3.4.2.5	4	J, E
84	Have referees use non-institutional or regularly updated contact email	3.4.2.4	0	R
85	Keep track of referee holiday/sabbatical times	3.4.2.4	0	E, R
86	Clear decision rules for accepting referee requests	3.4.2.4	4	J, E, R
87	Communicate expected response time to referee requests	3.4.2.4	0	J, E
88	Allow expedited reviews for clear rejections	3.4.2.6	5	J, R
89	Reduce default turnaround time for referee reports	3.4.2.7	10	J, E
90	Longer or adjustable turnaround times for referee reports	Comment	11	J, E
91	Set stronger norms around turnaround times	Comment	9	J, E
92	Cash payments to referees for timely reports	3.4.2.8	54.6%	35 J
93	Cash payments or other rewards to referees for quality reports	Comment	11	J
94	Choice of charity donations instead of referee payments	3.4.2.8	0	J
95	Submission fee waivers or discounts to referees	3.4.2.8	20.5%	2 J
96	Free journal access for referees	Comment	3	J
97	Incentives to referee at lower-ranked journals	Comment	1	J
98	Acknowledge names of all referees in annual report	3.4.2.9	6	J, E
99	Acknowledge names of exceptional referees in annual report	3.4.2.9	49.4%	18 J, E
100	Make acknowledgment of exceptional referees more exclusive	Comment	1	J, E
101	Personalized expressions of gratitude for good reviews	Comment	8	E
102	Certificates for refereeing	3.4.2.9	1	J
103	Give best referees extra publication opportunities at journals	Comment	2	J, E
104	Give better treatment to submissions from journal's referees	Comment	5	J, E
105	Faster turnaround time for best referees' manuscripts at journals	Comment	5	J, E
106	Allow best referees to appeal decisions on own submissions	Comment	2	J, E
107	Make referees co-authors for substantial contributions	Comment	2	J, E, R, A
108	Invite referees to conferences	Comment	1	J, E
109	Give best referees better editorial and refereeing opportunities	Comment	4	J, E
110	Reward best referees with fewer future requests	Comment	1	J, E
111	Better compensation for editors	Comment	2	J, E
112	Making individual turnaround times public	3.4.2.11	2	J, E
113	Assign referees' manuscript submissions to similar quality referees	3.4.2.10	10	J, E
114	Delay editorial review of late referees' manuscript submissions	3.4.2.10	2	J, E
115	Other negative incentives for bad referees in submission process	3.4.2.10	6	J, E
116	Limit manuscript length (e.g., short paper track)	3.4.2.12	76.3% ^c	12 J
117	More conditional acceptances (fewer R&Rs)	3.4.2.12	8	E
118	Limit number of review rounds	3.4.2.12	19	J, E
119	Notify referees about author withdrawals in timely manner	Comment	1	E
120	Streamlined process for receiving and submitting reports	Comment	8	J, E
121	Cap default manuscript revision period	3.4.2.13	0	J, E
122	Realign editorial incentives around revisions	3.4.2.15	26	E
123	Minimize or eliminate reject-and-resubmits	3.4.2.14	3	J, E

Section 3.5: Innovations in peer review

124	Solicit and test proposals to improve peer review	Comment	2	J, E
125	Share review history of published papers	3.5.2.1	52.1%	9 J, R
126	Share review history of rejected papers	3.5.2.2	6	J, R
127	Publish referees' names on accepted papers	3.5.2.3	4	J, R

Continued on next page

128	Publish referees' names and link to their reports	3.5.2.3	27.2%	1	J, R
129	Publish senior referees' names	3.5.2.4	22.8% ⁺	0	J, R
130	Referees sign their reports	3.5.2.4		0	R
131	Publish handling editor name	3.5.2.3		6	J, E
132	Public reputation scores for referees based on performance	Comment		18	J, E, R, A
133	Public evaluations of editorial performance	Comment		2	J, R, A
134	Publish more journal performance statistics	Comment		9	J, E
135	Better editorial explanations to authors about decisions	Comment		3	E
136	Remove editors who are late, biased, or inconsistent	Comment		5	J, E
137	Remove low-quality referees from the referee pool	Comment		6	J, E
138	Post-publication peer review	3.5.3.1		3	J, E, R
139	Fast-tracks for submissions	3.5.3.1		0	J
140	Set norms in economics emphasizing high-quality and bold research	Comment		11	All
141	Publish more papers	Comment		4	J, E, R
142	Peer review of registered reports	3.5.3.2		6	J
143	Modify evaluation criteria to encourage replication papers	Comment		2	J, E
144	Prediction markets for conducting/replicating research	3.5.3.6		0	J
145	De-emphasize use of peer review in scientific publishing	Comment		8	All
146	De-emphasize importance of top 5 in prestige and career	Comment		3	All
147	Change the ranking of journals to emphasize alternative metrics	Comment		2	All
148	Consolidate number of journals	Comment		2	J
149	Journal ownership by not-for-profit and academic organizations	Comment		1	J, A
150	Limit use of special issues and guest editors	Comment		1	J, E
151	Elect editors or define clear metric to become one	Comment		3	J, E
152	Professionalized full- or part-time referees	3.5.3.8		2	J, R
153	Use more referees (or editors) per paper	Comment		7	J, R
154	Use fewer referees (or editors) per paper	Comment		5	J, R
155	Crowdsourced peer review	3.5.3.3		5	J, E, R, A
156	Non-academics (e.g., policymakers) as referees	3.5.3.4		2	J, E, R
157	Non-economists as referees	3.5.3.5		0	J, E, R
158	Platform for authors to solicit feedback without journal submission	3.5.3.9		1	A
159	Referee reports as part of academic portfolio	3.5.3.7		2	R, A
160	Recognition of peer review in tenure and promotion decisions	3.5.3.8		32	R, A
161	More interactive review process	3.5.3.10		15	J, E, R, A
162	Allow discussion and evaluation between referees	3.5.3.10		5	J, E, R
163	Author- and third party-facilitated peer review	3.5.3.11		0	R, A
164	Centralized manuscript processing and allocation	3.5.3.12		0	All

⁺ The fourth column either reports the percentage of our respondents in favor of a proposal or the percent who find it useful; the latter is indicated by the ⁺ symbol.

^a The fifth column reports the number of respondents who we interpreted as being in favor of a given proposal from reading their comments. If a respondent supported a proposal across multiple comments, we only counted their support once.

^b This column indicates which peer review actors would need to take action in order to implement a given proposal: J = Journals; E = Editors; R = Referees; A = Authors; U = Universities.

^c We report the percentage of respondents in favor of the *AER: Insights* model as in favor of limiting manuscript length, but we note that the model has other key features, including a limit of one review round and the use of conditional acceptances in place of R&Rs.

A.2 Andre and Falk data and discussion (Tables 3 and A.2)

At various points in our report, we use data presented in Andre and Falk (2021) or statistics derived by Peter Andre from the same data source (henceforth, AF). First, in Paragraph 2.2.2 (Table 3), we compared the characteristics of our sample with those of (i) the AF study population of academic economists and (ii) the AF survey sample. Next, in Paragraph 3.2.3, we compared the study population with the subpopulations of (i) editors at the top 50 journals and (ii) referees for the top 5 journals. That was done using Table A.2, which is presented below. A discussion of the variables presented in these tables follows on the next page.

Table A.2: Comparison of study population to top 5 referees and editors

	Study population	Top 5 referee population ^a	Top 50 editor population ^b
Demographics			
Female	26.0%	16.8%	19.5%
<i>Location:</i>			
US/Canada	33.9%	60.2%	72.1%
Europe	40.4%	33.7%	22.4%
Asia/Oceania	21.4%	5.2%	4.9%
Other regions	4.3%	1.0%	1.0%
Professional Status			
Top 50 institution	12.1%	42.3%	46.7%
Number of unique co-authors	5.8	9.5	11.7
Number of co-authors with top 5 publications	0.5	4.0	4.1
Eigenvector centrality (index)	0.61	0.83	0.82
Publications			
Year of first publication (average)	2007	2002	1998
Average number of publications (in pub. sample) ^c	4.8	9.4	10.7
Average number of publications (overall)	17.1	24.0	32.8
Share of articles in econ journals	75.9%	85.1%	80.6%
Publications in top 5 Journal (in pub. sample) ^c	6.1%	60.4%	51.4%
Average journal rank 1-400 (in pub. sample) ^c	164.2	63.0	53.2
h-index	6.5	11.6	16.6
N	53,779	4,229	2,818

Notes: All of the data in this table is based on the data presented in [Andre and Falk 2021](#), derived by Peter Andre. For descriptions of the construction of these variables, see the next page.

^a “Top 5 referee population” consists of the researchers who repeatedly refereed for top 5 journals from 2015-2020. The data comes from the published list of referees or list of excellent referees from the top 5 journals. “Repeatedly” is defined as a researcher who appears on two or more of these lists of referees (at least two different top 5 journals or in at least two different years).

^b “Top 50 editor population” are those who served in editorial positions at the 50 highest-ranked journals in the 2020 Shanghai Academic Ranking of World Universities for Economics.

^c An explanation of the “publication sample” can be found in footnote 6 of Section [2.2.2](#).

Below we elaborate on the definitions of certain variables and note any instances where variables had to be reformulated in order to ensure the comparability of our data with the external data.

- **Location:** Found in Tables [3](#) and [A.2](#). From the AF statistics, we combine the “Asia” and “Australia and New Zealand” categories to create the “Asia/Oceania” category and we combine the “Latin America” and “Africa” categories to create the “Other regions” category.
- **Field of research:** Found in Table [3](#). The field categories from AF are based on the JEL codes:
 - Microeconomics = JEL D (Microeconomics)
 - Macroeconomics = JEL E (Macroeconomics and Monetary Economics) + JEL F (International Economics) + JEL G (Financial Economics)
 - Econometrics = JEL C (Mathematical and Quantitative Methods)
 - Development = JEL O (Growth and Development Economics)
 - Labor = JEL J (Labor and Demographic Economics)
 - Industrial Organization = JEL L (Industrial Organization)

- Public Economics = JEL H (Public Economics)
- Other fields = JEL Q (Agriculture and Environmental Economics) + Other fields
- **Position:** Found in Table 3. The “Full Professor,” “Postdoc/PhD,” and “Other” categories combine the AF categories of “Professor” and “Emeritus,” the categories of “Post-doc” and “Doctoral student,” and the categories of “Graduate student” and “Other,” respectively.
- **Average number of publications:** Found in Tables 3 and A.2. The number of publications is capped at 200.
- **Eigenvector centrality (index):** Found in Table A.2. Eigenvector centrality gives a relative score to each author based on the number and strength of co-author connections they have, with co-author connections with higher scores contributing more to that author’s score. Here, the eigenvector centrality index shows the share of authors with lower eigenvector centrality. For example, an author with an eigenvector centrality index of 80 has a higher eigenvector centrality (and thus a “stronger” network) than 80% of authors in the index. It also implies that this author is connected to other authors who have high scores themselves.
- ***h*-index:** Found in Table A.2. The variable *h*-index is a measure of the research impact of an author. The value of *h* is specific to each author and is the highest value for which the statement “this author has published at least *h* papers that have been each cited a minimum of *h* times” is true. For example, an author with 3 papers that have each been cited 3 times has an *h* of 3. If one of those papers had only received two citations, their *h*-index is only 2.

A.3 Top 5 table sourcing discussion

We also frequently refer to Table 4 in the report. Since it draws on information from a wide range of diverse sources, information about data sourcing and definitions is available here:

- The five journals are highly heterogeneous in the manner and extent to which they release information about submission and editorial activities. The *AER* and *Econometrica* release annual reports that provide a range of indicators. The *JPE* publishes rolling statistics from the past two years on its website, while *REStud* shares data online at about half-year time periods. Finally, the *QJE* shares a select few statistics about the previous year’s activity on Twitter each January. We contacted editors from each journal to confirm the information provided in the table and request additional data.
- The **Editorial Structure** section is based on journals’ editorial board pages as of mid-September 2021 ([AER-9](#), [ECMA-6-a](#), [JPE-5](#), [RES-6](#), [QJE-3](#)). We consulted with *REStud*’s and *JPE*’s editors to better understand their editorial structure. Per *REStud*, we count all Joint Managing Editors as Coeditors and all editorial board members (Chair, Board of Directors, Data Editor, Secretary and Business Manager, Editorial Board, Foreign Editors, and Journal Manager) as Associate Editors. Per *JPE*, we count their Advisory Board members and Managing Editor as Other Editors.
- In the **Submission Policies** section, information about submission requirements, fees, and discounts comes from each journal’s submission guidelines or editorial policy pages ([AER-6](#), [ECMA-4](#), [JPE-3-a](#), [RES-4](#), [QJE-1](#)). Membership or subscription fee amounts come from the journals’ pages to sign up for the services ([AER-7](#), [ECMA-5](#), [JPE-4](#), [RES-5](#), [QJE-2](#)).
- In the **Submission Statistics** section, we report annual submission, resubmission, R&R numbers, and desk rejection (DR) rates from the latest journal reports, when available ([AER-1-c](#), [ECMA-2-a](#), [RES-3-b](#)). *QJE*’s submission volume and DR rate were taken from an official [tweet](#), while their resubmission and R&R volumes were supplied by the editors. *REStud*’s DR rate and R&R volume were communicated to us by the editors. We also received *JPE*’s statistics directly from the editors. We calculate *Econometrica*’s DR rate from submission and decision volumes. Resubmissions at the *AER* are calculated as the sum of new revisions under review, revisions given acceptances, and conditional acceptances awaiting data approval in 2019-2020, from Table 3 in their report. The only R&R statistics available from the *AER* are from 2018-2019, as included in their latest report.

Report Years for Submission Statistics are from the most recent period available, but the exact date range varies between journals. The *AER*'s report year is November 1, 2019 - October 31, 2020. *Econometrica*'s is July 1, 2019 - June 30, 2020. *JPE*'s is January 1, 2020 - December 31, 2020. *REStud*'s is September 1, 2019 - August 31, 2020. Finally, *QJE*'s is October 1, 2019 - September 30, 2020.

- In the **Policies and Guidelines** section, we note whether there are editor- or reviewer-specific conflict-of-interest guidelines made publicly available by the journal and whether they offer advice or instructions to referees about how to write reports ([AER-2-a](#), [AER-3-a](#), [ECMA-7](#), [ECMA-9](#), [JPE-2](#), [RES-4](#), [QJE-5](#)). *Econometrica* provides its own official guidelines about what they expect from referee reports, while we label the *AER*'s unofficial because the guidelines they link to "do not represent official AEA policy."
- In the **Referee Incentives and Recognition** section, information mostly comes from journal information pages ([AER-3-a](#), [JPE-3-a](#)). According to its editors, the *JPE* offers reviewers who submit a timely report the choice between a \$100 cash compensation or a two-year subscription to the journal. The number of recognized and exceptional referees named each year typically comes from periodic reports or issues ([AER-1-c](#), [ECMA-8](#), [RES-7](#), [QJE-4](#)). The number of referees who assisted the *JPE*, *REStud*, and *QJE* in 2019-2020 was given to us by their respective editors.
- In the **Turnaround Statistics** section, numbers are mostly sourced from https://jcsuarez.shinyapps.io/journal_turnaround_app/ to ensure the comparability of the metrics for each journal. The exception is *JPE*, for which we use data they provided us instead, pursuant to their advice. Requested Referee Turnaround Time was taken from a cross-section of referee invitation emails obtained by the research team. We asked the journal editors to confirm that these are the default request times. Some noted that editors have the discretion to modify the requested deadline.
- For the **Author Appeal Policies** section, we scanned journal policies to see whether any official mention of appealing or contesting editorial decisions is made, and whether the mention makes it clear how authors can initiate the appeal process ([AER-8](#), [ECMA-3-a](#), [JPE-2](#)).

A.4 Journal population table sourcing discussion

In order to be precise about the population of journals under (primary) consideration, we present Table [A.3](#). At several points in the report, we refer to it in order to understand the prevalence of certain practices among economics journals, such as being a Publons partner. The table is on the next page and followed by a discussion on the page after that.

Table A.3: Journal population

Ranking ^a	Journal of Interest	Publisher	ORCID	Publons	2-year IF	5-year IF	h-index
1	<i>The Quarterly Journal of Economics</i>	OUP	✓		12.59	14.94	320
2	<i>Journal of Political Economy</i>	UChicago Press			9.65	8.24	307
3	<i>American Economic Review</i>	AEA			8.98	6.11	359
4	<i>Journal of Economic Perspectives</i>	AEA			6.95	7.77	194
5	<i>Journal of Economic Literature</i>	AEA			6.83	7.41	181
6	<i>Econometrica</i>	Wiley	✓		6.25	6.12	294
7	<i>American Economic Journal: Applied Economics</i>	AEA			6.13	6.34	85
8	<i>American Economic Journal: Macroeconomics</i>	AEA			6.11	6.7	73
9	<i>Review of Economics and Statistics</i>	MIT Press			5.47	5.54	183
10	<i>Review of Economic Studies</i>	OUP	✓		4.89	5.66	207
11	<i>American Economic Journal: Economic Policy</i>	AEA			4.76	5.26	66
12	<i>Journal of Labor Economics</i>	UChicago Press			4.71	5.22	134
13	<i>Journal of Finance</i>	Wiley	✓		4.3	5.49	295
—	<i>Journal of Economic Growth</i>	Springer	✓	✓	4.24	5.67	88
14	<i>Journal of Financial Economics</i>	Elsevier	✓		3.82	4.44	242
15	<i>Journal of Business & Economic Statistics</i>	TANDF	✓	✓	3.81	2.98	45
16	<i>American Economic Review: Insights</i>	AEA			3.79	3.79	8
17	<i>Journal of Monetary Economics</i>	Elsevier	✓		3.76	4.34	195
18	<i>Journal of International Economics</i>	Elsevier	✓		3.63	3.58	168
20	<i>Journal of the European Economic Association</i>	OUP	✓		3.52	4.28	26
23	<i>Review of Financial Studies</i>	OUP	✓		3.15	3.9	170
25	<i>Journal of Development Economics</i>	Elsevier	✓		3.06	3.22	160
26	<i>Journal of Public Economics</i>	Elsevier	✓		3.02	2.93	167
27	<i>Journal of Human Resources</i>	UW Press			2.97	5.61	110
28	<i>IMF Economic Review</i>	Springer	✓		2.96	2.94	41
29	<i>Journal of Economic Surveys</i>	Wiley	✓	✓	2.96	3.31	88
30	<i>Journal of Urban Economics</i>	Elsevier	✓		2.68	2.67	100
31	<i>Economic Journal</i>	OUP	✓		2.63	2.91	17
33	<i>JAERE^b</i>	UChicago Press			2.5	2.83	24
34	<i>Econometrics Journal</i>	OUP	✓		2.46	2.15	7
35	<i>Journal of Applied Econometrics</i>	Wiley	✓		2.45	2.68	112
36	<i>Journal of Accounting and Economics</i>	Elsevier	✓		2.38	2.65	117
37	<i>European Economic Review</i>	Elsevier	✓		2.35	2.6	151
41	<i>JEEM^b</i>	Elsevier	✓		2.24	2.42	119
42	<i>Quantitative Economics</i>	Wiley	✓		2.23	2.39	21
43	<i>Review of Finance</i>	OUP	✓		2.23	2.52	52
45	<i>Energy Economics</i>	Elsevier	✓		2.2	1.8	129
46	<i>Journal of Money, Credit and Banking</i>	Wiley	✓		2.16	2.49	36
48	<i>Journal of International Money and Finance</i>	Elsevier	✓		2.1	2.06	106
49	<i>Journal of Population Economics</i>	Springer	✓	✓	2.04	1.96	72
50	<i>JIFMIM^b</i>	Elsevier	✓		2.04	1.88	53
52	<i>Experimental Economics</i>	Springer	✓	✓	1.98	1.98	67
60	<i>Review of Economic Dynamics</i>	Elsevier	✓		1.84	2.28	78
64	<i>American Economic Journal: Microeconomics</i>	AEA			1.79	1.9	40
68	<i>Journal of Econometrics</i>	Elsevier	✓		1.66	1.85	199
84	<i>RAND Journal of Economics</i>	Wiley	✓		1.49	2.14	60
90	<i>Journal of Economic Dynamics and Control</i>	Elsevier	✓		1.44	1.49	105
119	<i>Journal of Economic Theory</i>	Elsevier	✓		1.28	1.34	143
138	<i>Journal of Economic Behavior & Organization</i>	Elsevier	✓		1.17	1.45	110
158	<i>Economics Letters</i>	Elsevier	✓		1.12	0.99	110
182	<i>Economic Inquiry</i>	Wiley	✓		1.00	1.14	44
196	<i>Management Science</i>	INFORMS	✓		0.97	1.4	151
198	<i>Theoretical Economics</i>	Wiley	✓		0.96	1.42	33
199	<i>Games and Economic Behavior</i>	Elsevier	✓		0.96	0.92	91
236	<i>Economic Theory</i>	Springer	✓	✓	0.78	0.68	62
—	<i>Journal of the Economic Science Association</i>	Springer	✓	✓	0.71	3.36	10
361	<i>PNAS</i>	USNAS	✓		0.49	0.49	7
563	<i>PLOS ONE</i>	PLOS	✓		0.12	0.12	25
	<i>The Lancet</i>	Elsevier	✓		79.32		762
	<i>Nature</i>	Springer	✓		49.96		1226
	<i>Science</i>	AAAS	✓		47.73		1186
	<i>British Medical Journal</i>	BMJ Publishing Group	✓		39.89		429
	<i>Nature Human Behaviour</i>	Springer	✓		13.66		46
	<i>Psychological Science</i>	SAGE	✓	✓	7.03		260
	<i>F1000 Research</i>	TANDF	✓				60
	<i>Frontiers</i>	Frontiers Media					

^a Notes and data sources: For information about the journals and statistics included in this table, please see this [discussion](#).^b Journal acronyms: JAERE = *Journal of the Association of Environmental and Resource Economists*; JEEM = *Journal of Environmental Economics and Management*; JIFMIM = *Journal of International Financial Markets, Institutions, and Money*.

There are many possible ways to construct a ranking of journals in economics. The discussion below explains our decisions and considerations about the information presented in the table.

- Our main inclusion criterion for journals was to be among the top 50 ranked by Impact Factor (IF) at <http://citec.repec.org/search.html> as of June 2, 2021. Any additional journals with ranks other than 1-50 were added at our discretion.
- The 2-year IF is calculated as the number of citations made in the current year of articles the journal published in the previous 2 years, divided by the number of articles. The 5-year IF is the same but calculated over 5 years instead.
- The h-index is a measure of the research impact of a journal. The value of h is the highest value for which the statement “this journal has published at least h papers that have been each cited a minimum of h times” is true. For example, an journal with 3 papers that have each been cited 3 times has an h of 3. If one of those papers had only received two citations, its h -index is only 2.
- Journals with ranks of “—” were omitted from the source’s ranking but had statistics that we took from elsewhere on the website. Ranks with blank spaces signify journals that do not publish economics output but were included for comparison purposes. The 2-year impact factor and h-index for the these journals were taken from Clarivate’s 2020 Journal Citation Reports (JCR) and from <https://www.scimagojr.com/journalrank.php>, respectively. *F1000Research* does not have an impact factor as it is not indexed in the Web of Science (**F1000-1**). We do not list statistics for *Frontiers* because it is actually split into dozens of distinct journals, primarily in the hard sciences, each of which have different statistics. None are particularly relevant to economics.
- We also applied an exclusion criterion to RePEc’s list to eliminate journals not sufficiently relevant to economics, namely the *Journal of Financial Intermediation*, *Journal of Multinational Financial Management*, *Journal of International Business Policy*, *Journal of African Business*, *Journal of Business Venturing*, *Research Policy*, *Journal of Financial Stability*, *Journal of International Business Studies*, *Journal of World Business*, and *African Development Review*.
- We note that the *IMF Economic Review* is published by Palgrave Macmillan, which is a subsidiary of Springer Nature.
- Journals with checkmarks in the ORCID column are those whose publishers are ORCID Member Organizations listed at <https://orcid.org/members>. Those with checkmarks in the Publons column are listed by Publons as Official Partners on their website at https://publons.com/journal/?order_by=reviews.

A.5 Journal sourcing table

Finally, we present Table A.4 on the next page as an exhaustive list of sources about journal statistics and policies that we cited throughout the report. We elected not to cite these sources in the References list, as they are not academic sources or news articles. Each source contains a reference to where it was used in the text and an external URL that links to its web location.

Table A.4: Sources for journal statistics and editorial policies

Journal Name	Source ID(s)	Source Name	Source Year	Source Link
<i>AEJ: Applied</i>	AEJ-1	Submission Guidelines	2021	AEJ-1
	AER-1-a , AER-1-b , AER-1-c AER-1-d	Report of the Editor	2021	AER-1
	AER-2-a , AER-2-b	AER: Editorial Policy	2021	AER-2
	AER-3-a , AER-3-b	AER: Information for Reviewers	2021	AER-3
	AER-4-a	Report of the Editor	2018	AER-4
<i>AER</i>	AER-5-a , AER-5-b , AER-5-c AER-6 AER-7 AER-8 AER-9 AER-10 AER-11	Report of the Editor AER: Submission Guidelines AEA Membership Dues and Information AER: FAQs Editors of the American Economic Review Data and Code Availability Policy FAQ Report of the Editor	2011 2021 2021 2021 2021 2021 2019	AER-5 AER-6 AER-7 AER-8 AER-9 AER-10 AER-11
<i>AER: Insights</i>	AERI-1 AERI-2	Submission Guidelines Report of the Editor	2021 2019	AERI-1 AERI-2
<i>BMJ</i>	BMJ-1 BMJ-2 BMJ-3	BMJ Article Transfer Service Resources for reviewers Instructions for reviewers	2021 2021 2021	BMJ-1 BMJ-2 BMJ-3
<i>Development</i>	DEV-1	Confidentiality	2021	DEV-1
<i>Econometrica</i>	ECMA-1-a , ECMA-1-b ECMA-2-a , ECMA-2-b , ECMA-2-c ECMA-3-a , ECMA-3-b ECMA-4 ECMA-5 ECMA-6-a , ECMA-6-b ECMA-7 ECMA-8 ECMA-9	Report of the Editors Report of the Editors Editorial Procedures and Policies Instructions for Submitting Articles Membership Editorial Board Conflict of Interest Policy Econometrica Referees 2019–2020 Guidelines for Referees	2011 2020 2021 2021 2021 2021 2021 2021 2021	ECMA-1 ECMA-2 ECMA-3 ECMA-4 ECMA-5 ECMA-6 ECMA-7 ECMA-8 ECMA-9
<i>Economic Inquiry</i>	EI-1	Manuscript Submission Instructions for Authors	2021	EI-1
<i>Economic Journal</i>	EJ-1 EJ-2-a , EJ-2-b EJ-3-a	Ethics Policies Instructions to Authors Economic Journal 2020 Annual Report	2019 2021 2020	EJ-1 EJ-2 EJ-3
<i>Economics Letters</i>	ECOLET-1	Guide for Authors	2021	ECOLET-1
<i>Energy Economics</i>	ENEC-1 ENEC-2	Guide for Authors Energy Economics	2021 2021	ENEC-1 ENEC-2
<i>Ergo</i>	Ergo-1	Peer Review Policy	2021	Ergo-1
<i>F1000Research</i>	F1000-1 F1000-2	FAQs How it Works	2021 2021	F1000-1 F1000-2
<i>Frontiers</i>	Front-1-a , Front-1-b , Front-1-c Front-2	Collaborative Peer Review Conflicts of Interest	2021 2021	Front-1 Front-2
<i>Harvard Law Review</i>	HLR-1	HOW TO SUBMIT	2022	HLR-1
<i>JDE</i>	JDE-1	JDE Pre Results Review Papers	3021	JDE-1
<i>JEEM</i>	JEEM-1	Special Issue: JEEM COVID-19 fast-track	3021	JEEM-1
<i>JEEA</i>	JEEA-1-a , JEEA-1-b , JEEA-1-c	Report of the Editors of JEEA	2021	JEEA-1
<i>Journal of Finance</i>	JF-1	Referee Guidelines for the Journal of Finance	2021	JF-1
<i>JFE</i>	JFE-1	Guide for Authors	2021	JFE-1
<i>JIE</i>	JIE-1	Guide for Authors	2021	JIE-1
<i>JME</i>	JME-1	The role of reviewers	2021	JME-1
<i>JPE</i>	JPE-1-a JPE-2 JPE-3-a , JPE-3-b JPE-4	JPE Turnaround Times, Previous Two Years Ethics Instructions for Authors Subscribe	2021 2021 2021 2021	JPE-1 JPE-2 JPE-3 JPE-4

Continued on next page

<i>JPE</i>	JPE-5	Editorial Board	2021	JPE-5
<i>JUE</i>	JUE-1	Announcing JUE: Insights	2021	JUE-1
<i>The Lancet</i>	LAN-1 LAN-2	Journey of a paper Publishing Excellence	2021 2021	LAN-1 LAN-2
<i>Management Science</i>	MS-1	Submission Guidelines	2021	MS-1
<i>Meta-Psychology</i>	MP-1	About the Journal	2021	MP-1
<i>Nature</i>	Nature-1 Nature-2 Nature-3-a , Nature-3-b , Nature-3-c Nature-4 Nature-5	Competing interests policy ORCID iDs at Springer Nature Peer Review Journal metrics Registered Reports	2021 2021 2021 2021 2021	Nature-1 Nature-2 Nature-3 Nature-4 Nature-5
<i>PLOS ONE</i>	PLOS-1 PLOS-2 PLOS-3 PLOS-4 PLOS-5 PLOS-6	ORCID at PLOS Submission Guidelines Our Peer Review Process Writing the review Editorial and Peer Review Process Guidelines for Reviewers	2021 2021 2021 2021 2021 2021	PLOS-1 PLOS-2 PLOS-3 PLOS-4 PLOS-5 PLOS-6
<i>PNAS</i>	PNAS-1-a , PNAS-1-b	For Reviewers	2021	PNAS-1
<i>Quantitative Economics</i>	QE-1	Submit A Paper	2021	QE-1
<i>QJE</i>	QJE-1 QJE-2 QJE-3 QJE-4 QJE-5	Manuscript Preparation Instructions Purchase Editorial Board Acknowledgement Journal Policies	2021 2021 2021 2021 2021	QJE-1 QJE-2 QJE-3 QJE-4 QJE-5
<i>REStud</i>	RES-1 RES-2 RES-3-a , RES-3-b RES-4 RES-5 RES-6 RES-7	Editors Report Editors Report Editors Report Submission Guidelines Purchase Editorial Board Excellence in Refereeing Award	2011 2014 2021 2021 2021 2021 2021	RES-1 RES-2 RES-3 RES-4 RES-5 RES-6 RES-7
<i>REStat</i>	REStat-1	Submission Guidelines	2021	REStat-1
<i>ROF</i>	ROF-1	Report from the Managing Editor	2020	ROF-1
<i>RFS</i>	RFS-1 RFS-2	Advice for Referees Change to Sub. Fees and Referee Payments	2021 2015	RFS-1 RFS-2
<i>Science</i>	SCI-1 SCI-2 SCI-3	Science: Information for authors Instructions for Reviewers of Research Articles Journal metrics overview	2021 2019 2021	SCI-1 SCI-2 SCI-3
<i>Theoretical Economics</i>	TE-1	Submitting a paper	2021	TE-1

Journal acronyms: *AER* = American Economic Review; *BMJ* = British Medical Journal; *JDE* = Journal of Development Economics; *JEEA* = Journal of the European Economic Association; *JEEM* = Journal of Environmental Economics and Management; *JFE* = Journal of Financial Economics; *JIE* = Journal of International Economics; *JME* = Journal of Monetary Economics; *JPE* = Journal of Political Economy; *JUE* = Journal of Urban Economics; *QJE* = Quarterly Journal of Economics; *REStud* = Review of Economic Studies; *REStat* = Review of Economics and Statistics; *RFS* = Review of Financial Studies.

Appendix B Description of variables and statistics

B.1 Variable codebook

Heterogeneity variables

- **Research field** - Based on the responses to [Q40](#) where respondents could select multiple fields (N = 1,459 respondents; 3,982 selections).
 - **Field:** Categorical variable with the following grouping of categories:
 - 1 - Behavioral/Experimental = Behavioral economics + Experimental economics
 - 2 - Applied microeconomics
 - 3 - Microeconomic theory = Decision theory + Game theory + Microeconomic theory
 - 4 - Macroeconomics = Financial economics + International economics + Macroeconomics
 - 5 - Econometrics = Applied econometrics + Econometric theory + Structural econometrics
 - 6 - Development economics
 - 7 - Labor economics
 - 8 - Industrial organization
 - 9 - Public economics
 - 10 - Political economy/economic history = Political economy + Economic history
 - 11 - Other = Urban economics + Other fields
 - * Used in Figure [B.4](#) (see below). Constructed as the percentage of respondents who selected each category (total is > 100% because of multiple selections).
 - * Used in Table [3](#) (removing 1 and pooling 2 & 3 under “Microeconomics” and 10 & 11 under “Other fields”). Constructed as the percentage of total selections (Total = 100%).
 - **Behav/Exp:** Binary variable = 1 if the respondent selected “Behavioral economics” or “Experimental economics.”
 - * Used in all heterogeneity and regression figures.
- **Gender identity** - Based on the responses to [Q41](#) (N = 1,459).
 - **Gender:** Categorical variable with 1 - “Male” and 2 - “Female” (excl. 3 - “Prefer not to say” (4.4%)).
 - * Used in Figure [12](#)
 - **Female:** Binary variable = 1 if the respondent selected “Female,” with the baseline being the combination of “Male” (72.0%) + “Prefer not to say” (4.4%).
 - * Used in Tables [3](#) & [C.2](#) and all heterogeneity and regression figures.
- **Professional status** - Based on the responses to [Q43](#) (N = 1,459).
 - **Position:** Categorical variable with 1 - “Full professor,” 2 - “Associate professor,” 3 - “Assistant professor,” 4 - “PhD/postdoc Candidate” (combination of “PhD candidate” and “Post-doctoral researcher” selections), and 5 - “Other category” (excl. 6 - “Prefer not to say”).
 - * Used in Tables [3](#) & [C.2](#) and Figure [19\(b\)](#).
 - * Used in Figures [E.7](#) and [E.8](#) (excl. 5 - “Other category”)
 - * Used in Figure [12](#) (pooling categories 3, 4, & 5 under “Other position”).
 - * Used in Figures [E.1](#) & [E.2](#) (with the combined categories 3, 4, & 5 as the baseline).
 - **Junior:** Binary variable = 1 if the respondent selected “PhD candidate” or “Post-doctoral researcher” or “Assistant professor.”
 - * Used in heterogeneity and regression figures.

- **Geographic location** - Based on the responses to [Q44](#) ($N = 1,459$). Distribution of responses shown in Figure [B.2](#).

N.b.: In the early rounds of data collection, the survey did not include “Prefer not to say” as an option and there are indications that some respondents selected countries that were not their actual location.

- **Location:** Categorical variable with 1 - “US/Canada,” 2 - “Europe,” 3 - “Asia/Oceania,” 4 - “Other countries” (excl. 5 - “Prefer not to say” (4.6%)).
 - * Used in Tables [3](#) & [C.2](#).
 - * Used in Figure [12](#) (pooling categories 3 & 4 under “Other regions”).
 - * Used in Figures [E.1](#) & [E.2](#) (with the combined categories 3 & 4 as the baseline).
- **US/Canada:** Binary variable = 1 if the respondent selected “United States” or “Canada.”
 - * Used in heterogeneity and regression figures.
- **Editorial responsibilities** - Based on the responses to [Q24](#) ($N = 1,483$). Heterogeneity of responses shown in Figures [12](#) & [E.1](#).
 - **Editor:** Binary variable = 1 if the respondent answered “Yes” to having held an editorial position.
 - * Used in Table [C.2](#) and all heterogeneity and regression figures.
- **Types of journals refereed for** - Based on responses to [Q23](#) ($N = 1,483$). Average responses shown in Figure [B.5](#).
 - **Journal type:** Categorical variable with 1 - “Top 5,” 2 - “Top field,” 3 - “Other economics,” and 4 - “Other discipline.”
 - * Used in Figure [4\(b\)](#).
 - **Percentage of top 5 reviews:** Continuous measure of the percentage of time respondents review for top 5 journals.
 - * Used in regression figures (demeaned).
 - **Top 5 reviewer:** Binary variable = 1 if the respondent indicated they had refereed for a top 5 journal ($> 0\%$).
 - * Used in Table [C.2](#), regression figure [E.2](#), and all heterogeneity figures.
- **Annual reports written** - Based on responses to [Q22](#) ($N = 1,483$). Distribution of responses shown in Figure [4\(a\)](#) (all respondents with ≥ 50 reports per year are grouped into the 50+ category).
 - **Reports written:** Continuous measure of the average number of reports written per year.
 - * Used in Figures [5\(b\)](#), [7\(a & b\)](#), & [8\(a\)](#).
 - * Used in Figure [26](#) (quartiles).
 - * Used in regression figures (demeaned).
 - **Active reviewer:** Binary variable = 1 if the respondent wrote more reports annually than the median respondent (> 8 reports).
 - * Used in all heterogeneity figures.
- **Annual average submissions** - Based on responses to [Q1](#) ($N = 1,484$). Because [Q1](#) asks for the number of submissions over a two-year timeframe, we divided the responses by 2 to get the annual average. Distribution and heterogeneity of responses shown in Panel (a) and (b) of Figure [2](#), respectively.
 - **Annual submissions:** Continuous measure of the annual average number of submissions.
 - * Used in Figure [4\(b\)](#) (quartiles).
 - **Active author:** Binary variable = 1 if the respondent made more submissions over the designated timeframe than the median respondent (> 3 submissions).
 - * Used in all heterogeneity figures.

Other descriptive variables

- **Age Range** - Based on responses to Q42 (N = 1,459). Distribution of responses shown in Figure B.1.
N.b.: To save space, we do not use it as a heterogeneity variable due to the high correlation between Under 40 and Junior ($\rho = 0.65, p < 0.01$).
 - **Age:** Categorical variable with 1 - “Under 40” (combination of “Under 30” (1.8%) and “30-39” categories), 2 - “40-49,” 3 - “50-59,” and 4 - “Over 60” (combination of “60-69” and “70+” (2.4%) categories) (excl. 5 - “Prefer not to say”).
 - * Used in Tables 3 & C.2.
 - **Under 40:** Binary variable = 1 if the respondent is under 40.
 - * Used in Figures E.1 & E.2.
- **Number of publications** - Based on responses to Q39 (N = 1,459). Distribution of responses shown in Figure B.3 (all respondents with ≥ 100 publications grouped into the 100+ category).
 - **Total publications:** Continuous measure of the total number of publications.
 - * Used in Table C.2 (median on full sample).
 - * Used in Table 3 (median with publications capped at 200).
 - * Used in regression figures (demeaned).

Selection of variables

For ease of exposition, we decided to present dichotomic measures for the various dimensions of heterogeneity presented in our heterogeneity figures. For the continuous measures that we dichotomized, we chose criteria that roughly split our sample in half. The percentage and number of respondents that belong to each category is shown in Table B.1. In our regression figures, we kept the dichotomized measures constructed from categorical variables in order to minimize the number of categories; on the other hand, we used the full information from our continuous measures and standardized them to facilitate comparisons and control the scale. We include the standardized continuous measure of **Total publications** along with the other heterogeneity variables. Some regression figures include additional covariates to control for other baseline differences (Figures E.15(b) & E.16). Two other regression figures (Figures E.1 & E.2), which each use a dimension of heterogeneity as the dependent variable, deviate from this standard by utilizing a categorical form of our **Position** and **Location** variables and by including the dichotomic **Under 40** variable. All the regression figures present the estimated coefficients with 95% confidence intervals.

- Heterogeneity Figures: 8(b), 13(b), 14, 15, 16, 19(a), 20, 21, 22(a), 23, 27, 30, 32(a & b), 33(b), & E.19(a & b)
- Regression Figures: E.1, E.2, E.3, E.4, E.5, E.6, E.9, E.10, E.12, E.13, E.14, E.15(b), E.16, & E.18

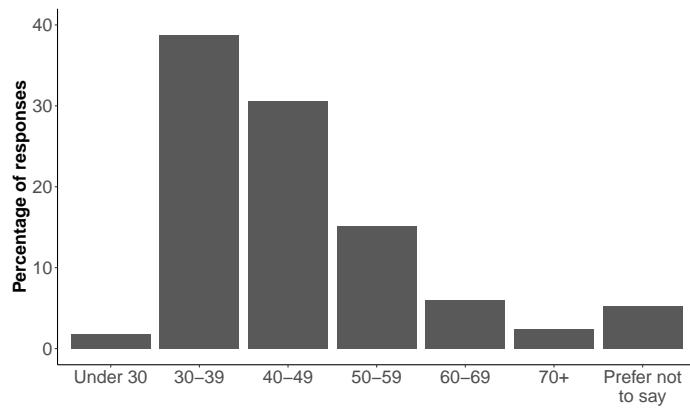
B.2 Descriptive statistics and figures

Table B.1: Descriptive statistics for the main dimensions of heterogeneity

Variable	N	%
Behav/Exp	802	55.0
Female	343	23.5
Junior	478	32.8
US/Canada	494	33.9
Editor	564	38.7
Top 5 reviewer	867	59.4
Active reviewer	657	45.0
Active author	578	39.6
All	1,459	100.0

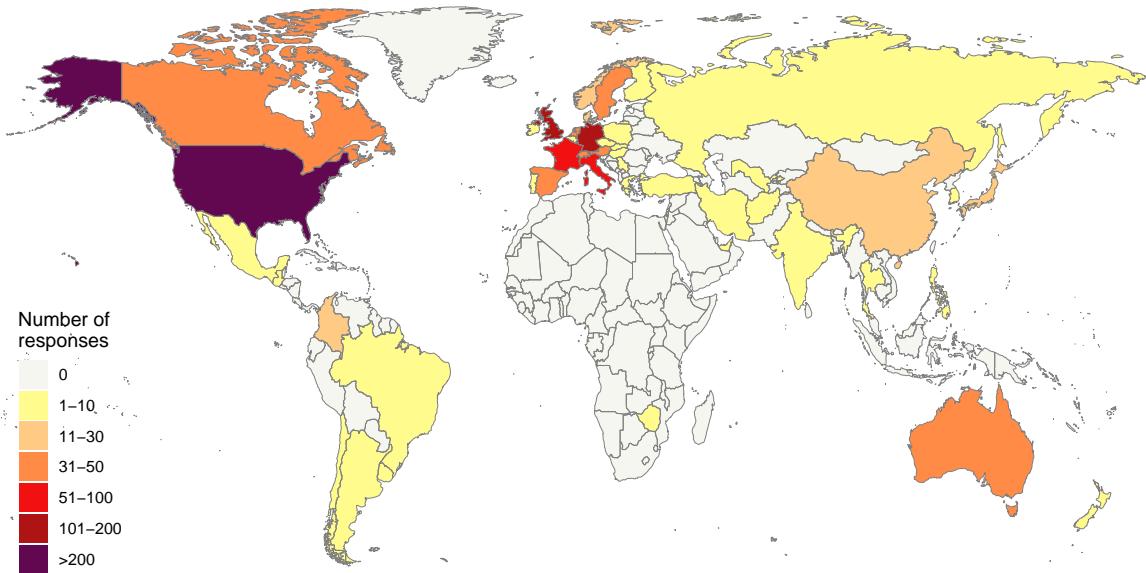
Notes: Variable descriptions can be found in the variable codebook above. [Return to section 3](#).

Figure B.1: Age of respondents



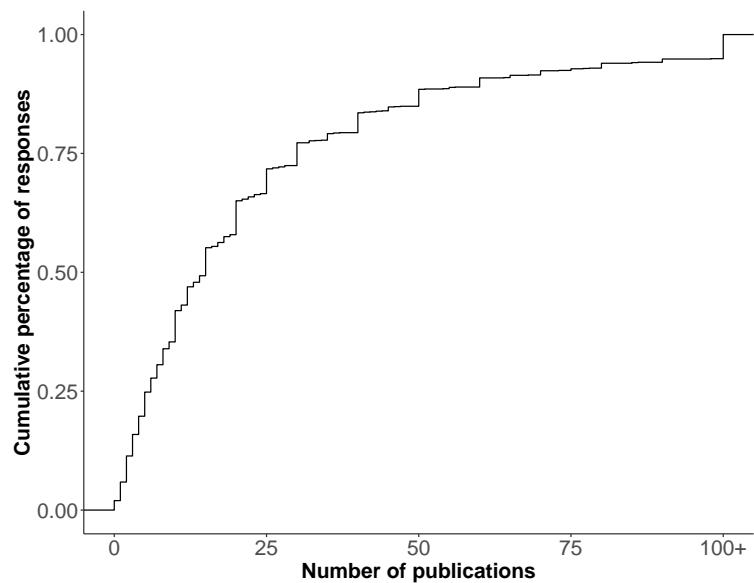
Notes: This figure is based on the responses to question Q42. N = 1,459. [Return to subsection 2.2.2](#).

Figure B.2: Location of respondents



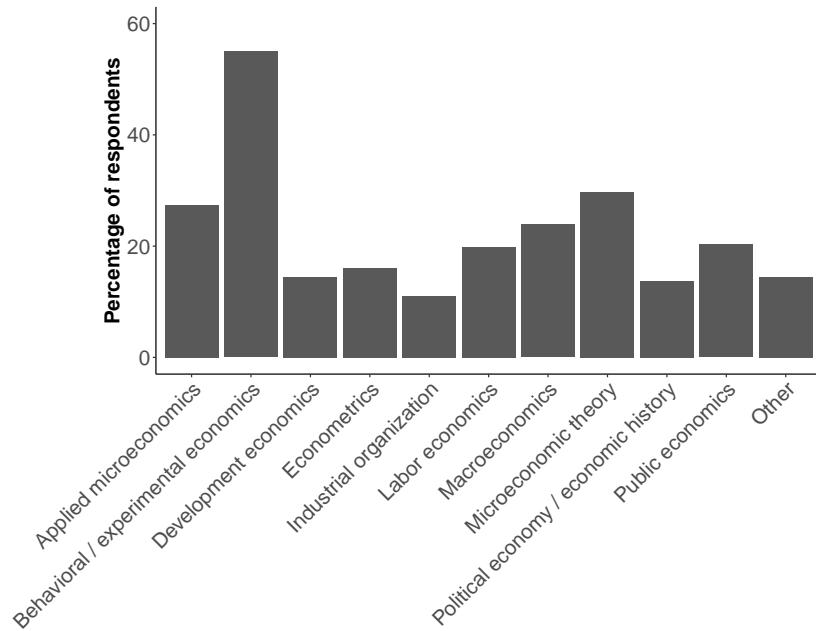
Notes: This figure is based on the responses to question Q44. Respondents who answered “Prefer not to say” are not included in this figure. N = 1,392. [Return to subsection 2.2.2.](#)

Figure B.3: Total number of papers published by respondents



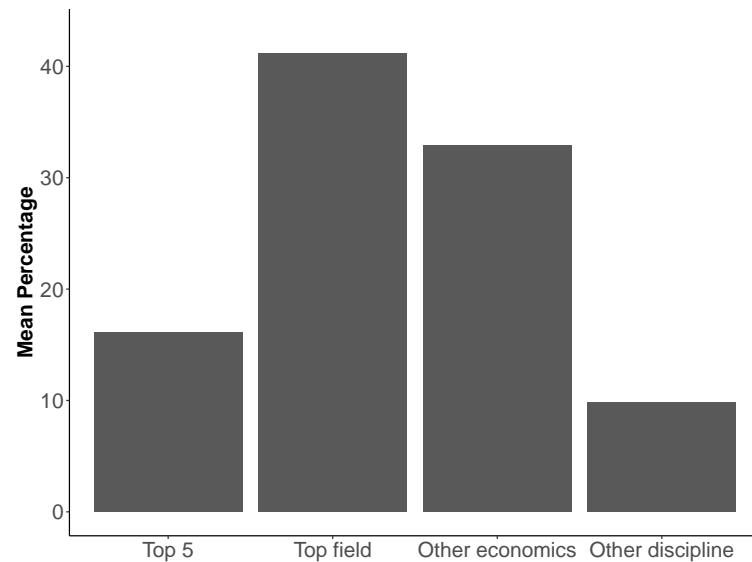
Notes: This figure is based on the responses to question Q39. N = 1,459. [Return to subsection 2.2.2.](#)

Figure B.4: Distribution of research fields of respondents



Notes: This figure is based on the responses to question Q40. N = 1,459. [Return to subsection 2.2.2.](#)

Figure B.5: Types of journals that respondents referee for



Notes: This figure is based on the responses to question Q23. N = 1,483. [Return to subsection 2.2.2.](#)

Appendix C Recruitment

To increase participation in the survey, we conducted multiple stages of outreach and recruitment among different communities of academic economists. A timeline and summary of these efforts can be found in Table C.1. The first part (Wave 1) lasted from July 2020 to October 2020. In this phase, we targeted groups of behavioral and experimental economists. An initial pilot study was sent to a select group of these researchers in order to gather feedback on the content and structure of the survey. After that, a link to the survey was posted on the Economic Science Association ([ESA](#)) discussion forum. Next, we created a database of email addresses of behavioral and experimental economists in order to reach a larger group of potential respondents (mail merge 1). Email addresses in the database came from a wide variety of sources, including lists of behavioral and experimental economists on the RePEc database and participation lists from various conferences and seminars i.e., the ESA conference, the Early-Career Behavioral Economics Conference ([ECBE](#)), and the Stanford Institute for Theoretical Economics conference ([SITE](#)). A few emails were also added individually. The entire email database was sent an initial email in early August, which was followed by a reminder email in late September, sent only to those who had not provided their email address after taking the survey (i.e., in order to be considered for the prize drawing). We sent emails out to 1,802 researchers, for which we received 655 responses (36.3%). A separate but similar email was also sent to the participants of the Virtual East Asia Experimental and Behavioral Seminar ([VEAEBES](#)).

The second part (Wave 2) was conducted from November 2020 to January 2021. In this stage, we shifted our efforts to recruiting economists from outside of behavioral and experimental economics. We did this by reaching out to communities of economists that are not specific to any subfield, as well as conducting efforts targeted at some particular subfields. Posts on the European Economic Association ([EEA](#)) website and Twitter feed advertised the survey to general groups of economists. Emails sent to the [CESifo](#) and [CEPR](#) networks targeted similarly varied groups. Subfield-specific outreach efforts included a post on the Decision Theory ([DT](#)) forum and emails sent to Health Economics at Lancaster ([HEAL](#)) seminar series members. Finally, we constructed another database of email addresses targeted at non-behavioral and experimental economists, with a particular emphasis on reaching out to underrepresented fields like macroeconomics (mail merge 2). The database was partly constructed with participant lists of conferences hosted by various organizations, including the Society for Economic Dynamics ([SED](#)), the American Economic Association/Allied Social Science Associations ([AEA/ASSA](#)), the Society for Judgment and Decision Making ([SJDM](#)), and the Society for the Advancement of Economic Theory ([SAET](#)). We also included contact details collected from the [NBER](#) database and some emails added individually. The entire email database received a survey request in mid December, followed by a reminder email in mid January (once again to those we could not confirm took the survey). We sent emails out to 3,618 researchers, for which we received 269 responses (7.4%). Informal recruitment efforts took place throughout the entire recruitment period, including via individual emails and social media posts sharing the survey link.

As mentioned elsewhere, 1,497 (1,459) individuals at least partially responded (fully responded) to our recruitment efforts. The median response time for fully completed surveys was 15.5 minutes (see Figure C.1 for the full distribution). Most respondents were recruited from the two mail merges, from which we received 875 completed responses (60.0% of our sample). Four other recruitment channels garnered > 100 completed responses each (see Figure C.2).⁵⁷ Characteristics of the sample broken down by recruitment channel can be found in Table C.2. Some clear demographic differences are worth noting. First, the CESifo and CEPR channels have particularly high percentages of respondents over 50 (44.3% and 39.6%, respectively), respondents who are full professors (57.4% and 58.6%), and respondents with editorial experience (49.7% and 56.3%). Additionally, the percentage of US/Canada-based respondents from mail merge 2 (62.9%) is much higher than the other channels, which tend to have relatively more Europe-based researchers. Finally, while all channels have a fairly high percentage of researchers who have refereed for top 5 journals, this percentage is particularly high in the CEPR and mail merge 2 channels (84.0% and 71.4%, respectively).

⁵⁷A recruitment channel is based on the survey link used. Some survey links were used in multiple methods of recruitment; these are considered to be one recruitment channel.

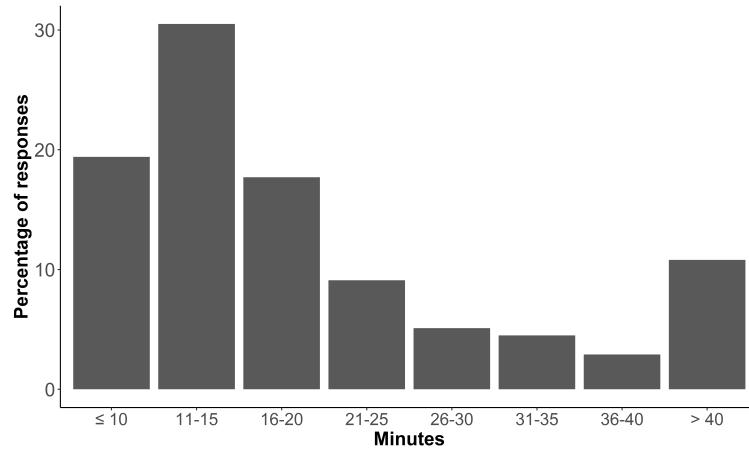
Table C.1: Recruitment strategy summary

Method	Date	Population characteristics	Recruitment channel
Wave 1			
Pilot (individual emails)	Early July 2020	Behavioral economists (initial feedback)	Pilot
ESA discussion forum	16 July 2020	Behavioral and experimental economists	ESA forum
First mail merge	8 August 2020 (main) 25 September 2020 (reminder)	Behavioral and experimental economists (Combination of the RePEc database and conference programs of ESA, AEA/ASSA, ECBE, and SITE + a few additional)	Mail merge 1
Emails to VEAEBES seminar series members	18 September 2020	Behavioral and experimental economists	Mail merge 1
Wave 2			
EEA (post on website and Twitter)	13 November 2020	Various fields	EEA website + social media
DT forum	15 November 2020	Theorists and experimentalists	DT forum
Emails to CESifo members	24 November 2020	Various fields	CESifo network
Emails to CEPR members	26 November 2020	Various fields	CEPR network
Emails to HEAL seminar series members	7 December 2020	Health economists	EEA website + social media
Second mail merge	16 December 2020 (main) 12 January 2021 (reminder)	Economists from various fields (Combination of NBER database and conference programs of SAET, SED, AEA/ASSA, and SJDM + a few additional emails)	Mail merge 2
Personal emails	Sporadically	Various fields	EEA website + social media ^a
Social media posts	Sporadically	Various fields	EEA website + social media

Notes: See previous page for more information about the meaning of the various acronyms.

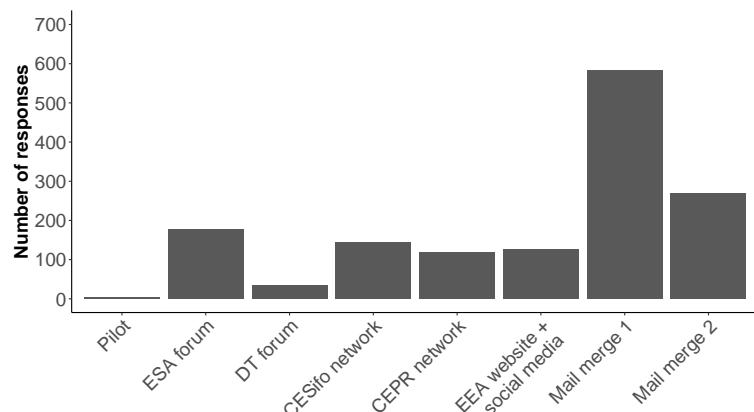
^a Some personal emails were sent using different survey links, but the majority of them used the link for “EEA website + social media.”

Figure C.1: Survey response time



Notes: Time variable converted from seconds to minutes to provide greater clarity. This figure only includes respondents who fully completed the survey. N = 1,459. [Return to subsection 2.1.](#)

Figure C.2: Distribution of responses across recruitment channels



Notes: This figure only includes respondents who fully completed the survey. N = 1,459. [Return to subsection 2.2.1.](#)

Table C.2: Characteristics across recruitment channels^a

	<u>ESA forum</u>	<u>CESifo</u>	<u>CEPR</u>	<u>EEA / social media</u>	<u>Mail merge 1</u>	<u>Mail merge 2</u>
Demographics						
Female	27.5%	17.9%	28.4%	28.5%	23.8%	25.9%
<i>Age:</i>						
Under 40	55.1%	30.0%	26.7%	50.8%	40.1%	48.6%
40-49	27.5%	25.7%	33.6%	35.5%	38.1%	26.1%
50-59	11.4%	26.4%	22.4%	9.7%	15.1%	15.7%
60 and over	6.0%	17.9%	17.2%	4.0%	6.6%	9.6%
<i>Location:</i>						
US/Canada	32.4%	27.1%	23.3%	28.7%	30.5%	62.9%
Europe	57.8%	69.3%	74.1%	61.5%	53.8%	30.9%
Asia/Oceania	6.9%	2.9%	2.6%	8.2%	13.5%	4.6%
Other countries	2.9%	0.7%	0.0%	1.6%	2.2%	1.5%
Position						
Full Professor	22.6%	57.4%	58.6%	25.2%	39.0%	36.5%
Associate Professor	25.6%	13.5%	21.6%	22.8%	25.7%	16.9%
Assistant Professor	26.8%	16.3%	12.1%	29.3%	26.4%	34.2%
Postdoc/PhD Candidate	17.9%	5.0%	0.0%	15.4%	4.9%	6.9%
Other Position	7.1%	7.8%	7.8%	7.3%	4.0%	5.4%
Professional Experience						
Editorial Experience	29.3%	49.7%	56.3%	29.7%	37.8%	40.9%
Average number of publications ^b	20.7	40.3	26.6	25.4	28.0	20.3
Referees for top 5 journals	46.2%	54.5%	84.0%	43.0%	57.5%	71.4%
N^c	190	145	119	128	606	269

^a This table only looks at the recruitment channels (i.e., survey links) that received > 100 partially- or fully-completed responses (= 96.5% of the total sample).

^b Full range used (no top coding).

^c These sample sizes refer to all partially- or fully-completed surveys for each recruitment channel. For the individual statistics, sample sizes may differ from the stated N as unanswered questions and “Prefer not to say” responses were removed from these calculations.

Appendix D Survey questions

[[Back to Section 2](#)]

Consent Form

Principal Investigators: Gary Charness (UCSB), Anna Dreber (Stockholm School of Economics), and Séverine Toussaert (Oxford)

Description: This is a survey on peer review, which should take about 15-20 minutes of your time. We are interested in your view of the current peer review process and how it can be improved.

Eligibility Criteria: You are eligible to participate in this survey if, **over the last two years**, (i) you **completed at least one peer review**; and (ii) you **received referee reports** on a paper you submitted for publication in a peer-reviewed journal.

Risks and benefits: There are no physical or emotional risks associated with this study that would go beyond the risks of daily life. Your participation in this study may improve the peer review process and, therefore, benefit the scientific community. In addition, we will give \$100 (cash or gift certificate) to two people drawn randomly from the respondents; you will be asked to leave your email address in a separate survey link if you wish to be entered in the lottery.

Confidentiality: The information collected in this survey may be published in a report or a journal article and presented to interested parties, including possibly, but not exclusively, members of editorial boards or scientific committees. In no circumstances will your identity or personal involvement in this study be disclosed. No personal data (e.g., your IP address) will be collected, except for your email address if you wish to be emailed the report and/or participate in the prize draw (this information will not be connected to your survey responses and will be destroyed after the prize draw). Other information (e.g., survey responses, time of the survey) will be kept by the researchers and may be used for future studies.

Your rights as a participant: Participation is entirely voluntary. You may leave the survey at any time without any penalty or prejudice.

Ethics approval: This research has been reviewed according to the ethics procedures for research involving human subjects of the University of Oxford (approval # ECONCIA-21-21-20). If you wish to raise any concerns about this study to the ethics committee, please email ethics@economics.ox.ac.uk.

Please indicate below that you have read the above, that you meet the eligibility criteria, and that you are willing to participate in this online survey.

Yes, proceed to the survey YES/NO

Your experience of the peer review process as an author

[Q1]: Over the last two years, how many times did you submit a paper to an economics journal? Please include only first-time submissions (not revisions), with submissions of the same paper to different journals counted separately. [Dropdown with numbers] (SEE FIGURES 2, 8(A&B), E.3, E.11(B))

[Q2]: How would you rate the overall quality of the referee reports you received over this period? Please indicate what approximate percentage of reports were of the following quality (total should sum to 100):

Very low	[]
Fairly low	[]
Average	[]
Fairly high	[]
Very high	[]
Total	[100]

(SEE FIGURES 16, E.12)

[Q3]: What were the characteristics of the low-quality reports? Please tick all that apply:

- Inaccurate statements about what the paper does or does not do
- Overly short report
- Very vague and unconstructive comments
- Written with an aggressive tone
- Personal insults
- Unrealistic demands
- Inconsistent demands
- Other - please specify: [TEXT BOX]

(SEE FIGURES 17, E.13)

[Q4]: A referee report can achieve multiple objectives. How important do you consider each of the following objectives? Please rank 1-4 in order of importance (with 1 being most important) by dragging and dropping the various items: [1 = most important, 2, 3; 4 = least important]

- Help editor reach an informed decision on the paper
- Give general comments that improve the paper
- Provide detailed feedback on the paper
- Make precise suggestions that improve the paper

(SEE FIGURES 3(B), 6(A), E.4(B), E.18)

[Q5]: As an author, what do you expect from the peer-review process? Please rank 1-3 in order of importance (with 1 being most important): [1 = most important, 2; 3 = least important]

- Getting useful feedback on my work
- A timely decision (whether good or bad)
- Getting a reasonable and well-substantiated decision

(SEE FIGURES 3(A), E.4(A))

Improving the quality of peer reviews

[Q6]: Below is a list of proposals to improve peer reviews. On a scale from 1 to 5, how useful do you find each of them? [1 = not useful at all; 2, 3, 4; 5 = extremely useful]

- i. Providing a set of guidelines for writing referee reports.
- ii. Providing doctoral training on how to write peer reviews.
- iii. Making the history of (anonymous) reviews and authors' responses publicly available.
- iv. Removing the anonymity of senior referees.
- v. Removing the anonymity of associate editors.

- vi. Somehow grading reports and rewarding referees for high-quality reports.
- vii. Encouraging the use of a platform that tracks referee activity in a centralized way.
- viii. Making all reports available to all of the reviewers and making sure reviewers know this is being done.
(SEE FIGURES 13(A), 19(A), 20, 21, 22(A), 33(A), E.19(A&B), E.20)

Guidelines for writing a report

[Q7]: What type of comments do you find most useful or would you like to see more of? Please make 3 selections from the following list:

- Comments about the presentation of the results
- Suggestions to improve the existing analysis
- Suggestions about possible extensions
- Comments that help me clarify the contribution of the paper relative to the literature
- Comments about shortening/restructuring the paper
- Comments that put in perspective the assumptions made in the paper
- Comments about missing previous work and references
- Robustness checks

(SEE FIGURE 18)

[Q8]: Do you think journals or associations should provide a template for referee reports? [YES/NO] (SEE FIGURE 19(B))

Information disclosure

[Q9]: In other disciplines, such as public health/medicine, many journals have an open peer review process: referees sign their reports and the entire review history (including responses to referees) is disclosed. On a scale from 1 to 5, how favorable would you be to an open review policy? [1 = not favorable at all; 2, 3, 4; 5 = very favorable] (SEE FIGURE 32(A))

[Q10]: What if this only applied to senior reviewers? [1 = not favorable at all; 2, 3, 4; 5 = very favorable] (SEE FIGURE 32(B))

[Q11]: Another recent trend is to make the history of reports/responses to referees publicly available in an anonymized way unless the reviewers choose to disclose their identity; see e.g., [Nature Communications](#). On a scale from 1 to 5, how favorable would you be to such a policy? [1 = not favorable at all; 2, 3, 4; 5 = very favorable] (SEE FIGURE 33(A&B))

Tracking referee activity

[Q12]: At the moment, there is no centralized system that would allow journal editors to:

- check how many peer review requests a researcher has recently received across all journals.
- find suitable referees who might be currently available to provide a peer review.

One platform called [Publons](#) allows researchers to document their (verified) peer review activity and to register their interest in doing peer reviews for journals. However, it is not widely used at the moment in economics.

On a scale from 1 to 5, how favorable would you be to the more widespread use of Publons or a similar type of platform? [1 = not favorable at all; 2, 3, 4; 5 = very favorable] (SEE FIGURE 13(A&B))

Recognition

[Q13]: Do you think that referees would do a better job if they were better rewarded for their work? [YES/NO] (SEE FIGURE 29)

[Q14]: How should referees be rewarded? Please tick all that apply:

- Excellence in refereeing awards based on specific criteria
- Payment for timely completion e.g., as at the *American Economic Review*
- Discount on submissions to the publisher
- Other - please specify: [TEXT BOX]

(SEE FIGURE 29)

Improving the peer review process more generally

[Q15]: What do you think is an appropriate time length to give to reviewers to submit their reports (in weeks)? [Dropdown: From 1 to 16+ weeks] (SEE FIGURES 25(A&B), E.11(A), E.16)

[Q16]: How do you feel about the policy of having desk rejections? [1 = not favorable at all; 2, 3, 4; 5 = very favorable] (SEE FIGURE 27)

[Q17]: The American Economic Association started a new journal in 2017 called *AER: Insights*. This journal follows a model close to the one of medicine, with the endeavor to accept or reject papers without having to go through a lengthy revision process. Like the papers that *AER: Insights* is looking to publish, reports are supposed to be short and to the point. The whole process is supposed to be fast.

How favorable are you to this type of model? [1 = not favorable at all; 2, 3, 4; 5 = very favorable] (SEE FIGURE 30)

[Q18]: In the case of a rejection, the norm is not to challenge the decision made by the Editor or the views of the referees. This norm is not always followed in practice.

How favorable would you be to a policy allowing the authors to submit a (single) response to the referees and the Editor? The referees would be under no obligation to provide additional comments; a “cooling period” could be required before the authors can send their response. There would be no guarantee of the referees taking this rebuttal into account, and the decision would be final after the comment period. [1 = not favorable at all; 2, 3, 4; 5 = very favorable] (SEE FIGURE 23)

[Q19]: At journals such as *Management Science*, the review process is double-blind i.e., the identity of both the authors and the referees is kept anonymous. How favorable are you to double-blind reviewing? [1 = not favorable at all; 2, 3, 4; 5 = very favorable] (SEE FIGURE 15)

[Q20]: In some fields, authors are allowed to suggest that certain reviewers should be disqualified from reviewing their work. How favorable are you to this possibility? [1 = not favorable at all; 2, 3, 4; 5 = very favorable] (SEE FIGURE 14)

[Q21]: Are there other proposals you would like to make to improve the quality of peer reviews or the peer review process more generally? [TEXT BOX]

Your experience of the peer review process as a referee

[Q22]: On average, approximately how many referee reports do you write per year? [Dropdown with numbers] (SEE FIGURES 4, 5(B), 7(A&B), 8(A&B), E.3, E.6(A), E.11(A))

[Q23]: What percentage of the time do you write referee reports for the following types of journals? (total should sum to 100):

top 5 journal	[]
Top field journal	[]
Other journal in Economics	[]
Journals in other disciplines	[]
Total	[100]

(SEE FIGURES 4(B), B.5, E.2)

[Q24]: Have you occupied or are you currently occupying an editorial position? [YES/NO] (SEE FIGURES 12, E.1)

[Q25]: Usually, how much time do you spend on a referee report, including reading the paper and writing the report? [Dropdown: Less than one hour, 1 or 2 hours, Half a working day, 1 day, 2 days, More than 2 days] (SEE FIGURE 5(A&B))

[Q26]: Over the past two years, what percentage of the time were you late submitting a referee report? [0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100] (SEE FIGURES 24(A), 26, E.15(A&B))

[Q27] [If Q26 answer > 0]: On average, what was your delay? [Dropdown: 1 day, More than 1 day & less than 1 week, 1-2 weeks, 3-4 weeks, More than a month] (SEE FIGURES 24(B), 25(B), E.15(A), E.17)

[Q28]: What do you think is a reasonable number of reports to be assigned per year? [Dropdown with numbers] (SEE FIGURES 7(A&B), E.6(A))

[Q29]: Did you reject a request to referee over these past two years? [YES/NO] (SEE FIGURES 9(A&B), E.6(B), E.11(A&B), E.17)

[Q30]: How many times did you reject a request to referee? [Dropdown with numbers] (SEE FIGURES 9(A&B), E.6(B), E.11(A&B), E.17)

[Q31]: What were the main reasons? Please tick all that apply:

- Conflict of interest
- Inability to judge the paper
- Too remote from your research field
- Lack of time
- Low quality paper
- Lower-ranked journal
- Other - please specify: [TEXT BOX]

(SEE FIGURES 10, E.9(A&B))

[Q32]: How many times did you feel tempted to decline a report even if you ended up fulfilling the request? [Dropdown with numbers]

[Q33]: When you were tempted to decline a report, what were the main reasons? Please tick all that apply:

- Conflict of interest
- Inability to judge the paper
- Too remote from your research field
- Lack of time

- Low quality paper
 - Lower-ranked journal
 - Other - please specify: [TEXT BOX]
- (SEE FIGURES 10, E.9(A&B))

[Q34]: How do you feel about people refereeing papers by co-authors or friends?

- This should never happen.
- This should happen as little as possible but cannot be avoided sometimes.
- This is not a problem as long as the editor is aware of the potential conflict of interest.
- This is not a problem and there is no reason to inform the editor.

(SEE FIGURES 11, E.10)

[Q35]: What do you see as the biggest benefits of being a referee? Please rank 1-5 in order of importance (with 1 being most important) by dragging and dropping the various items: [1 = most important; 2, 3, 4; 5 = least important]

- i. I can help to ensure the right papers are published or rejected
- ii. I can get to know the editors and make myself known.
- iii. I can learn from the opinion of the other referees and the editor.
- iv. I can attentively read papers I would never read otherwise.
- v. Being a referee makes me a better writer.

(SEE FIGURES 6(B), E.5)

[Q36]: How important do you consider your role as a referee? [1 = most important; 2, 3, 4; 5 = least important]

[Q37]: How could your experience as a referee be improved? Please rank 1-4 in order of importance [1 = most important; 2, 3; 4 = least important]

- i. There is a global annual limit on how many papers I am requested to review.
- ii. The editors give clear guidance of what they would like to learn from my report.
- iii. The editors systematically share their decision and the other reports.
- iv. The editors assign me only papers that are related to my research.

(SEE FIGURES 22(B), E.14)

[Q38]: Please enter below any additional suggestion(s) to improve your experience as a referee: [TEXT BOX]

A little more about you

[Q39]: How many papers have you published in your career up to now? Please indicate a ballpark estimate. [TEXT BOX] (SEE FIGURE B.3)

[Q40]: What are your key areas of research? Please select all that apply:

- Applied econometrics
- Applied microeconomics
- Behavioral economics

- Decision theory
- Development economics
- Economic history
- Econometric theory
- Experimental economics
- Financial economics
- Game theory
- Industrial organization
- International trade
- Labor economics
- Macroeconomics
- Microeconomic theory
- Political economy
- Public economics
- Structural econometrics
- Urban economics
- Other - indicate: [TEXT BOX] (SEE FIGURE B.4)

[Q41]: What is your gender? [Dropdown: Male, Female, Other, Prefer not to say]

[Q42]: What is your age? [Dropdown: Under 30, 30-39, 40-49, 50-59, 60-69, 70+, Prefer not to say] (SEE FIGURE B.1)

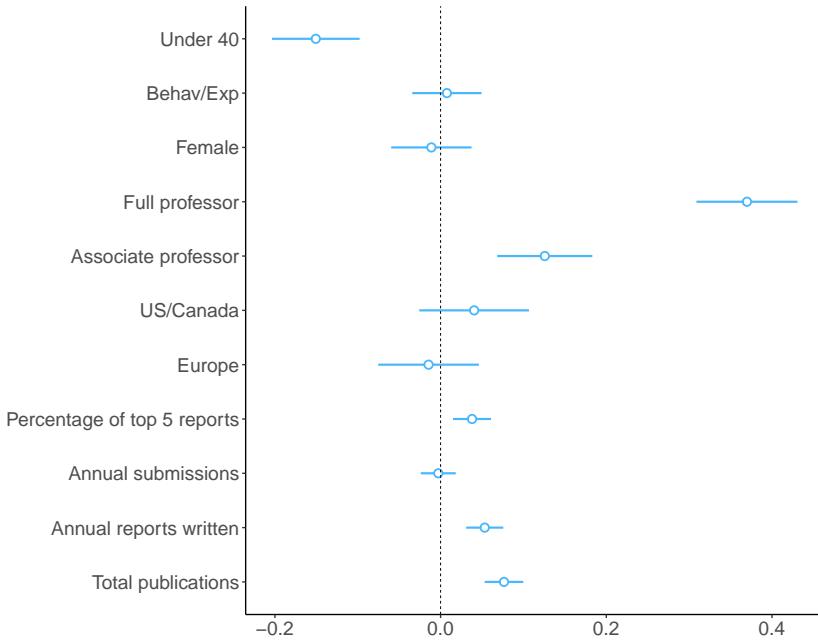
[Q43]: What is your position? [Dropdown: PhD candidate, Post-doctoral researcher, Assistant professor, Associate professor, Full professor, Prefer not to say]

[Q44]: In what country is your job located? [Dropdown] (SEE FIGURE B.2)

[Q45]: Finally, if you have any comments about the survey itself, feel free to add in the text box below: [TEXT BOX]

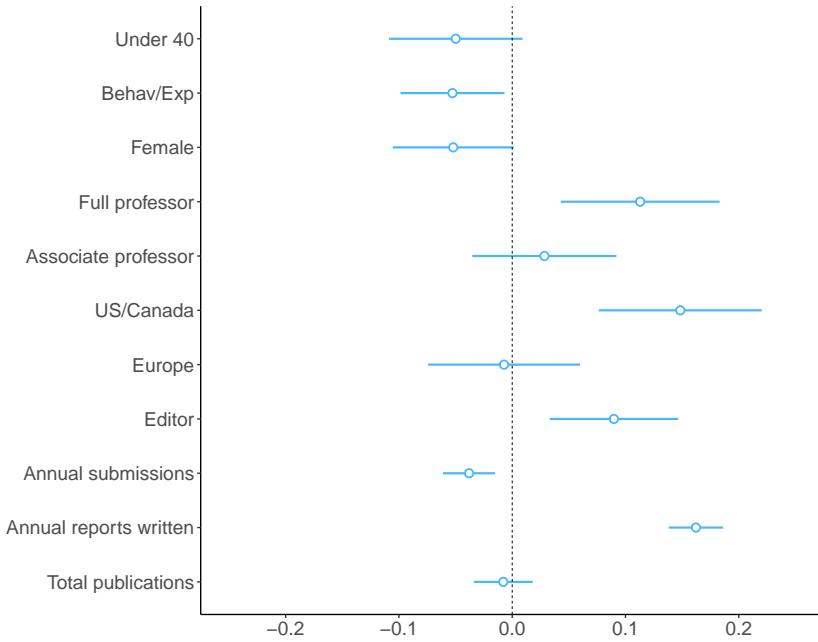
Appendix E Additional figures

Figure E.1: Regression of editorial experience on respondent characteristics



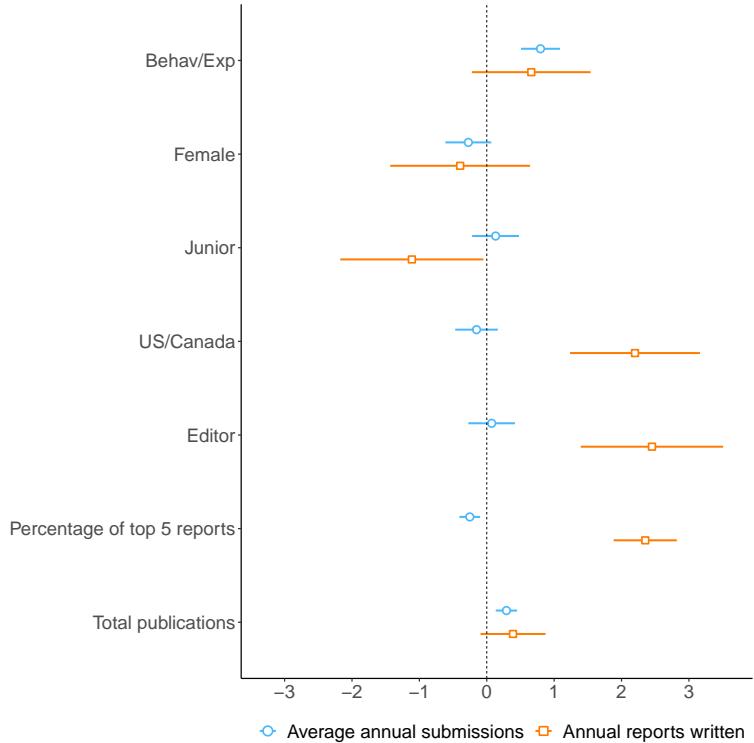
Notes: This figure is based on the responses to question Q24. N = 1,459. **Return to subsection 3.1.2.**

Figure E.2: Regression of top 5 journal refereeing experience on respondent characteristics



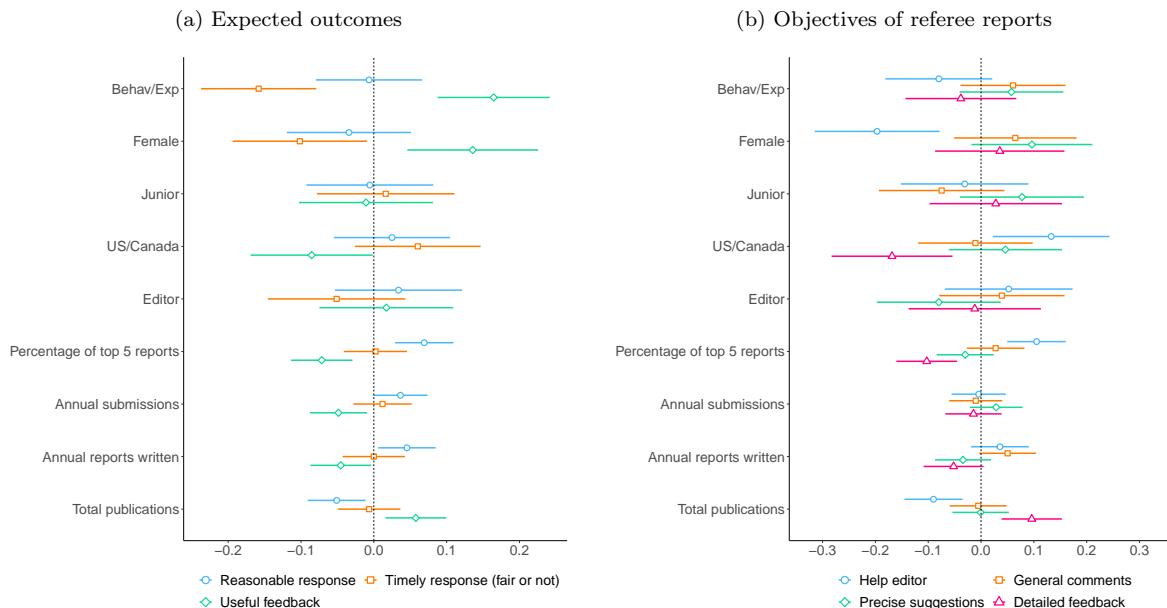
Notes: This figure is based on the responses to question Q23. The outcome variable is the binary indicator Top 5 reviewer; using instead the continuous variable Percentage of top 5 reviews yields very similar results (except for Total publications now being significant at the 5% level). N = 1,459. **Return to subsection 3.1.2.**

Figure E.3: Regression of reports written and submissions made on respondent characteristics



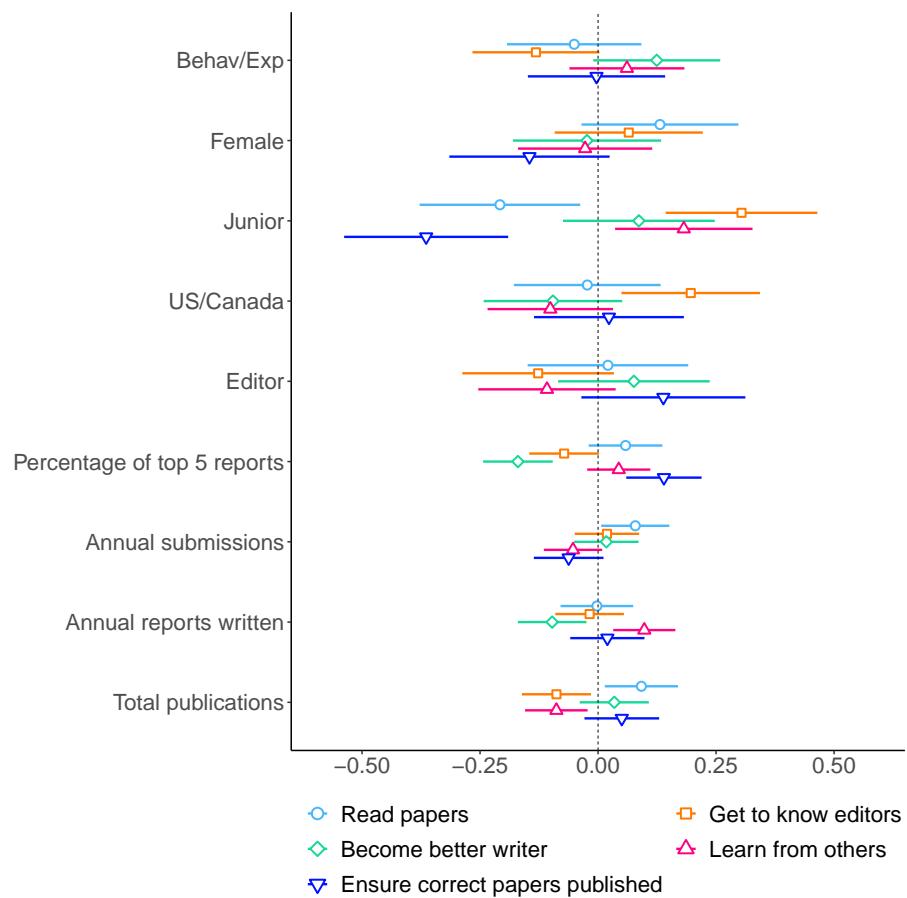
Notes: This figure is based on the responses to question Q1 and Q22. For both submissions made and reports written, the annual amount is regressed. N = 1,459. **Return to subsection 3.1.1 (submissions) or 3.1.2 (reports).**

Figure E.4: Regressions of rank assigned to stated peer review outcomes on respondent characteristics



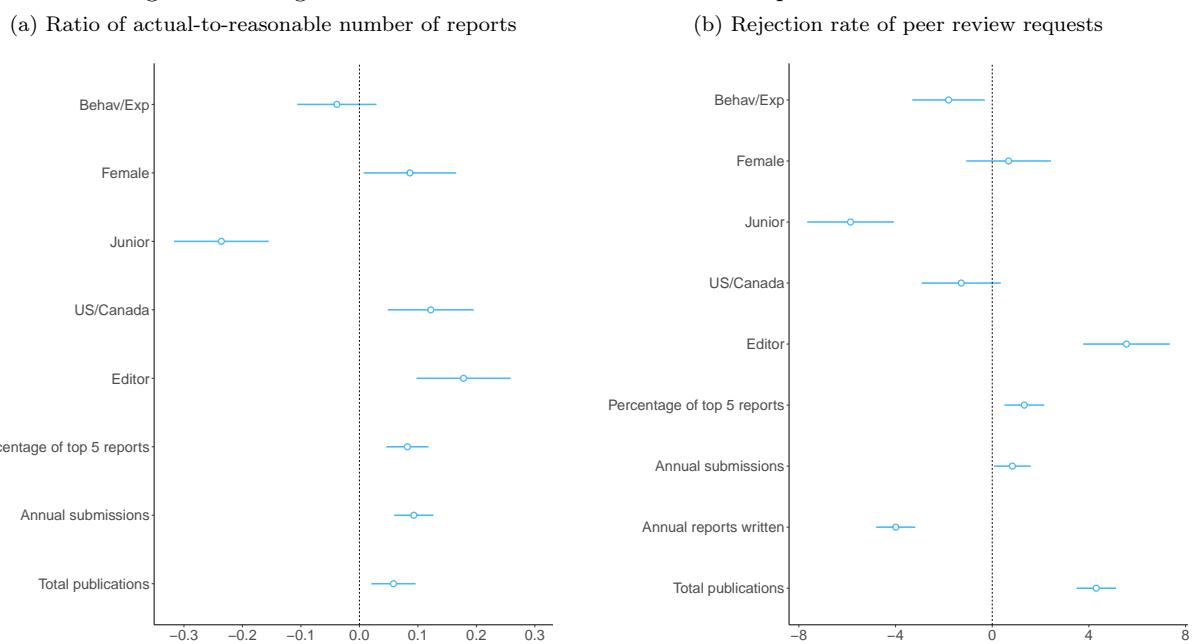
Notes: Panel (a) is based on the responses to question Q5 and Panel (b) is based on the responses to question Q4. For both panels and each outcome, the negative numeric value of the rank assigned by respondents is regressed, so that positive coefficients indicate a better rank on average. N = 1,459 for both panels. **Return to subsection 3.1.1.**

Figure E.5: Regression of rank assigned to stated benefits of refereeing on respondent characteristics



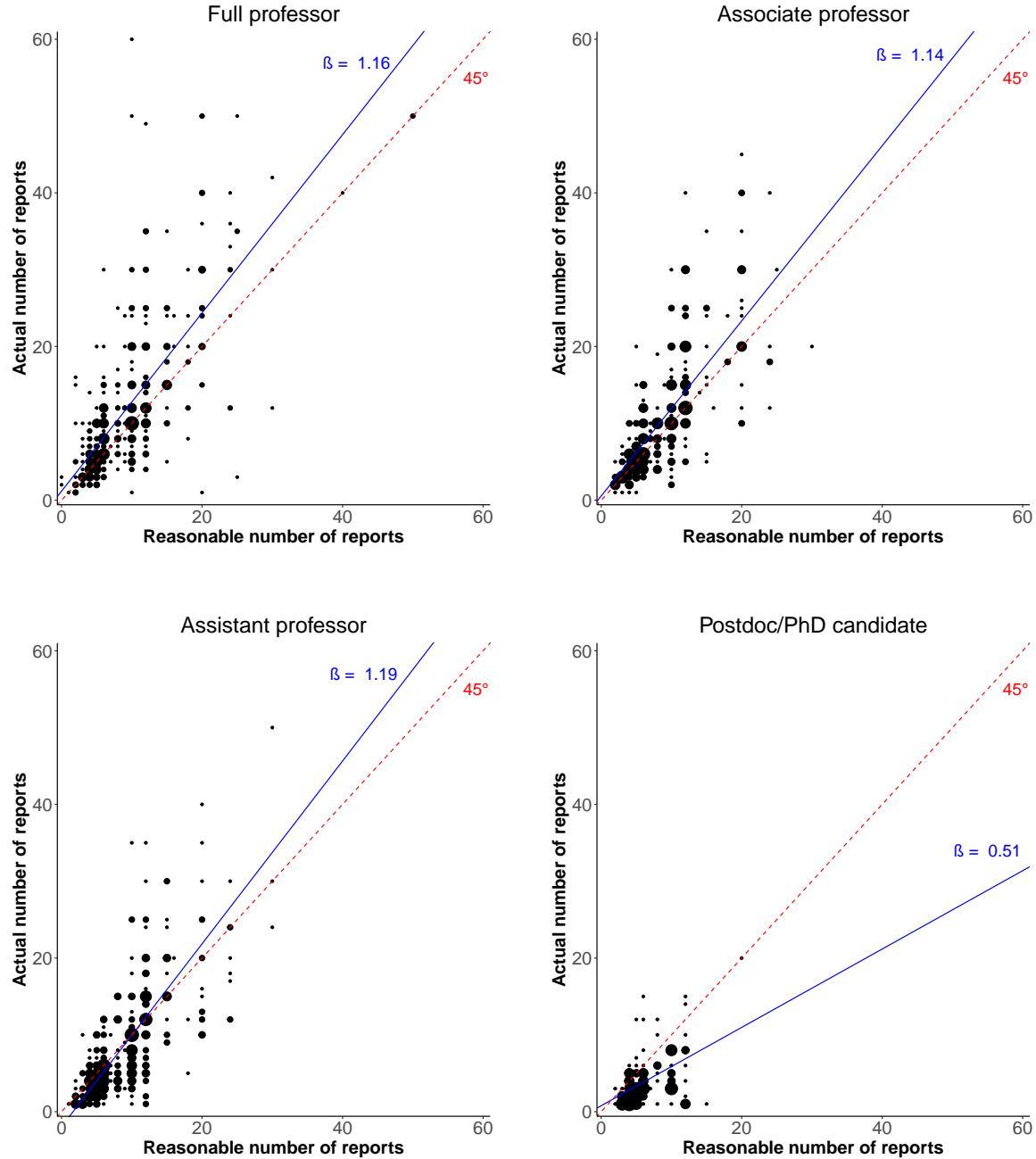
Notes: This figure is based on the responses to question Q35. For each potential benefit, the negative numeric value of the rank assigned by respondents is regressed, so that positive coefficients indicate a better rank on average. N = 1,459. **Return to subsection 3.1.2.**

Figure E.6: Regressions of misallocation variables on respondent characteristics



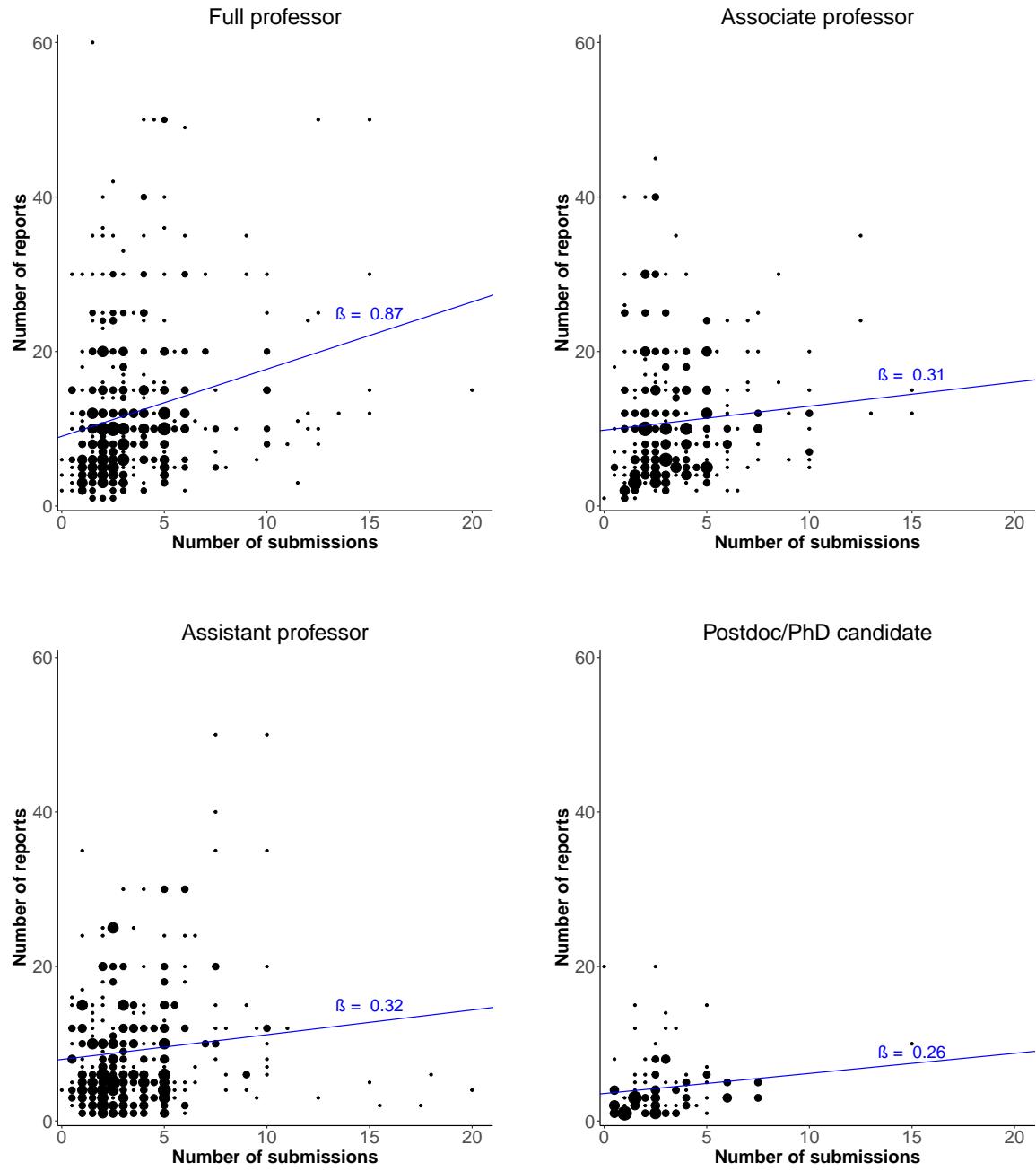
Notes: Panel (a) is based on the responses to questions Q22 and Q28 with the ratio determined by dividing the actual number of reports written by the number of reports considered reasonable; using the difference instead of the ratio gives similar results. Panel (b) is based on the responses to Q29 and Q30. Rejection rate is determined by dividing the number of rejections by the total number of requests (number of requests = number of rejections + number of reports written). N = 1,452 for Panel (a) and N = 1,459 for Panel (b). [Return to subsection 3.2.1.1 for Panel \(a\)](#) and [subsection 3.2.2.4 for Panel \(b\)](#).

Figure E.7: Actual vs. reasonable number of referee reports per year (by position)



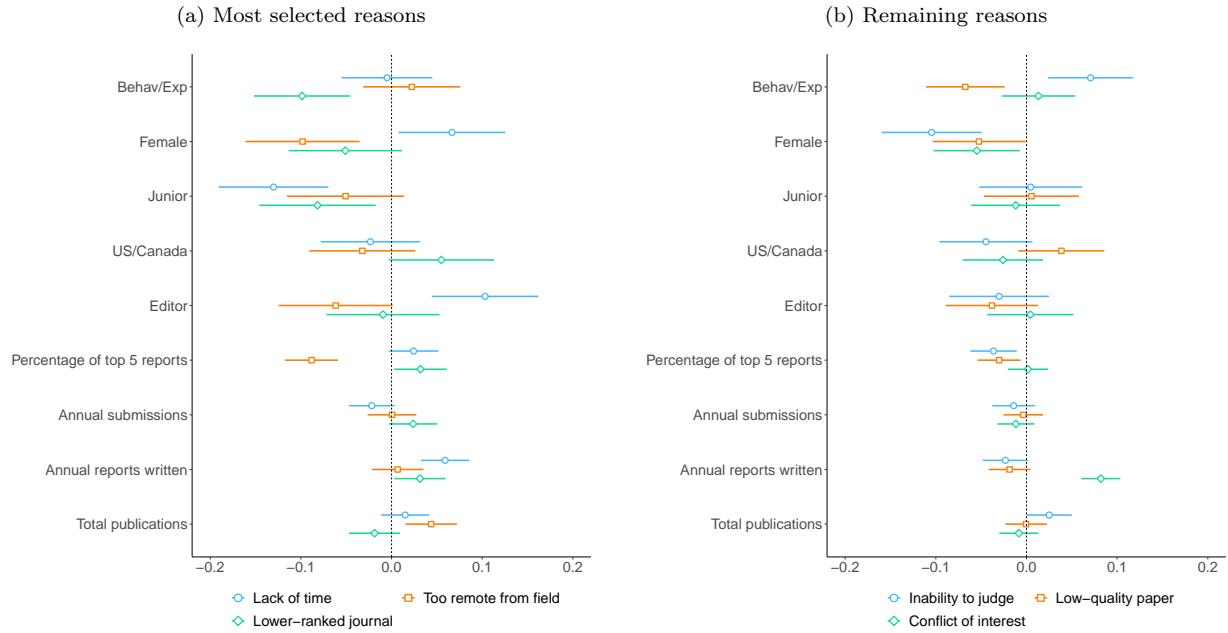
Notes: This figure is based on responses to questions Q22 and Q28 broken down by position (omitting respondents who selected “Other” or “Prefer not to say” as their position). The blue line in each panel is the linear relationship between the actual and reasonable number of reports, with β being the regression slope coefficient. The intercepts for the regression lines are 1.18 for full professors, 0.54 for associate professors, -1.95 for assistant professor, and 0.77 for postdoc/PhD candidates. N = 537 for full professors, N = 308 for associate professors, N = 371 for assistant professors, and N = 107 for postdoc/PhD candidates. [Return to subsection 3.2.1.1.](#)

Figure E.8: Referee reports vs. submissions per year (by position)



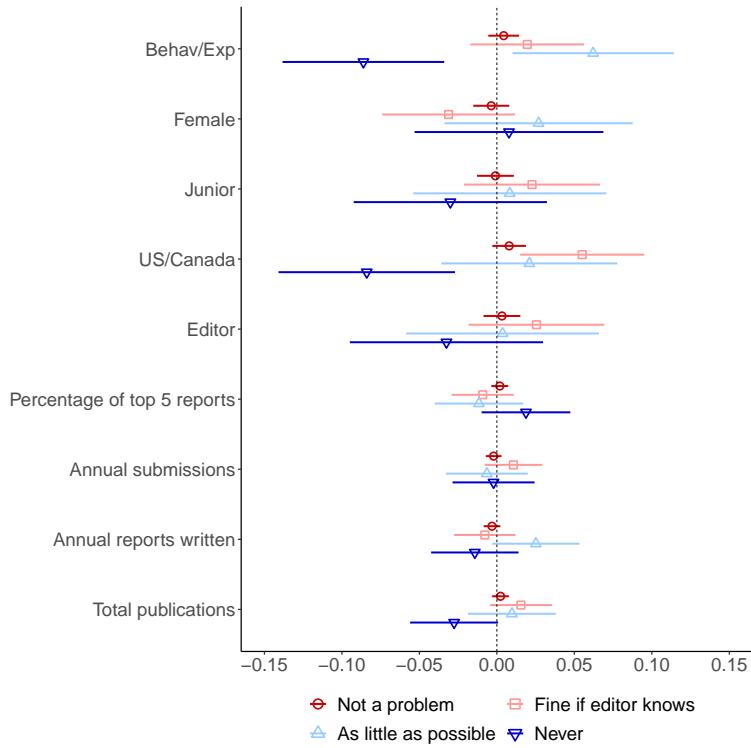
Notes: This figure is based on responses to questions Q22 and Q1 broken down by position (omitting respondents who selected “Other” or “Prefer not to say” as their position). The blue line in each panel is the linear relationship between the number of referee reports written and number of submissions, with β being the regression slope coefficient. The intercepts for the regression lines are 9.02 for full professors, 9.82 for associate professors, 7.97 for assistant professor, and 3.56 for postdoc/PhD candidates. N = 537 for full professors, N = 307 for associate professors, N = 368 for assistant professors, and N = 106 for postdoc/PhD candidates. **Return to subsection 3.2.1.2.**

Figure E.9: Regressions of reasons for (considering) rejecting requests on respondent characteristics



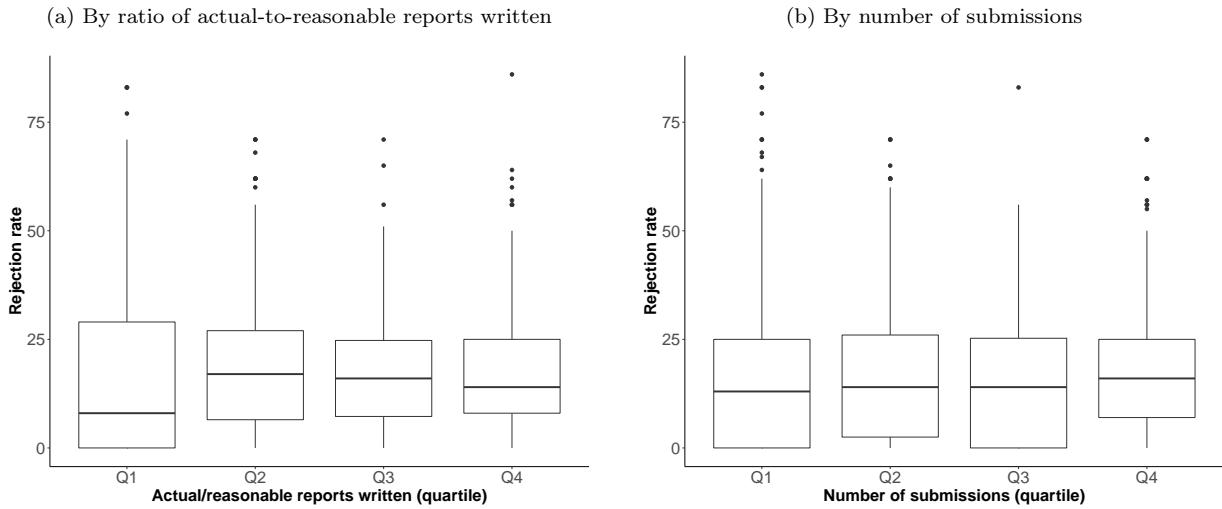
Notes: Both panels are based on the responses to questions Q31 and Q33. Each regression outcome is a binary indicator for whether the stated reason was selected. Since no respondent answered both questions (by design of the survey), the responses were combined into one variable. Panel (a) shows the three characteristics chosen most often and Panel (b) shows the remaining three characteristics. N = 1,359 for both panels. **Return to subsection 3.2.2.2.**

Figure E.10: Regression of attitudes towards friends/co-authors refereeing on respondent characteristics



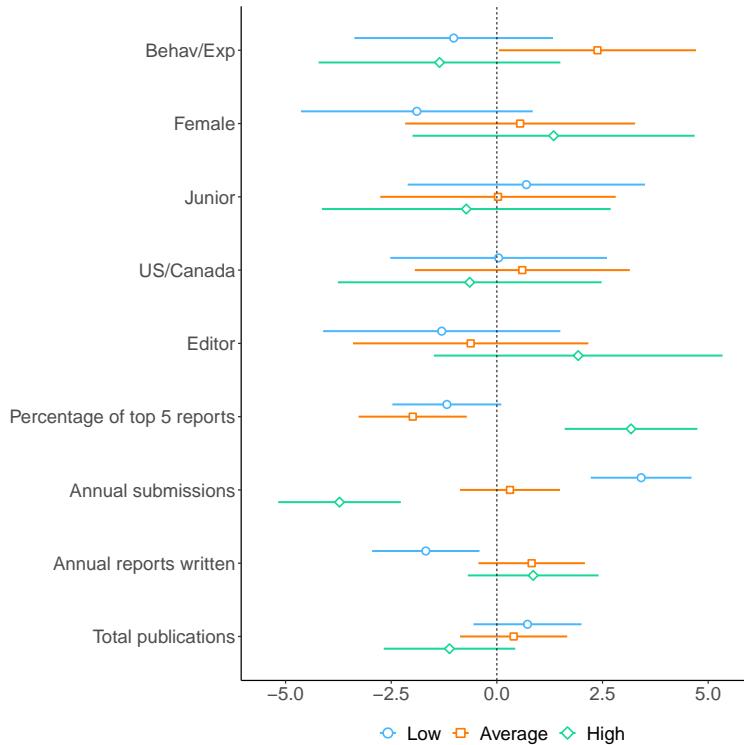
Notes: This figure is based on the responses to question Q34. N = 1,459. Each regression outcome is a binary indicator for whether the stated answer was selected. **Return to subsection 3.2.2.3.**

Figure E.11: Heterogeneity in rejection rates



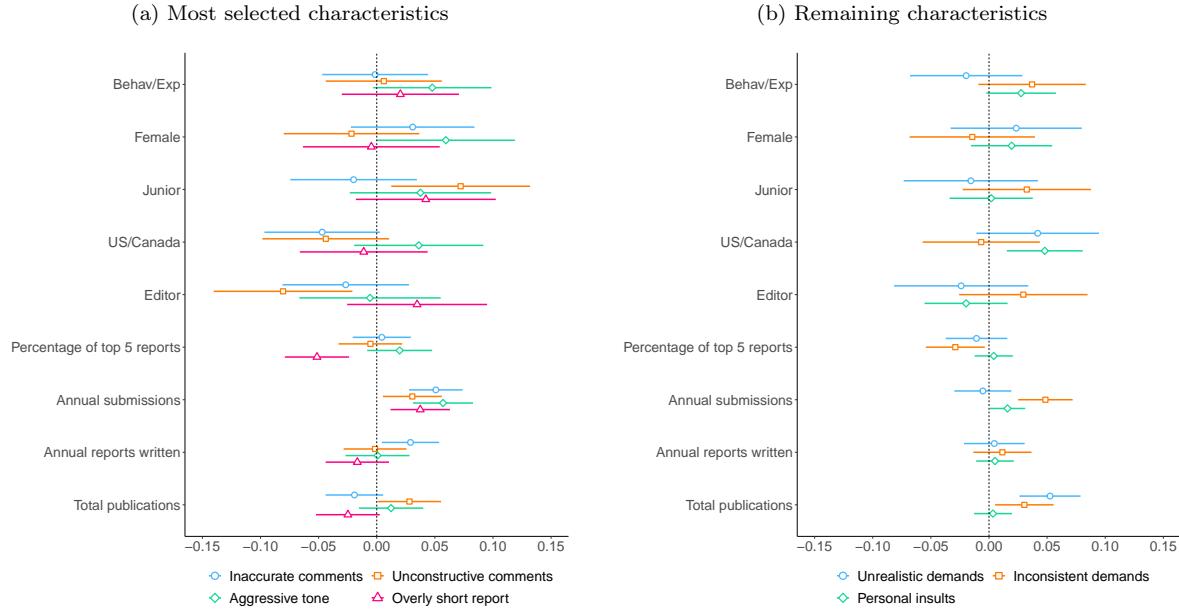
Notes: Both panels combine responses to questions Q29 and Q30 to get the number of requests rejected. The percentage of requests rejected was calculated by taking the proportion of requests rejected from the total of requests received (= requests rejected + reports written) $\times 100$. Panel (a) combines responses from questions Q22 and Q15 to get the ratio of the actual number of reports written to the number of reports considered reasonable and split respondents into quartiles based on this variable. Panel (b) uses the responses to question Q1 to divide respondents into quartiles based on their submission activity. N = 1,476 for both panels. **Return to subsection 3.2.2.4.**

Figure E.12: Regression of perceived quality of reports received on respondent characteristics



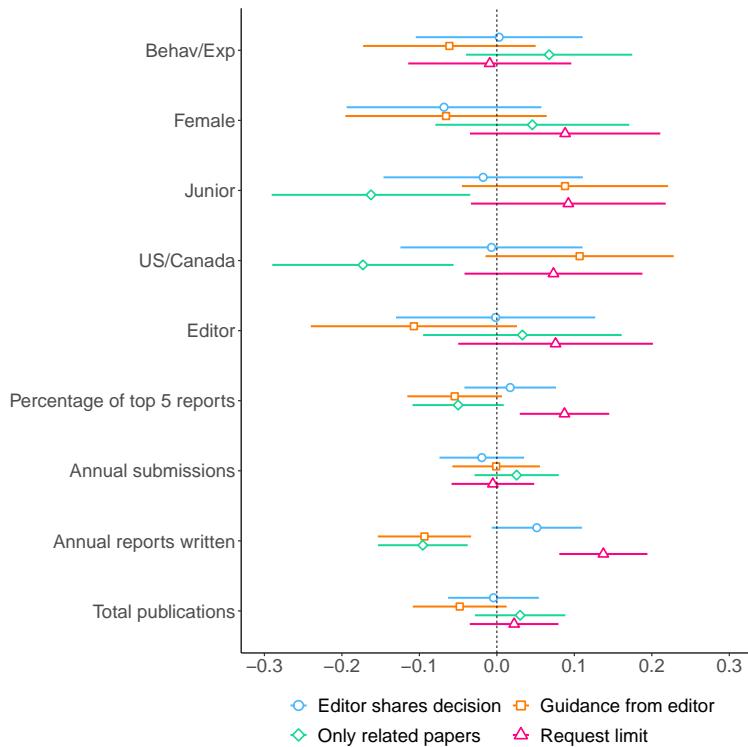
Notes: This figure is based on the responses to question Q2. The three outcome variables correspond to the percentage of referee reports considered to be of low, medium, and high quality (combining the category “very low” with “fairly low,” and the category “very high” with “fairly high”). N = 1,459. **Return to subsection 3.3.1.**

Figure E.13: Regressions of characteristics of low-quality reports on respondent characteristics



Notes: Both panels are based on the responses to question Q3. Each outcome variable is a binary indicator for whether the stated reason was selected. Panel (a) shows the four characteristics chosen most often and Panel (b) shows the remaining three characteristics. N = 1,459 for both panels. **Return to subsection 3.3.1.**

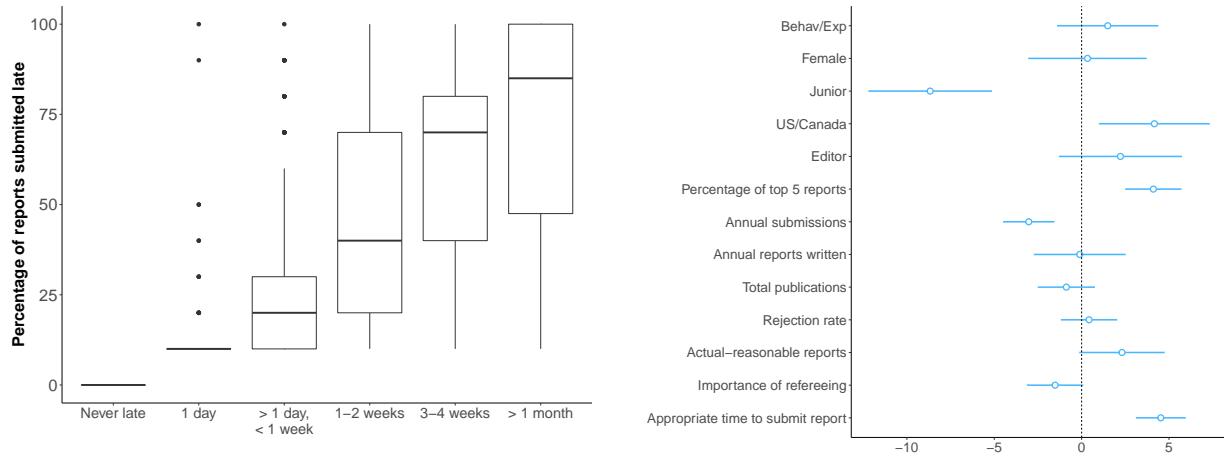
Figure E.14: Regression of rank assigned to proposals for improving refereeing on respondent characteristics



Notes: This figure is based on the responses to question Q37. For each proposal, the negative numeric value of the rank assigned by respondents is regressed, so that positive coefficients indicate a better rank on average. N = 1,459. **Return to subsection 3.3.3.**

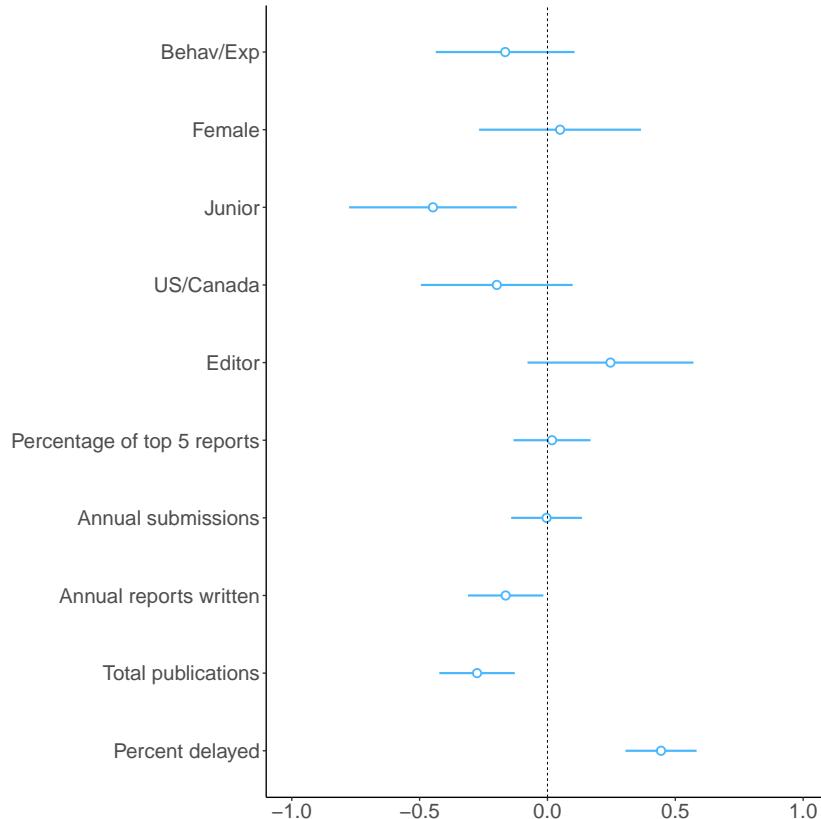
Figure E.15: Heterogeneity in percentage of late reports

(a) Percentage late and average delay (b) Regression of percentage late on respondent characteristics



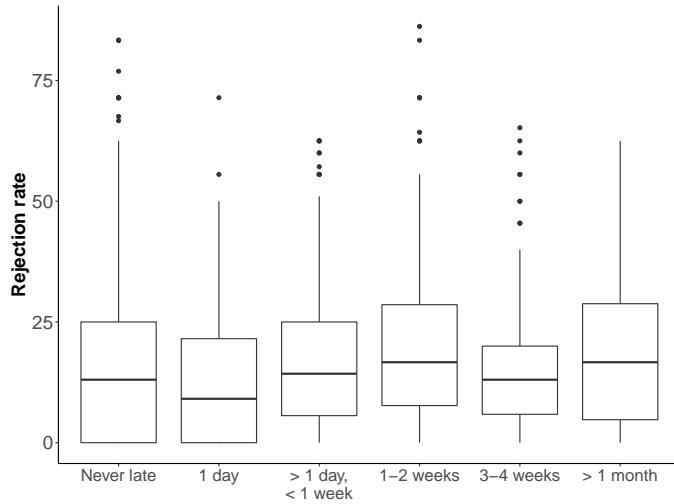
Notes: Both panels are based on the responses to question Q26. Panel (a) also uses responses to the survey question Q27. N = 1,483 for Panel (a) and N = 1,459 for Panel (b). **Return to subsection 3.4.1.3.**

Figure E.16: Regression of reasonable time for reports (in weeks) on respondent characteristics



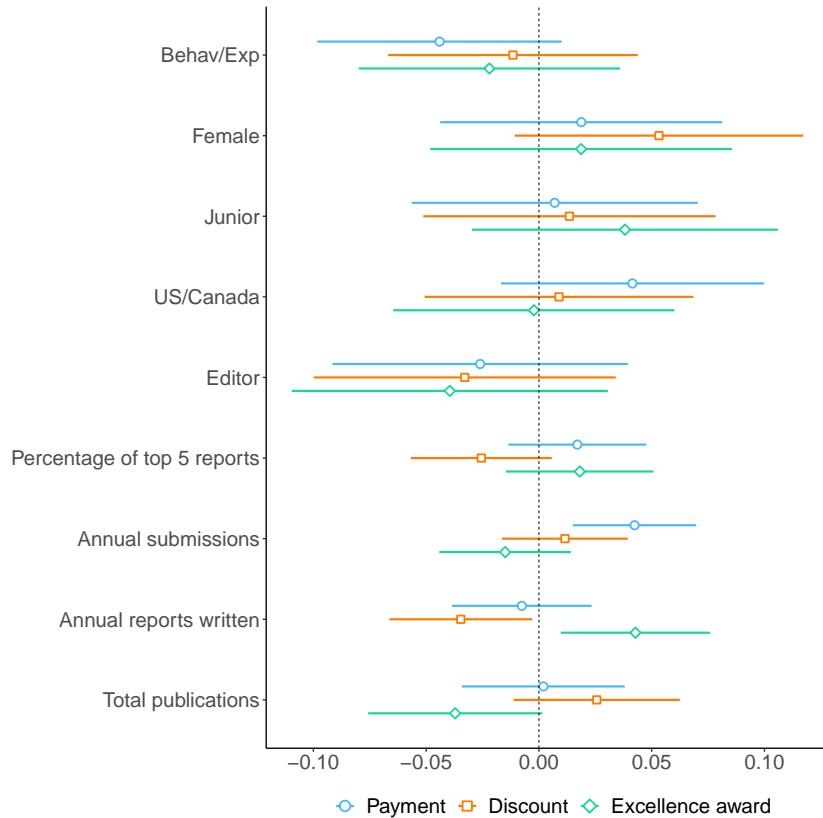
Notes: This figure is based on the responses to question Q15. N = 1,467. **Return to subsection 3.4.1.3.**

Figure E.17: Average delay and rejection rate of review requests



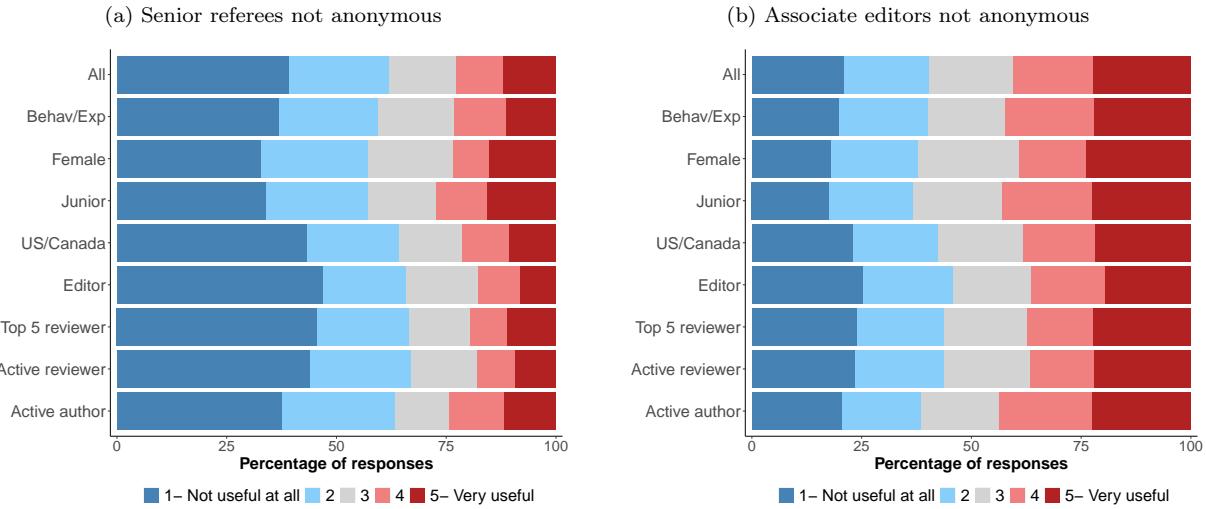
Notes: This figure is based on responses to question Q27 (average delay). It also combines responses to questions Q29 and Q30 to get the number of requests rejected. N = 1,483. **Return to subsection 3.4.1.3.**

Figure E.18: Regression of referee incentives on respondent characteristics



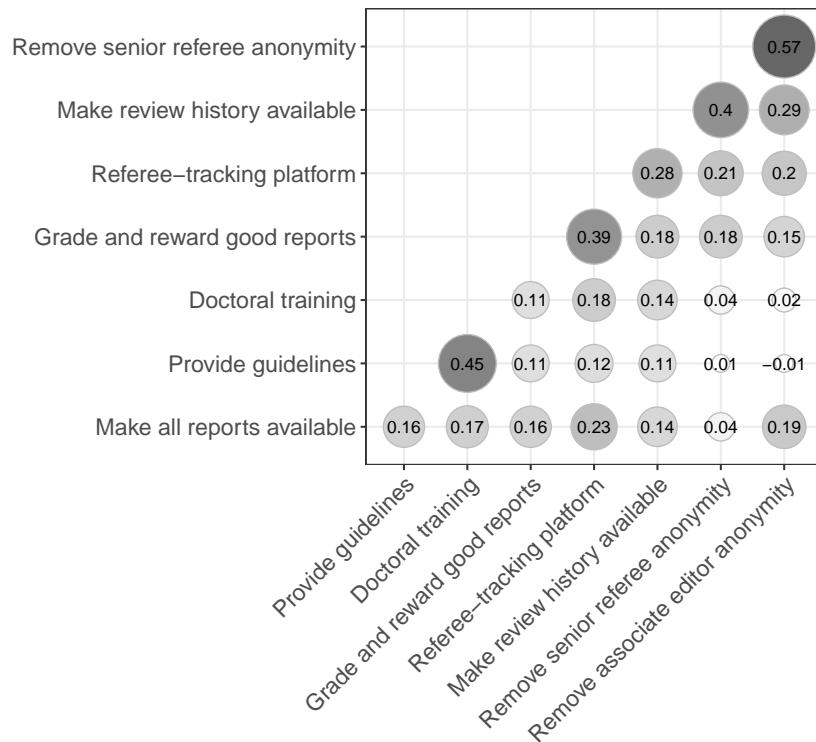
Notes: This figure is based on the responses to question Q4. Each regression outcome is a binary indicator for whether the stated incentive was selected. N = 1,075. **Return to subsection 3.4.2.8 (payments/discounts) or 3.4.2.9 (excellence rewards).**

Figure E.19: Perceived usefulness of removing reviewers' anonymity



Notes: Both panels are based on the responses to question Q6. Panel (a) is based on item *iv*. Panel (b) is based on item *v*. N = 1,459 for both panels. **Return to subsection 3.5.2.4.**

Figure E.20: Correlations between the perceived usefulness (1-5) of various proposals



Notes: This figure is based on the responses to question Q6. The color and size of the circle correspond to the magnitude of the correlation. All correlations above 0.04 are significant at the 5% level. N = 1,469. **Return to subsection 4.1.2.**

Figure E.21: Word cloud of top 100 most common words



Notes: This figure shows the 100 most frequent words used in the survey's comment section. The size of a word's font indicates its relative frequency, where the biggest word ("referee") occurs 144 times and the smallest words occur 7 times (several instances). We allowed for bigrams and lemmatized all words. We removed NLTK's stop words, ciphers, punctuation and special characters. Further, we removed the following common words that carry little meaning for our purposes: ['able', 'actually', 'almost', 'already', 'also', 'always', 'ask', 'b', 'come', 'could', 'eg', 'especially', 'etc', 'even', 'everyone', 'example', 'far', 'feel', 'find', 'first', 'get', 'give', 'go', 'happen', 'know', 'least', 'less', 'let', 'like', 'lot', 'love', 'make', 'many', 'might', 'much', 'n', 'need', 'never', 'nt', 'obviously', 'often', 'one', 'overall', 'part', 'perhaps', 'put', 'rather', 'really', 's', 'say', 'second', 'see', 'seem', 'something', 'sometimes', 'take', 'thing', 'think', 'try', 'two', 'use', 'want', 'way', 'well', 'whether', 'would', 'x']. **Return to subsection 4.1.3.**