Peer Review in Economics: Where Are We? Where Can We Go?*

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

June 2023

Abstract

Peer review is the cornerstone of scientific publishing in economics, but data issues have inhibited past efforts to study it. To address this challenge, we combine insights from a novel survey dataset documenting the submission and review activity of over 1,400 researchers with observational evidence on how peer review in economics compares to the evaluation processes in other disciplines. In this paper, we present the main takeaways of our efforts, with an emphasis on exploring three key trade-offs that govern any attempts at reform: (i) the appropriate balance between relying on expert opinions and seeking input from a diverse pool of referees; (ii) the challenge of reducing turnaround time while maintaining rigor in peer review; and (iii) the difficulty of increasing the transparency of evaluations while preserving confidentiality. We conclude by fleshing out possible steps towards reform on these sets of issues.

*Acknowledgments: The article below represents an adaptation of our earlier report on peer review in economics. We would like to thank all of the associations and networks that helped us advertise the survey to their members, including CEPR, CESifo, the ESA, and the EEA. Our appreciation also goes out to the editors at various journals for the statistics provided and to Peter Andre and Fulya Ersoy for the statistics and datasets they shared with us. This project 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).

1 Introduction

Peer review occupies a central place in the lives of researchers. As authors, they navigate the system in hopes of publishing impactful research and securing their professional standing. In their roles as referees and editors, they spend countless hours evaluating the importance and scientific merit of new research. Given the prominence of peer review in academia, it is striking that our understanding of its inner workings remains limited. Our lack of knowledge partially relates to a scarcity of suitable data (Squazzoni et al., 2020). The expectation of confidentiality (COPE Council, 2017a) often limits the questions that can be explored with individual-level data. Meanwhile, the willingness of journals to provide aggregate-level statistics varies widely, with the most transparent outlets periodically summarizing their activity in public reports and others providing no information at all. As such, guided by data availability, much prior research has focused on a single aspect of peer review in isolation. But since the individual components of peer review are closely interconnected, studying them in isolation may miss out on important trade-offs and the downstream effects of reforms. With this in mind, we sought to study peer review in economics in a more holistic manner - one that acknowledges interdependence across journals and the stages of the peer review process. Our efforts culminated in a comprehensive report on the state of peer review in economics (Charness et al., 2022).

The present study, which is based on the report, contributes to the state of knowledge about peer review in three ways. First, we examine the *supply-side* (reports written) and *demand-side* (manuscripts submitted) of peer review jointly, and across a range of journals. We surveyed over 1,400 researchers in economics to collect individual-level data on their submission and review activities, in their dual roles as authors and referees. Second, we present evidence on how our respondents perceive the current system and how they reason about certain proposals for reform. Crucially, we link their views to their demographic characteristics and participation in the system, in order to explore the heterogeneity of attitudes. Third, we complement our survey findings with a review of the existing literature and observational data on practices across disciplines. With these datasets in hand, we discuss the state of peer review in economics and potential avenues for reform. Our discussion is organized around three key themes, each of which highlights a trade-off in peer review or a challenge for attempts at reform. The first asks to what extent editors should increase the diversity of refereeing pools and distribute responsibilities more evenly across researchers, as opposed to continuing to solicit evaluations from a small group of trusted experts. The next explores the challenges of reducing publication lags while maintaining the rigor of evaluation. The final one concerns how journals can balance demands for transparency with their obligation to protect the confidentiality of stakeholders.

We acknowledge that our study is not the first to investigate the state of peer review. From an interdisciplinary perspective, Tennant et al. (2017) leverage an extensive survey of the literature on peer review to offer an account of its history and trajectory. Other interdisciplinary studies have surveyed researchers about proposals to change peer review (Mulligan et al., 2013) and experiences with unprofessional referee reports (Silbiger and Stubler, 2019). Relative to this work, our study focuses on the topics most relevant to peer review in economics and offers an economic perspective on issues common across disciplines. Within economics, studies on peer review have been relatively scarce and limited in scope. One early example comes from Mason et al. (1992), who surveyed 304 economists about delays in manuscript evaluations. More recently, observational studies have delivered insights on referee and editorial decision making (Card et al., 2020b,a) and the revision process in economics (Hadavand et al., 2022). While these efforts have been highly instructive, data availability issues have usually limited them to evidence from a handful of (elite) journals. By contrast, the present work leverages a large survey sample to (i) jointly study the review and submission activities of researchers across all journals that they interact with and (ii) explore an array of novel questions, such as the allocation of refereeing resources and the trade-off between transparency and confidentiality. Beyond that, our discussion of potential reforms and the normative considerations surrounding them contributes to the ongoing conversation around improving peer review in economics.[1]

We structure our paper as follows. In Section 2, we introduce the survey design and present basic descriptive statistics about our respondents and their experiences with peer review. The next three sections are organized thematically around the "who," "when," and "how" of peer review. Section 3 considers *who* should be responsible for evaluating manuscripts, tackling the issues of referee expertise and diversity. Section 4 addresses the *when* of peer review, by examining proposals to reduce publication lags and modify the timing

[1] Examples of authors who have offered their perspectives on the current system and possible improvements include Frey (2003), Berk et al. (2017), Akerlof (2020), and Siemroth (2023), among others.

of evaluations. Section 5 considers *how* peer review should operate, focusing on the extent to which it should be conducted in the public eye. Next, in Section 6, we contextualize our findings with respect to the peer review systems in other disciplines, in the hope of understanding what reforms are possible within economics. Finally, in Section 7, we conclude with a discussion on the limitations of our study and possible paths forward.

2 Survey on Peer Review in Economics

The present paper draws on evidence from a number of sources, but the survey represents the cornerstone of our efforts. In this section, we first describe the content of our survey and how we recruited our respondents. We next use our unique survey data on respondent submission and review activities to characterize the "market" for peer review. The survey dataset, as well as further regressions and figures of interest, is available on our OSF page (https://osf.io/eczkv/).

2.1 Survey Design and Sample

We conducted an anonymous survey of economists between July 2020 and January 2021 in order to take a snapshot of the state of peer review in our discipline. Researchers were eligible to participate if, over the prior two years, they had (i) written at least one referee report and (ii) received reports on at least one journal submission. Generally, we inquired about their experiences with the system, their opinions about its current performance, and their attitudes towards certain proposals for reform. The survey was divided into three sections: the first two asked about peer review from their perspective as a reviewer and as an author (with the order randomized across respondents), while the third concerned their demographic characteristics. The full questionnaire can be found in Appendix D.

Our recruitment strategy was designed to maximize the number and diversity of responses (see Appendix C). In the end, most of our responses came from (i) outreach to the members of various associations and (ii) mail merges sent to researchers in two recruitment waves (Appendix Figure B.1).[2] Channel (i) took a variety of forms, including posts made on the forums of the Economic Science Association (ESA). For channel (ii), the first recruitment wave targeted behavioral/experimental economists and was sent to 1,802 contacts, yielding 655 responses (36.3%). The second targeted economists from all fields and was sent to 3,618 contacts, which yielded 269 responses (7.4%).[3] In total, 1,459 (1,497) researchers fully (at least partially) completed the survey. We make use of responses from partially-completed surveys and the pilot survey when possible.

We next tried to understand the representativeness of our sample. To the best of our knowledge, general statistics on the world population of economists are nonexistent. In light of this, we compare our respondents' characteristics to those of the weighted sample in Andre and Falk (2021).[4] Compared against this baseline, our sample seems to be broadly representative of the profession, with the following exceptions: behavioral/experimental economists and Europe-based researchers were oversampled, while older researchers, associate professors, and macroeconomists were undersampled. Descriptive statistics for our survey respondents and those of Andre and Falk (2021) can be found in Table A.1.

2.2 The Peer Review Market

Before proceeding with the discussion-based sections, we will present some descriptive statistics on the peer review activity of our respondents. In the "market" for peer review, authors generate demand for referee reports by submitting manuscripts to journals (demand side). Referees meet this demand by evaluating these manuscripts and writing reports (supply side). Unlike in traditional markets, participants are usually active on both the demand and the supply side. Using our survey data, we can examine our respondents' level of involvement on each side of the market and what they expect from their involvement.

[2] Different survey links were generated each time, allowing us to track the number of responses by recruitment channel. A downside of this 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.

[3] These response rates likely represent lower bounds, as some researchers may have already responded via other channels, such as the ESA forum. This is particularly likely to have been the case for the first wave of emails.

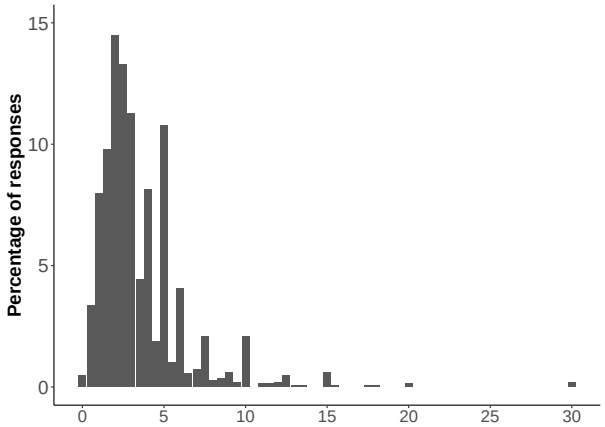
[4] Andre and Falk (2021) elicited survey responses from a database of economists that consisted of almost all active economists with publication data on EconLit. As such, their sample is weighted with respect to an underlying study population.

Starting with the demand side, we see that over a period of two years, our respondents made an average of 5,234 new manuscript submissions per year ($N = 1,484$), for a mean (median) of 3.5 (3) per person (Panel (a) of Figure 1). However, participation is highly uneven, with 50% of manuscripts being submitted by only about 20% of authors (Panel (b) of Figure 1). Although submissions may not perfectly correlate with demand for referee reports, this finding still suggests that demand is high but non-uniform. Given this, it is important to understand whether authors generally agree on the purpose of peer review, and whether their views differ based on their level of involvement. We asked respondents to rank the relative importance of three objectives of the peer review process. Authors indicated that receiving "*a reasonable and well-substantiated decision*" from the editor is more important than other objectives, with 62% prioritizing this option. Receiving "*a timely decision*" and "*useful feedback*" were ranked as secondary objectives, with about 21% and 17% of respondents prioritizing them, respectively. This suggests that our respondents might prefer retaining rigor over reducing delays in peer review, to the extent that these are in tension; we will revisit this trade-off in Section 4. These views do tend to vary with authors' reported participation in peer review. Top 5 referees, active reviewers, and active authors give higher priority on average to getting a reasonable response, and lower priority to useful feedback. Meanwhile, behavioral/experimental economists and female researchers put more weight on getting useful feedback and less weight on a timely decision (Panel (a) of Appendix Figure B.2).

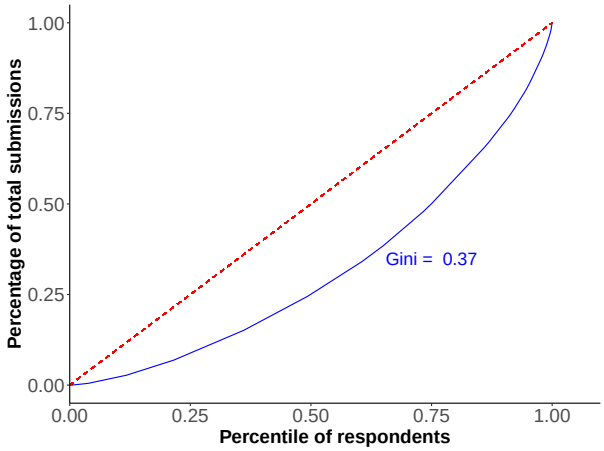
Similarly, we asked respondents to rank the importance of four objectives of referee reports. About 67% chose "*[helping] editors reach an informed decision*" as the most important objective. Comparatively, receiving reports which "*make precise suggestions that improve the paper*" seems to be of secondary importance, but was still preferred over providing general comments or very detailed feedback. This signals a demand for clear direction and guidance during the revision process, an issue that comes up again in Section 4. Researchers based in the US/Canada and top 5 reviewers tended to put more weight on helping the editor decide and less weight on detailed feedback, while the opposite holds for those with more publications (Panel (b) of Appendix Figure B.2).

Figure 1: Submissions to an economics peer-reviewed journal

(a) Annual number of submissions (average over 2 years)



(b) Lorenz curve of submissions



Notes: 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.

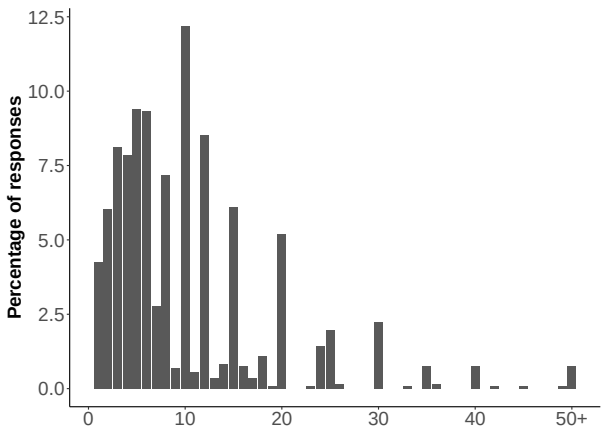
Shifting to the supply side, our survey respondents wrote about 15,060 referee reports per year ($N = 1,484$) over the same time period, for a mean (median) of 10.2 (8) reports per person (Panel (a) of Figure 2). Participation is again skewed, with the top 10% most active referees writing 20 or more annually (Panel (b) of Figure 2). More active referees tend to be researchers who are working in the US/Canada, who have editorial experience, and who occupy more senior positions (Appendix Figure B.3). We also find heterogeneity in

the journals that respondents wrote reports for, with about half never having refereed for a top 5 journal. Moreover, 80% of top 5 reviews were produced by only 25% of respondents (Panel (b) of Figure 2).

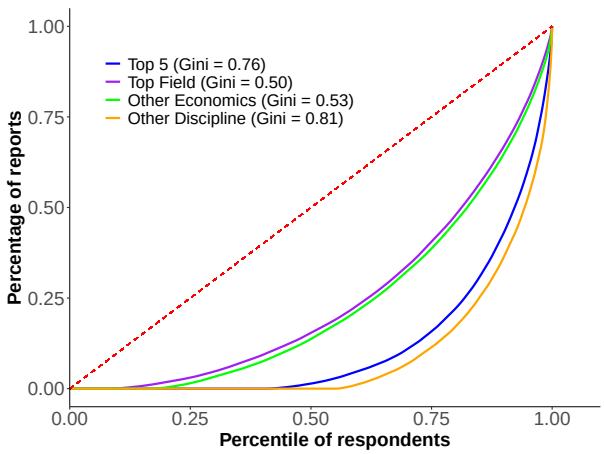
Researchers reported spending a non-trivial amount of time on these reports, with the median respondent spending a total of 9 working days per year, and those at the 90th percentile spending 25 days or more. In line with this, researchers considered their role as referees to be quite important. To better understand what motivates our respondents to serve as referees, we asked them to rank five possible benefits of refereeing. They emphasized most the opportunity to "*ensure the right papers are published*" and to "*read papers attentively*," with about 43% and 28% ranking each as the top benefit, respectively. Frequent top 5 referees tended to assign more importance to ensuring that the right papers are published, and less to becoming a better writer or getting to know the editors. Meanwhile, junior researchers put more weight on getting to know the editors and learning from the opinions of others, and less on ensuring the right papers are published (Appendix Figure B.4).

Figure 2: The peer review efforts of referees

(a) Distribution of annual number of reports written



(b) Lorenz curve for reports written by journal type



Notes: 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.

3 Who should be the gatekeepers?

In the previous section, we saw that a small number of referees write a large share of referee reports. It is not ex-ante obvious whether this degree of concentration is optimal. On the one hand, relying on a handful of trusted, experienced referees could allow editors to elicit dependable signals of manuscript quality. It might also be more efficient in terms of turnaround time. Finally, allowing senior referees to take on more responsibility likely relieves pressure on the workloads of junior researchers, who might rather focus on producing research and securing tenure.

On the other hand, the current distribution might be putting excessive strain on the workloads of some referees. It also restricts the breadth of signals editors receive, which could stifle the publication of bold and innovative research, if incumbent gatekeepers tend to give conservative and homogeneous assessments. In consideration of these trade-offs, we use this section to explore the current allocation of responsibilities, the possible margins of reallocation (i.e., quantity and quality), and potential reforms to improve the status quo.

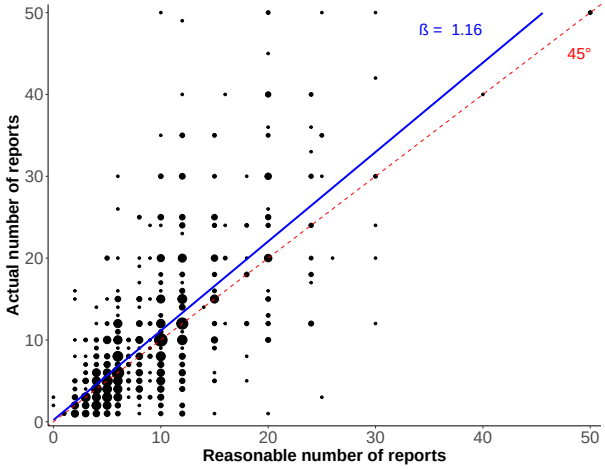
3.1 Allocation of Resources

3.1.1 Quantity of reports written.

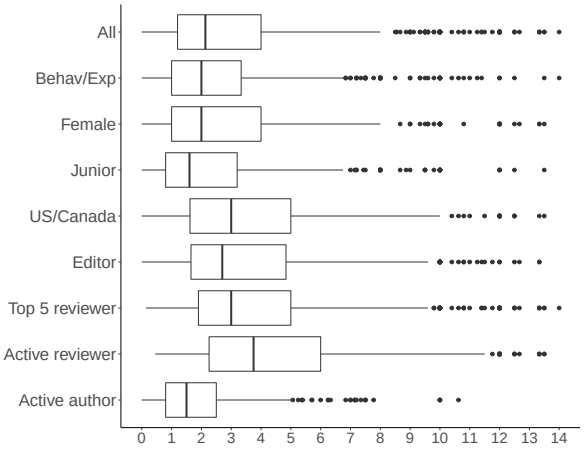
We begin by examining the current distribution of refereeing responsibilities and our respondents' level of satisfaction with it. One indication that many researchers might feel overburdened with the number of requests they receive is that about 66% of respondents that rejected a referee request (or were tempted to reject one) cited a "*lack of time*" as a reason for doing so.[5] To evaluate this more systematically, we compare the number of reports that respondents reported writing with the number they think is reasonable to write.[6] There is a strong correlation between these quantities at the aggregate level (Spearman $\rho = 0.73$, p-val < 0.01), but this obscures the fact that only 25% write the number of reports that they deem reasonable (Panel (a) of Figure 3). Among the remaining 75%, there was significant heterogeneity in whether they reported writing more or fewer reports than is reasonable. On average, tenured professors said they write more, while PhD candidates and postdocs said the opposite. These findings (cautiously) suggest that some refereeing work could be reallocated to early-career researchers who might benefit from more practice and exposure to emerging research.

Figure 3

(a) Actual vs. reasonable number of referee reports per year (individual level)



(b) Reports written to submissions per year



Notes: In Panel (a), 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). In Panel (b), the ratio was created by dividing annual reports written by average annual submissions. In order to improve clarity, outliers with a ratio > 14 were removed (57 respondents - 3.9%). $N = 1,478$ (1,474 without outliers) for Panel (a) and $N = 1,459$ (1,402 without outliers) for Panel (b).

An alternative approach to examining the current distribution of responsibilities is to compare the number of submissions made by researchers to the number of reports they write. While we acknowledge that equal supply and demand for all researchers would not necessarily be socially optimal, the low level of correlation between the two quantities is striking (Spearman $\rho = 0.22$, p-val < 0.01). The ratio of average reports written to average submissions is 2.8, with substantial heterogeneity: while about 40% of respondents wrote at least 3 times as many reports as they made submissions, around 15% wrote fewer than 1 report per

[5] Respondents were allowed to make multiple selections. 1,383 respondents indicated that they rejected or were tempted to reject a referee request at least once over the two years prior to answering the survey.

[6] In survey Q28, we asked respondents "What do you think is a reasonable number of reports to be assigned per year?" This might have been ambiguous, if respondents did not know whether to answer for themselves or for a representative referee. We asked respondents in a follow-up survey ($N = 117$; more details in Appendix C) 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-val < 0.01). In the follow-up survey, average differences between *actual* and *reasonable* are similar to those from the original survey, both when using the number of reports respondents deemed reasonable for themselves and for a representative referee. These similar relationships also hold for the subgroup of junior researchers. As a result, we proceed with using our original survey results.

submission. Researchers based in the US/Canada, active referees, and referees for top 5 journals tended to have higher reports-to-submissions ratios (Panel (b) of Figure 3). We note that the optimal distribution of these ratios is unclear, since it depends on what is assumed regarding desk rejection rates, the number of reports solicited per manuscript, and normative considerations based on reciprocity and fairness.

3.1.2 Quality of reviewer-manuscript matches.

For any given distribution of the peer review workload, mismatches between a manuscript and the referees assigned to review it can cause problems for authors and reviewers. These *quality* mismatches can take many forms e.g., if researchers are requested to review manuscripts that are (i) outside of their area of expertise, (ii) on topics that they no longer work on, or (iii) from authors with whom they have a conflict of interest. By ensuring high-quality matches between referees and manuscripts, journals can enhance the informativeness of referee evaluations and improve the efficiency of the peer review system.

Attesting to the prevalence of this issue, our respondents reported frequent mismatches between review assignments and their skills/interests. About 60% (30%) of them rejected more than 1 in 10 (1 in 4) requests over the surveyed period. Rejection rates and peer review workloads are only weakly correlated, suggesting that match quality issues are also salient. Corroborating this, our respondents who rejected or were tempted to reject a referee request often cited receiving manuscripts that are "*too remote from [their] research field*" (50%) and/or for which they were "*[unable] to judge the paper*" (25%) as reasons for this.

Another potential source of quality mismatch relates to referee conflicts of interest, although there are competing perspectives on this issue. Carrell et al. (2022) and Colussi (2018) present evidence suggesting that network effects for authors connected to referees and editors exist in various settings. As such, restricting the referee pool to senior academics is likely to help incumbent researchers. Brogaard et al. (2014) find similar results, but they also note that papers from authors connected to editors receive more subsequent citations than others. This latter finding may suggest some role for a reduction in the search costs of finding good papers, as opposed to pure nepotism (Laband and Piette, 1994b). Regarding our respondents' attitudes on this issue, 1 in 6 of those who rejected or considered rejecting a request mentioned a "*conflict of interest*" as a reason for doing so. However, there is disagreement on what constitutes an incurable conflict. In fact, 40% think that reviewing friends/coauthors should generally not happen but is sometimes unavoidable. Complicating matters, journals in economics generally do not offer guidance on how to handle referee conflicts of interest, with the *Journal of Political Economics* (*JPE*) being the only top 5 journal that does so (JPE-2). The issue of conflicts further highlights the apparent trade-off between soliciting reports from possibly biased experts versus obtaining more neutral ones from potentially less qualified scholars.

3.1.3 Understanding the current allocation.

There are at least three (non-exclusive) mechanisms that could help to explain the patterns observed above. First, there might be informational frictions that prevent editors from learning about the interests and workloads of referees. Second, increases in manuscript production and enhanced pressure on researchers to publish well could be generating excess demand for (superfluous) review efforts in the system. Third, there might be increasing returns to editors from working with the same referees. The first two mechanisms are potentially indicative of inefficiencies. The first suggests that editors may want to make better manuscript-reviewer matches or distribute refereeing responsibilities more evenly, but the search costs of finding qualified referees or referees with excess capacity are prohibitively high. The second mechanism proposes that elevated levels of demand for reports are potentially overburdening the system. Azar (2004) estimated that a paper of average quality is likely to be submitted to journals between 3.2 and 5.5 times before being published somewhere. We suspect that this has risen since then, in tandem with increases in manuscript production and journal selectivity (Card and DellaVigna, 2013). To some extent, this may also be caused by the strong career incentives of researchers to "shop around" at publication outlets, especially at the top 5 journals (Heckman and Moktan, 2020), regardless of the quality of the fit. As such, editors must make increasingly many allocation and desk rejection decisions. Excessive editorial workloads may limit the extent to which they can search for new referees or ensure the quality of reviewer-manuscript matches.

The third mechanism is not necessarily indicative of inefficiency. Rather, it suggests editors do not expand referee pools because they prefer working with a small group of reviewers whom they can trust. This can be beneficial to both authors and editors, to the extent that trusted referees are reliable, return

reports on time, and offer carefully-considered evaluations. Additionally, repeated editor-referee interaction could allow editors to account for referee-specific factors, like a distaste for certain methods or a generally critical disposition, and better filter out any noise in referee signals. Extracting reliable signals is extremely important, as observational evidence suggests that editors tend to make manuscript decisions based on the number of positive recommendations received (Welch, 2014; Card et al., 2020a).[7]

We note that this pattern of concentration might be somewhat self-fulfilling. If editors do not contact researchers that they do not know, then they may never develop the requisite relationships and experience necessary to become reliable referees. Alternatively, if editors are somehow prompted to elicit opinions from outside of their trusted network, they may start to place less weight on referee evaluations and rely more on their own judgment. Comments left by our respondents suggest that they were split on whether more independent editorial evaluations would be a positive development or not.[8]

3.2 Improving allocation

3.2.1 Addressing imperfect information and decentralization.

In view of the challenges outlined above, we consider a range of possible solutions. Moves towards a centralized platform could be beneficial in many respects. Such a platform would allow editors to track referee workloads, availability, interests, skills, and networks in one place. This would be a great improvement in information availability relative to the status quo, wherein referee characteristics are typically unobservable outside of an editor's network. Our respondents were generally favorable towards such a platform (about 49%), but were less convinced that it would actually be useful for improving referee reports (about 37%; see Appendix Figure B.6). Garnering buy-in for centralization efforts could be challenging, but past attempts have succeeded, as with the job market facilitated by the American Economic Association (AEA) (Coles et al., 2010). Moreover, certain institutions (i.e., Elsevier and the Web of Science) have already achieved some degree of centralized information within a set of participating journals and reviewers.

Given the costs of setting up a centralized platform, some intermediate steps might be useful to consider. Journals could start by allowing researchers to document their current and recent peer review activity via ORCID, as is done in other disciplines. Another idea could be for the AEA or other professional organizations to run recurring surveys asking reviewers about their interests and availability, and share the information publicly. This could also be organized on a volunteer basis, in a way similar to recent efforts to develop centralized lists of potential seminar speakers from underrepresented groups (see example here).

Another type of centralization might be useful for addressing growing submission and review volumes: centralized manuscript submission and processing. Prufer and Zetland (2007) model a centralized clearing house for articles, where authors submit their manuscript and journals bid for the right to publish it. Many variations on this idea are possible, including having the platform conduct peer review itself. In any case, an appropriately designed system might reduce superfluous review efforts, mitigate "shopping around," and relieve pressure on overworked editors and reviewers. Again looking at more intermediate steps, authors might appreciate a mechanism that allows them to solicit opinions on where to submit their manuscript, without having to make multiple submissions. A platform or discussion forum could be established for this purpose. Other possibilities include facilitating greater coordination within journal families (e.g., as is done to some extent by the AEA and Econometric Society journals) and re-using reports from past submissions more often. Both of these practices are more common in other disciplines, such as in neuroscience where around 60 journals have entered into an agreement to share review materials.

3.2.2 Expanding referee pools and maintaining informativeness.

The next set of reforms addresses the third mechanism i.e., the tendency of editors to rely on small networks of referees. As we note above, there are plausible benefits to editors soliciting opinions from referees with

[7] These papers show that editors rarely overrule referees when their recommendations are unanimously positive or negative. In Welch (2014), manuscripts that receive strictly positive reports are rejected between 0.0% and 14.2% of the time. Those with negative reports are rejected 97.1% of the time. In the case of two referees, most papers with mixed assessments are rejected.

[8] Eight respondents argued in open-text comments that a close mapping between recommendations and decisions makes publication outcomes less arbitrary. But 28 others wished that editors would take a more hands-on approach and make their own independent evaluations more often, even if that means overruling the recommendations of referees.

whom they already have a working relationship. But increasing the number and variety of signals could also help editors make more informed decisions. In this section, we explore proposals that might allow editors to expand their referee pools without losing the informativeness of reviewer evaluations.

We first consider the possibility of simply soliciting more referee evaluations per manuscript. Editorial decisions are currently made with very few signals from referees: even in the most ambiguous cases, it is rare for editors to contact more than three referees (Card et al., 2020a). This is despite the low level of inter-reviewer correlation in economics (Blank, 1991; Welch, 2014; Card et al., 2020a), which could render the opinions of 2-3 referees too idiosyncratic to be useful. In other settings, increasing the number of opinions solicited to 5-10 has been shown to yield a substantial improvement in the average informativeness of signals (DellaVigna and Pope, 2018; Otis, 2022). But this logic has limits, as the marginal benefit of additional signals appears to be declining in the number of reviewers, at least in the setting of grant proposal evaluation (Forscher et al., 2019). Relatedly, signal quality could decline quickly if editors struggle to find referees with the appropriate expertise, although editors would retain the discretion to discard low-quality signals. Another concern around increasing the number of referees is that this could further exacerbate editorial and referee workloads, as editors will have to solicit and sift through several signals of varying quality. Finally, some argue that relying on a few noisy referee signals is actually beneficial, as retaining a stochastic component might enable the publication of bold research that may otherwise be rejected (Armstrong, 1997).

In consideration of these challenges, one relatively low-cost way of increasing the number of evaluations would be to employ "shadow reviewers." In this setup, editors would solicit secondary evaluations from one or more currently underutilized reviewers, on top of the usual 2-3 opinions. The new reviewers would be acknowledged as official referees, have access to the editor's decision letter and other referees' reports, and benefit from the journal's referee incentive schemes. However, there would be no requirement for the editor to consider or transmit the reports of shadow reviewers. This would allow underutilized referees the opportunity to gain valuable experience that could make them more visible and appealing to editors in the future. Concurrently, editors would be able to test whether there is value in soliciting additional opinions and assess the quality of each reviewer's report, without committing to using them. In terms of possible downsides, this would likely require more work from editors and journals in terms of finding shadow reviewers and managing interactions with them. However, external assistance could be provided e.g., via professional associations running surveys among their members to ask about refereeing capacity. Another limitation of this proposal is that it does nothing to decrease the current burden on overutilized referees, unless editors begin to substitute incumbent referees with those who started on the shadow track.

3.2.3 Rethinking the role of referee.

Instead of simply scaling up the number of referees, we now consider changes to the nature of refereeing. One possibility is for editors to request specialized opinions from each referee. In this scenario, each referee is only responsible for evaluating one part of the manuscript, depending on their expertise. For example, a senior researcher could be asked to judge the contribution and importance of a paper, while an expert in applied econometrics could assess its identification strategy. This division of labor could allow overutilized referees to share their workload with junior researchers, who would gain some valuable refereeing experience. Some editors have reported efficiency gains from editorial specialization (McAfee, 2010), so similar effects could be present under referee specialization.

On the other end of the spectrum from specialization, editors could leverage the wisdom of the crowds. This has worked well in other realms, such as the aggregation of product information in online retail (Reimers and Waldfogel, 2021) or the use of prediction markets to assess the replicability of research (Dreber et al., 2015; Nosek et al., 2022). Collecting crowdsourced opinions could mitigate concerns around noisy taste-based signals and low inter-reviewer reliability. But dramatically scaling up the number of referee opinions solicited per manuscript necessarily implies that each reviewer will need to review many more manuscripts and/or that the referee pool will need to be expanded. To keep this system tractable, reviewers would likely have to reduce the dimensionality of their assessments and condense evaluations into manageable statistics, like "upvotes" or prediction market prices, for use by editors. In turn, this could end up making individual signals noisier than they were initially. Additionally, asking referees to make many snap judgments could exacerbate biases and induce them to use heuristics more often.

These problems might be minimized if crowdsourced peer review is used as a complement to peer review,

rather than as a substitute. To make progress, a journal could pilot the crowdsourcing of evaluations alongside traditional peer review for a subset of manuscripts, to see if editors find it useful. Moreover, in a manner similar to Gordon et al. (2022), the outcomes of these manuscripts could be tracked over some period of time and the relative performance of each system could be compared.

3.2.4 Changes to the content of reports and recommendations.

Another class of reforms that we consider would restructure how referee opinions are elicited. As it stands, most journals (at least in Welch (2014) and Card et al. (2020a)) request referees to write a report and give a recommendation on a grid with several options, which range from "definitely reject" to "accept." This requires referees to reduce their qualitative, multidimensional assessment into a single categorical recommendation. That is, they must decide (i) how to weigh the individual components of quality and (ii) how to set the quality threshold sufficient to recommend an R&R. These decisions should vary based on journal-specific standards, but editors rarely provide explicit guidance on (i) and (ii). Additionally, when mapping their assessments into recommendations, referees may rely on heuristics and biases unrelated to the content of the manuscripts such as author institution (Ersoy and Pate, 2021), status (Huber et al., 2022), and gender (Card et al., 2020b; Clain and Leppel, 2017). A possible symptom of this problem is that, while referee recommendations are predictive of citation impact (Card et al., 2020a), some algorithmic models of manuscript characteristics appear to perform better (Abramo et al., 2019).

Instead of just providing a recommendation, reviewers could be asked to rate manuscripts on more objective criteria, as a complement or substitute to existing recommendation grids. For example, the *Journal of Economic Behavior and Organization* (*JEBO*) asks reviewers about the novelty of the contribution and the carefulness of the analysis, among other attributes. This could allow editors to extract the exact information they want while leaving minimal room for misinterpretation. One downside is that referees might find the scales constraining or might retroactively adjust them to match their subjective recommendation. Tying this to our earlier discussion, journals could run prediction markets on particular characteristics of the manuscripts, such as the replicability and generalizability of their findings (Delios et al., 2022).

4 Can we speed up peer review without sacrificing quality?

Alongside allocation and evaluation issues, long turnaround times are another challenge facing the field. As we will discuss again in Section 6, comparative research indicates that economics is particularly slow relative to other disciplines (Huisman and Smits, 2017; Hadavand et al., 2022). Moreover, the field is slower than it used to be decades ago (Ellison, 2002a,b), with little evidence of improvement in more recent years (Hadavand et al., 2022). Pressures on turnaround times include growing submission volumes (Card and DellaVigna, 2013), the increasing length and technical nature of papers (Schwert, 2021), and enhanced documentation requirements, such as pre-analysis plans (Olken, 2015) and reproducibility practices (Vilhuber, 2019).

Our respondents placed some importance on the speed of peer review decisions, but most prioritized receiving a reasonable and well-substantiated decision over timeliness (Section 2.2). This highlights the potential challenge we explore in this section i.e., that reductions in turnaround time are not entirely costless and the optimal time between submission and decision is unclear.[9] We note that while additional revision rounds do seem to be associated with increased citation counts, longer delays within each round appear to be counterproductive (Hadavand et al., 2022). Moreover, observational evidence shows that researchers from cohorts who face longer publication delays than previous cohorts end up publishing fewer papers (Conley et al., 2013). These findings suggest that the length of the peer review process may be harming researcher productivity, though as we discuss below, there might be ways of improving efficiency without sacrificing on quality.

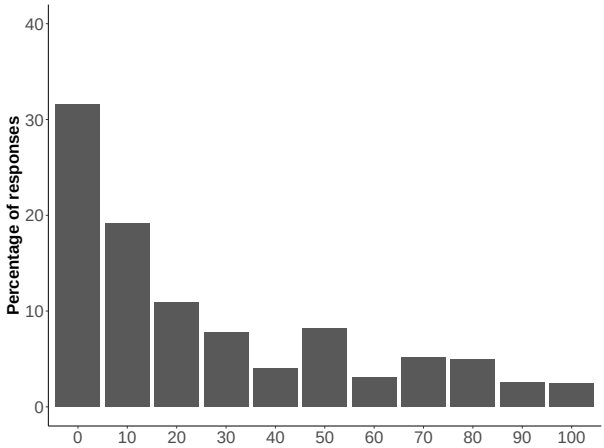
4.1 Understanding publication lags.

Since most submissions yield a rejection and first response times are a material contributor to the review duration of accepted papers, we begin our discussion by examining first-round lags. They can be decomposed

[9] To take a humorous example illustrating this trade-off, the *Journal of Universal Rejection* reports a turnaround time of just a few hours. The cost of this is that it desk rejects all manuscripts it receives.

into four stages: first, the editor decides whether to desk reject a paper or not; next, submissions are matched to reviewers; after that, the editor waits to receive all completed reports; and finally, the editor makes a decision after all reports are received. While delays occur at all of these stages, third-stage lags tend to be the most substantial (Charness et al., 2022).[10] In our survey data, only about a third of respondents reported always being punctual in returning their reports, while around a quarter were late most of the time (see Figure 4). These delays are also non-trivial, with the median length being 1-2 weeks past the report due date. But if editors wait for all reports, then the latest referee matters more than the median one. For context, the 90th percentile of respondents reported being late by 3 weeks or more. As such, individual referee performance alone can delay decisions by several weeks in excess of the agreed-upon timelines.

Figure 4: Percentage of time respondents had a delay



Notes: $N = 1,483$.

To a limited extent, lateness may be driven by some referees abiding by their own ideal deadline, rather than the official one. Our respondents generally considered 1-2 months to be a reasonable turnaround time, as most indicated that they find either 6 weeks (26%), 4 weeks (25%), or 8 weeks (20%) to be most appropriate. There is little difference between the preferred timelines of referees who are never late and those who tend to be 1-2 weeks late; the median deadline preferred by each group was 6 weeks (Appendix Figure B.7). However, among those with substantial lags (3+ weeks), the median preferred deadline was 8 weeks.

Workloads might help to explain delays as well, although the relationship between reports written and lateness is generally weak (see Appendix Figure B.5). The exception is among the least active reviewers, who are rarely late. Interestingly, more delayed respondents do not turn down requests at a significantly higher rate than others (Spearman $\rho = 0.08$, $p < 0.01$), despite writing more reports than they deem reasonable.

After an initial decision is made, further revision rounds can more than double total submission-to-publication lags (Hadavand et al., 2022). For the manuscripts with the most extreme lags in Hadavand et al. (2022), the cause appears to be delays in authors' implementation of requested revisions. In our survey, several respondents expressed immense frustration with revision rounds, both for the delays they induce and for the demands made by referees (20 open-text comments). Moreover, others lamented the lack of editorial direction about which revisions suggested by referees are necessary to implement (27 comments).

4.2 Improving turnaround times.

4.2.1 Holding reviewing volumes constant.

In light of these issues, we will consider a range of reform proposals aimed at reducing turnaround times. One challenge to these ideas is that shorter turnaround times effectively reduce the cost of making new

[10] There are some notable exceptions to this tendency. For example, second-stage lags might differ substantially by journal. Lower-profile journals might have an especially hard time finding referees for review assignments, as 41% of respondents who rejected a referee request (or were tempted to reject one) cited the lower rank of the journal as a reason for doing so.

submissions. That is, they could induce authors to "shop around" at journals more often than they already do, which we argued earlier could lead to a glut of review efforts across journals. Some studies have explored ways of reducing turnaround time while holding submission costs constant, including by increasing submission fees (Azar, 2005, 2006; Cotton, 2013; Zheng and Kaiser, 2015). Journals might also consider increasing desk rejection rates as a low-cost method of improving turnaround times and reducing the volume of superfluous reviews (Cherkashin et al., 2008). Approximately 75% of our respondents approved of the practice of desk rejections in general, but there are many factors to consider in determining the optimal rate. First, journals should carefully weigh how changes to the desk rejection rate affect the risks of Type I error (sending manuscripts to review that should have been desk rejected) and Type II error (desk rejecting manuscripts that should have been considered). Next, journals must decide which additional manuscripts to desk reject. There is some evidence that desk rejection decisions are associated with author characteristics, such as submission history and gender (Clain and Leppel, 2017; Card et al., 2020b; Schwert, 2021). If editors are forced to desk reject more manuscripts, this could exacerbate concerns around equity in peer review.

4.2.2 Reducing turnaround times.

With these caveats in mind, we start by considering proposals to address referee turnaround times. In a field experiment, Chetty et al. (2014) tested three treatments with the aim of reducing these lags: (i) a *social condition*, where referee response times were made publicly available; (ii) a *deadline condition*, where referees were requested to return their report within 4 weeks instead of 6 weeks; (iii) and a *financial condition*, where referees were paid $100 for returning their report within 4 weeks. The latter two treatments were the most effective at reducing turnaround times. The social treatment reduced the acceptance rates of referee invitations, while the financial treatment increased them. No negative effects on report quality were observed for any treatment, although the authors could only use report length and referee-editor agreement as proxies for this. In sum, changing the report deadline and paying referees appear to be effective with minimal downsides; the former even appears to be nearly costless to editors and journals.

The effects of payment on turnaround times are consistent with those seen in observational studies (Hamermesh, 1994; Thompson et al., 2010). Offering financial incentives was also popular among our respondents. Among the 1,080 respondents who thought that referees would perform better if better rewarded, about 75% thought that referees should receive a payment for timely completion. Economics stands out in this regard, as paying referees for their reports is almost unheard of in other disciplines (see Section 6).

4.2.3 Addressing revision rounds.

Given that revision rounds often drag on much longer than initial submission rounds (Hadavand et al., 2022), more structural changes to the review process might need to be considered. One simple reform could be to cap the number of rounds that a manuscript can go through. This might reduce total submission-to-publication lags and force editors and referees to consider the revisions they request more carefully. The other effects of this policy are more ambiguous. If referees and editors are not entirely satisfied with a manuscript after the maximum number of rounds, they might be more inclined to reject it than to publish it as-is. This policy might also affect the quality of published manuscripts. Qualitatively, many have argued that more revision rounds can worsen papers, as excessive demands from referees can make them unnecessarily long and obscure their main findings (Berk et al., 2017). For what it is worth, most research has found that additional revision rounds have a positive effect on manuscripts in terms of subsequent citation counts (Laband, 1990; Rigby et al., 2018; Hadavand et al., 2020, 2022) and quality assessments (Goodman, 1994); however, benefits are heterogeneous. In any case, it is not clear that these purported benefits of revisions are worthwhile, as additional rounds of review often imply massive time and labor costs for authors, referees, and editors.

One journal that has successfully implemented a limited-revision model is *AER: Insights*. Their short-paper, accept-or-reject, one-revision round format was very well-liked among our respondents, with approximately 75% reporting favorable attitudes towards it. This makes it one of the most popular proposals considered in our survey. Certain components of it, especially short-paper tracks, have also been implemented at other journals. However, these tracks are not suited to all types of articles, due to significant constraints on manuscript content.

Another possibility could be to make revision rounds more interactive by opening up further communication channels between authors, reviewers, and editors. In this spirit, the *Frontiers* journals host an

additional round of "interactive review" where the parties can communicate directly in an online forum after a traditional round of revision. This approach could eliminate back-and-forth discussions that lead to more revision rounds by allowing for questions and immediate resolutions. However, it would require more attention from reviewers at specific times.

4.2.4 Rethinking the timing of peer review.

More generally, we could consider proposals that shift the *timing* of peer review. Economics largely adheres to a "pre-publication" review process i.e., completed manuscripts are fully peer reviewed before publication (COPE Council, 2017a). As we have seen, this system is associated with massive submission-to-publication lags. This can cause problems, especially when (i) researchers are racing against the tenure clock and (ii) the publication contains information of urgent interest to the general public. Other possible systems include a "preprints" process where reviewers can comment on manuscripts in public archives and a "post-publication" process where peer review happens (in some form) after a journal publishes the manuscript.

These approaches are ideal for the rapid dissemination of research. Many authors in economics appear to appreciate the preprints approach, as sharing pre-publication manuscripts on platforms like NBER, IZA, CEPR, RePEc, and SSRN is already quite common (Ellison, 2011). This appears especially prevalent among post-tenure academics who do not face the same pressures to publish as junior researchers (Heckman and Moktan, 2020). While post-publication peer review is largely unheard of in economics, journals such as *F1000Research* practice it (F1000-1-a). Each of these systems can mechanically bring dissemination times down to (nearly) zero, but they come with risks. Many are concerned that manuscripts widely distributed prior to peer review can lead to the spread of misinformation, as updates and retractions often go unnoticed (Özler, 2011; Bolboacă et al., 2019; Gonçalves et al., 2021). If post-publication peer review is too risky, then fast-tracks for academics with upcoming tenure deadlines or for research on pressing topics (e.g., for COVID-19, as was done at the *Journal of Urban Economics* (JUE-1)) could be a useful replacement.

Another proposal is for journals to establish pre-results review tracks. In this setup, authors can submit a detailed project design and data analysis plan to the journal; the journal can then reject the submission or commit to publishing it regardless of the results. An example of this comes from the *Journal of Development Economics*, which allows authors to submit Registered Reports for peer review (JDE-1-a).[11] This could reduce turnaround times if it helps to prevent later complications and miscommunications. Further purported advantages include improvement in the eventual quality of papers (Foster et al., 2019) and the possibility that it could facilitate the pursuit of riskier research. However, the *JDE* editors report that they have not yet noticed any such between-track differences in "riskiness" (JDE-1-a). A final benefit, at least for Registered Reports, is that they appear to be useful for combating publication bias (Scheel et al., 2021). In view of this, a number of journals in economics have recently established Registered Report tracks (e.g., at the *Journal of Political Economy Microeconomics* (List, 2023) and *Oxford Open Economics* (OOE-1), and at the *Journal of the Economic Science Association* with a Registered Report track for replications (JESA-1).

Not all are convinced of these purported benefits, however. Referees and editors may struggle to assess a manuscript without seeing the results, especially if the authors fail to fully flesh out their proposed analyses, perhaps relying too much on the peer review team to resolve their own doubts. Additionally, imposing new review requirements (if in addition to previous ones) may instead exacerbate turnaround times (Vilhuber, 2019). Others also warn that reviewing projects pre-results actually disincentivizes the pursuit of bold research (Gross and Bergstrom, 2021), but whether this is the case is not clear to us. Further research into Registered Reports and pre-results review in general is likely needed to resolve these disputes.

Although it is not yet common, journals might also consider experimenting with a results-blind review process. In this setup, the results are known to the authors but editors and referees are asked to evaluate manuscripts without consideration of or access to them.[12] One issue is that authors might select into results-blind review after generating lackluster results, a concern that does not exist with Registered Reports.

[11] See Arpinon and Espinosa (2023) for more discussion and practical guidance on Registered Reports.

[12] One way of implementing this was proposed by Dufwenberg and Martinsson (2019). They suggest submitting information about empirical analyses and results in a sealed envelope. The envelope can only be opened after a manuscript decision is made.

5 How should we balance transparency and confidentiality?

Regardless of the form of peer review that we practice, difficult decisions about the confidentiality of authors and reviewers must still be made. Movements towards a system of "open science" have seen success in recent years, such as with pre-analysis plans (Olken, 2015) and reproducibility policies (Vilhuber, 2019). But open *peer review* practices i.e., reforms that increase the level of transparency in the evaluation of research, have seen much less adoption (Wolfram et al., 2020). A general exception to this has come from the Nature journal family. In 2016, *Nature Communications* began publishing referee reports subject to author consent (Nature Communications, 2015). The opt-in rate rose from 60% in 2016 to 70% in 2022, after which they started publishing all reports without exception (Nature Communications, 2016, 2022). Other journals in the family have since then followed with their own opt-in policies. As part of a parallel initiative, *Nature* and other affiliated journals have also begun publishing reviewer identities alongside manuscripts, subject to author and referee consent (Nature, 2019). Most authors and referees have been open to this practice as well, with around 80% of *Nature* manuscripts featuring at least one referee identity during the trial period.

Given the success of these initiatives, we sought to understand the arguments for and against increased transparency in economics, as well as our respondents' attitudes towards them. Our findings in this section suggest that the optimal level of transparency is unclear. Following the *Nature* initiatives, we explore proposals that (i) modify the identifiability of stakeholders and (ii) allow for the publication of peer review documents. These reforms could promote accountability and provide readers with useful information about manuscripts. At the same time, they might have real costs to the extent that the piercing of confidentiality affects stakeholders' (perceived) ability to deliver honest assessments. We offer perspectives on these issues from our respondents and from the experiences of other disciplines, when possible.

5.1 Author anonymity.

Historically, journals in economics have switched between a single-blind (authors blinded to referees) and a double-blind (authors and referees blinded to each other) model of refereeing (Hengel, 2021). While double-blind reviewing is common at top journals in other social sciences (see Table 1), the current consensus among top 5 journals in economics is single-blind. However, we note that other journals in economics deviate from this norm, such as *Economic Inquiry* (EI-1). Extending this further, one could even make peer review triple-blind i.e., by blinding editors to author identity as well. Some, such as Jung et al. (2017), have proposed ways of implementing this. The main argument in favor of blinded review is that it could protect authors from the identity-based biases of reviewers and editors. The state of the evidence on double-blind review is quite mixed (Blank, 1991; Laband and Piette, 1994a; Snodgrass, 2006; Kolev et al., 2019; Ersoy and Pate, 2021; Carrell et al., 2022), with many studies finding a null effect of it as a means of reducing bias.

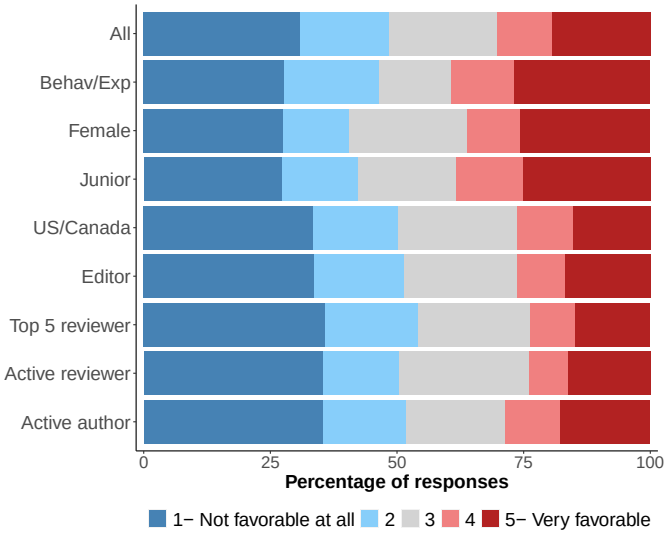
One reason to be skeptical about blinded review is that it is often infeasible in practice. Small research networks and the widespread distribution of preprints make it difficult to preserve author anonymity. This may be especially relevant in economics, where long turnaround times incentivize authors to seek maximum visibility for their preprints before they are published. In disciplines where authors typically keep their projects confidential until journal submission, blinded review may be more effective.

Attesting to these concerns, the editors of the *AER* declared their belief that double-blind anonymity had become nearly impossible in the age of search engines (Goldberg, 2012). Another reason that it is uncommon in economics may simply be that it is unpopular: only 20% of our respondents were very favorable to a system of double-blind reviewing, while nearly half were unfavorable (see Figure 5). In terms of heterogeneity, female and junior respondents tended to be the most supportive. To accommodate differing views, journals could adopt an opt-in model of double-blind review, similar to *Nature Human Behaviour* (NHB-1-a). However, authors who choose this option might signal to referees that they belong to an underrepresented group, thereby mitigating the benefits of anonymity. As such, this option would need to be used with caution.

5.2 Referee anonymity.

Referee anonymity is generally easier to preserve, which makes its benefits more plausible. One piece of evidence that referees appreciate the veil of anonymity comes from Chetty et al. (2014). In the experiment, the referees who were told that their turnaround times would be publicly shared were more likely to turn

Figure 5: Support for double-blind peer review



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

down refereeing requests. One can imagine similar effects if journals announce they will publish referee names alongside manuscripts. But in terms of the downsides of the status quo, a lack of identifiability among referees means that there is limited accountability for what they write. Reports with vague, inaccurate, and inappropriate content are therefore sometimes transmitted to authors who have little recourse. While our respondents were overall moderately satisfied with the quality of the reports they receive, a non-negligible percentage find their reports to be unsatisfactory (see Figure 6). The most common characteristic of low-quality reports was that they make "*inaccurate statements about what the paper does*," as reported by about 75% of our respondents. "*Vague and unconstructive comments*" were also cited by about 65%. Such problems are not unique to our field (Silbiger and Stubler, 2019), but research from other settings in economics and related disciplines suggests that underrepresented groups can be particularly affected by unprofessional behavior (Allgood et al., 2019; Wu, 2019; Dupas et al., 2021; Milkman et al., 2015).

If anonymity is what enables referees to write unconstructive reports, then increasing transparency might be a solution. We first consider a policy of publishing referee identities alongside manuscripts without linking them to a specific report.[13] This could have at least two additional benefits. First, identifying referees can help to expose unrecorded conflicts of interest (Benos et al., 2007). Second, it could increase the visibility and prestige of refereeing. It would now be possible for referees to highlight specific reviewing assignments on their CVs. This could also facilitate greater weight being given to peer review contributions in tenure, promotion, and hiring decisions, as suggested by several respondents (34 open-text comments). But as we argued above, greater transparency could have many downsides, including decreased participation in peer review and a diminished willingness to write critical reports, as referees may wish to avoid angering colleagues and provoking retaliation. Our respondents seemed to have these arguments in mind, as they generally opposed increasing the identifiability of referees and did not think such changes would be useful.[14]

As such, more conservative approaches to these issues might be worth considering. One possibility is to allow referees to opt-in to revealing their identity, after peer review is complete and conditional on acceptance. Both *Nature* and *Nature Human Behaviour* have such an opt-in policy (NHB-1-b). This type of identifiability may not improve accountability, but referees could use it as a commitment device to write quality reports and it could help to shift norms around identifiability over time. Next, in lieu of policies that directly reveal referee identities, authors could be allowed to nominate or oppose the assignment of particular referees to their submissions. This may give authors some added confidence that their manuscript received a "fair trial" from referees without actually knowing who reviewed it. Moreover, it allows them to communicate their

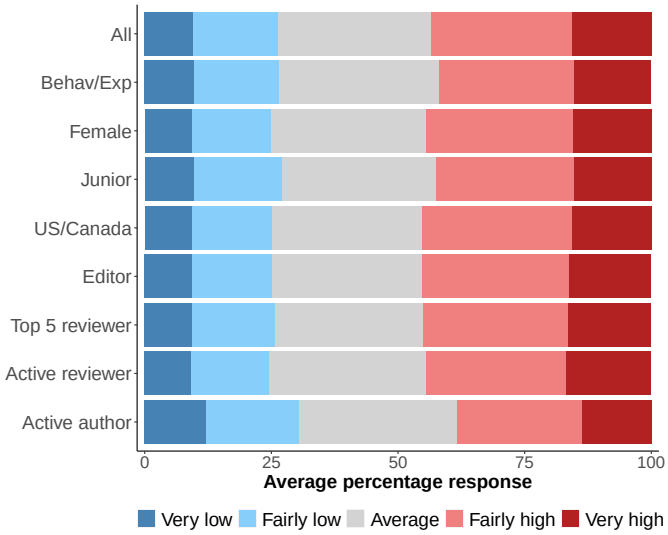
[13]This would still allow for some plausible deniability about report authorship. However, it becomes considerably easier for authors to guess who wrote low-quality reports if referee names are attached to a manuscript.

[14]For example, about 45% of our respondents did not think lifting the anonymity of senior referees would be useful.

concerns about certain referees in advance of manuscript decisions, which editors may find more credible than ex-post complaints about the neutrality of specific referees under a policy of identifiability. This practice is commonplace in other disciplines, but none of the top 5 journals in economics have official procedures for it (see Table 1). Although it could give authors an advantage in the publication process (Teixeira da Silva and Alkhatib, 2018; Moore et al., 2011; COPE Council, 2016), more than half of our respondents were favorable to a policy of allowing them to oppose certain referees (see Appendix Figure B.8).

Another possibility could be to give authors the right to respond to referee reports. This would create a mechanism for authors to challenge unconstructive reports while still conserving referee anonymity. Additionally, the prospect of an appeal process might motivate reviewers to write better reports. The majority of our respondents were in favor of such a policy, which we described as allowing authors to submit a single response to the referees and the editor without any "guarantee of the referees taking this rebuttal into account" (see Appendix Figure B.9). Interestingly, support for this policy from editors and active reviewers is only mildly weaker, despite the potential extra work implied by an appeal process.

Figure 6: Perceived quality of reports received



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

5.3 Publication of peer review documents

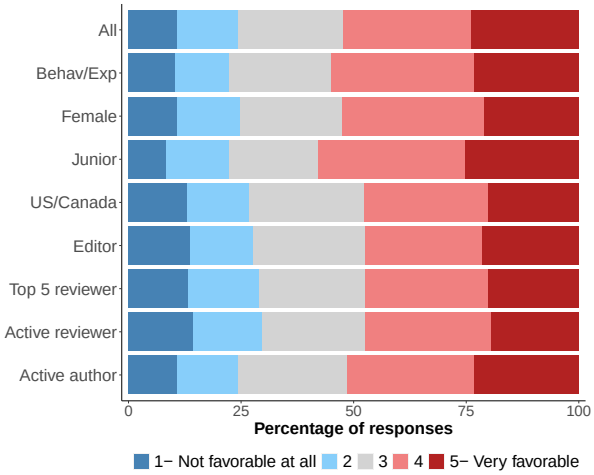
We now seek to investigate the publication of peer review documents separate from identifiability. The sharing of these documents could enhance our understanding of how the revision process shapes research. In particular, publishing referee reports, editorial letters, and prior manuscript versions would allow us to observe how manuscripts evolve in response to referee and editorial input. Further metadata on manuscripts could also be disclosed, such as initial submission dates and turnaround times. Encouragingly, evidence from trials at Elsevier and *Nature* suggests that releasing anonymized reports did not have any negative impacts on reviewer behavior (Bravo et al., 2019; Nature, 2022). We asked respondents for their views on publishing these documents "*in an anonymized way, unless the reviewers choose to disclose their identity.*" A near-majority expressed favorable views, with junior researchers being the most favorable (see Panel (a) of Figure 7). This would suggest that our respondents would prefer some increase in transparency in the peer review process. However, they were not especially convinced that this practice would actually improve review quality, with only about a third of respondents expecting it to be useful. Moreover, the feasibility of maintaining referee anonymity when reports are published is debatable. Report layout and recommended citations are both potential identifiers of reviewers, even if names are not explicitly revealed.

Finally, we consider a policy of publishing editorial decision letters and signed referee reports. This type of openness would strongly promote accountability, as referees and editors would have to write their evaluations with the knowledge that their content will become public. But it might also provoke the most

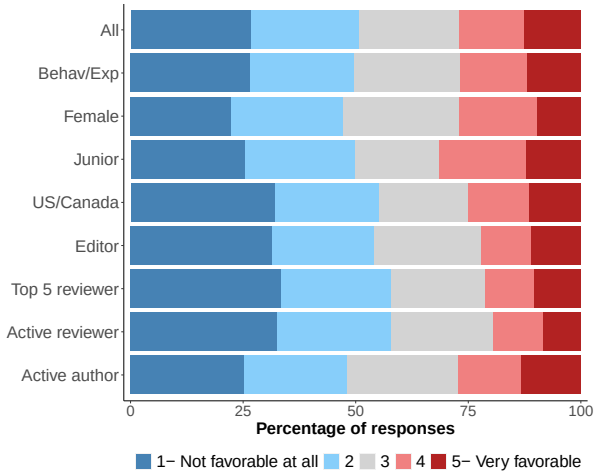
severe reduction in their willingness to give honest assessments, relative to the previous policies considered (if there are any such reactions). Some evidence on this comes from van Rooyen et al. (2010). They find that telling reviewers their signed report may be published reduces invitation acceptance rates and causes longer turnaround times, with no discernible effect on report quality. Consistent with this, our respondents were skeptical about a system where "*referees sign their reports and the entire review history (including responses to referees) is disclosed*," whether this applies to all referees (Panel (b) of Figure 7) or to senior referees only. None of the journals in our interdisciplinary sample have embraced this extent of transparency (Table 1). In fact, only a handful of journals have done so in general, such as *F1000Research* (F1000-1-b).

Figure 7

(a) Support for disclosing the review history



(b) Attitudes towards open peer review (for all referees)



Notes: $N = 1,459$ for both panels.

6 Is Economics Unique?

Having now explored the peer review process in economics in detail, it will be instructive to contextualize our system with respect to the evaluation processes in other disciplines. There are two main benefits of doing this. First, it will help us to understand to what extent economics' problems are common across disciplines or, alternatively, whether some issues we face are unique or unusually severe. Second, it may shed light on whether solutions that work in other fields could be imported into economics and the extent to which field-specific characteristics might diminish the effectiveness of certain policies.

To guide this discussion, we gathered comparative data on norms, policies, and performance in peer review across five disciplines. We contrast top journals in economics with general science journals (e.g., *Nature*, *Science*) as well as top journals in medicine, political science, and psychology. We chose these disciplines in order to have a mix of journals from the hard and social sciences, as well as to highlight the unique policies and initiatives pioneered by journals in the general science category.

6.1 Economics within context.

Our findings are summarized in Table 1. The first result is that economics is a relatively small discipline within the comparison group. Among those sampled, the average top journal in economics published 241 articles in 2019-2021, compared to an average of 1,718 (6,259) articles at top medicine (general science) journals. Small scale seems to be a common feature of the social sciences, as the political science and psychology journals published similar averages of 295 and 281 articles, respectively. Similar patterns hold when looking at journal impact factor and h-index as alternative measures of reach. One key implication of these facts relates to the transferability of journal policies that rely on scale to be effective. As we will see, the general science journals seem to perform especially well on metrics like turnaround time, and they also

tend to be innovative. This may partially be a product of their organizational scale and market power. It may also be indicative of more general differences between the social sciences and the hard sciences.[15]

There are three more areas in which the general science and medicine categories stand out as compared to the social sciences. One is their utilization of professional full-time editors, as opposed to part-time editors who also hold academic positions. The returns to professionalization are plausibly increasing with submission and publication volume; for this reason, editorial professionalization might make less sense in economics than it does in other disciplines. Consistent with this, the only journals in our sample where full-time editorial staff are responsible for publication decisions are in the two high-impact categories. Second, the hard science journals appear to enforce stricter requirements on manuscript length than journals in the social sciences. The general science and medicine journals request (on average) that articles contain no more than 5,050 and 3,400 words, respectively. By contrast, the three economics journals with length restrictions request no more than 20,500 words, with exceptions explicitly allowed by each one; the two other journals do not mention any specific restrictions. We noted earlier that increasing manuscript length in economics might be contributing to the field's long-turnaround times (Section 4). While political science journals seem somewhat more stringent, their average maximum word count is still quite high at 13,500. The psychology journals in the sample do not state any explicit length restrictions, but it is unclear whether this instead reflects stronger implicit norms around article length in psychology. Finally, there appears to be more willingness to experiment with transparency policies among the general science journals. *Nature* is the only journal in our sample that publishes the full contents of referee reports, subject to author and referee consent (Section 5). *Nature* and the *Proceedings of the National Academy of Sciences* (*PNAS*) are also the only ones to publish referee identities alongside manuscripts, at least in some situations.

Economics is also a unique category in some respects. It is the only discipline in the sample with journals that pay referees for timely reports, a policy embraced by our respondents (Section 4). Additionally, its top journals are unanimous with respect to conducting single-blind review, in contrast with the other social science journals in the sample, each of which use *double*-blind review (i.e., the political science and psychology journals). Our respondents were generally supportive of this status quo, as well (Section 5). Finally, it is the only discipline in our sample where none of the journals allow authors to nominate or oppose reviewers. In this case, our respondents would actually be open to a change in policy (Section 5).

6.2 A performance evaluation for economics.

Given the differences in policies outlined above, how does the performance of economics compare to that of other disciplines? We must caveat that the available data on this question is inherently limited. Objective measures of the quality and impact of manuscripts do not generally exist, let alone such measures for entire disciplines. One (narrow) comparison we can make is of turnaround times across disciplines. As many authors have noted and as Table 1 confirms, economics is unusually slow (Huisman and Smits, 2017; Hadavand et al., 2022). Both submission-to-acceptance and submission-to-publication lags at economics journals are substantially longer than those at other social science and general science journals. Among the ten fields compared in Huisman and Smits (2017), economics has the longest average first response time, total response time, and time needed for a revision, as reported by authors.[16] Furthermore, economics had the lowest level of satisfaction with the quality of referee reports and the most reported difficulty in implementing revisions.

The tough review process may have some benefits, however. In Huisman and Smits (2017), economics (tied with the humanities) had the highest level of reported improvement that came from implementing revisions. We consider this to be suggestive evidence that the rigor of the peer review process in economics adds value (in addition to studies from Hadavand et al. (2020) and Hadavand et al. (2022)).

[15] Informally, we include medicine and the general science journals in the hard sciences category. The general science journals do not exclusively publish on these topics, but it does represent the majority of their content.

[16] The 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 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.

Table 1: Comparison of top journals across disciplines

Reach	*Economics*	*General Science*	*Medicine*	*Political Science*	*Psychology*
Articles Published (2019-2021) [*AVG.*]	241	6,259	1,718	295	281
Impact Factor (two-year) [*AVG.*]	9	16	23	5	14
h-index [*AVG.*]	241	1,151	908	174	362
Administration and Policies					
Professional Editorial Board [*FRAC.*]	0 of 5	2 of 3	1 of 3	0 of 3	0 of 3
Restriction on Manuscript Word Count [*AVG.*]	20,500[2]	5,050	3,400	13,500	
Authors Can Suggest/Oppose Referees [*FRAC.*]	0 of 5	3 of 3	1 of 3	3 of 3	2 of 3
Referee Report Guidelines Available [*FRAC.*]	2 of 5	3 of 3	0 of 3	2 of 3	0 of 3
Referees Paid for Reports [*FRAC.*]	2 of 5	0 of 3	0 of 3	0 of 3	0 of 3
Desk Rejection Rate [*AVG.*]	50%	69%[1]		43%	
Journal Metrics					
Average Days to First Post-review Decision [*AVG.*]	109	46[2]			
Median Days to First Post-review Decision [*AVG.*]	98	53[1]		74[2]	
Average Days from Submission to Acceptance [*AVG.*]	752[3]			437[2]	437[1]
Median Days from Submission to Acceptance [*AVG.*]		189[1]			
Confidentiality and Transparency					
Double-Blind Refereeing [*FRAC.*]	0 of 5	1 of 3	0 of 3	3 of 3	3 of 3
Public Referee Identity [*FRAC.*]	0 of 5	2 of 3	0 of 3	0 of 3	0 of 3
Public Referee Reports [*FRAC.*]	0 of 5	1 of 3	0 of 3	0 of 3	0 of 3

Note 1: These statistics are based on a small sample of journals from each discipline. [*AVG.*] indicates that this statistic is averaged *across* the sampled journals for each discipline. [*FRAC.*] indicates the *fraction* of sampled journals for which the characteristic applies.
Note 2: The superscripts [1], [2], [3] indicate the number of journals for which a particular statistic is missing. A blank cell indicates that no data is available. Missing values for *Restriction on Manuscript Word Count* imply no stated restriction on the word count.
Note 3: For more information about the data sources and calculations made to produce this table, see the appendix discussion.

7 Next Steps

The great engagement we received from our respondents attests to the importance of our peer review system and the urgency of making changes. We are especially grateful for the comments describing their experiences and ideas. Having now incorporated these insights into the article, we hope to spur further discussions and initiatives aimed at improving our system. We will briefly mention the limitations of our study before enumerating some concrete steps that journals, researchers, and institutions could take to pursue reform.

The first limitation to mention is that our respondent pool is not entirely representative of the profession, even based on observables. 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. Fourth, we cannot guarantee that respondents interpreted our question wording uniformly, potentially leaving some interpretational gaps. Finally, the data produced in this study is merely correlational and provides limited information about the factors that may drive the patterns we observe. This once again emphasizes the importance of running trials and gathering causal evidence on the reforms we have discussed.

We previously noted that some journals e.g., in the *Nature* family, have been especially active in rethinking peer review practices and running the types of trials we are calling for. An example from within economics comes from the *JDE*, which piloted their Registered Reports track in March 2018 in advance of launching the full version (JDE-1-b). Certainly, we understand that it is not easy to trial wide-scale initiatives. As such, below we propose a handful of concrete actions that can be taken as stepping stones toward broader reform. We also highlight a few ideas that are more aspirational but that seem particularly promising or popular. These steps are organized thematically according to the topics in each section.

1. **Allocation of resources.** We explored in Section 3 the possibility of editors contacting "shadow reviewers" as a first step towards expanding referee pools. This policy seems particularly ripe for experimentation, as it could be implemented with fairly minimal disruption to existing evaluation procedures. This proposal would be greatly facilitated by coordination with other institutions, such as professional associations and academic departments, that could run programs to give report-writing feedback to junior referees or promote more frequent co-reviewing between supervisors and their students. Although it would not be a low-hanging fruit, the AEA or a similar institution could start to collect survey data on reviewer availability and interests for use in a central database of referees. This would open up many other possibilities for reform and could catalyze further centralization efforts.

2. **Publication lags.** In Section 4, we noted that our respondents are generally supportive of desk rejections, but that the optimal desk rejection rate is unclear. To address this challenge, journals could start systematically tracking the outcomes of all papers they reject. The eventual publication and citation outcomes of desk rejected papers could be informative about potential Type II error. Similarly, the proportion of positive referee reports for manuscripts on the margin of desk rejection (but sent to review) would be reflective of Type I error. To assist with making desk rejection decisions, editors could experiment with eliciting other types of evaluations, such as those from prediction markets and forecasting surveys or from asking ChatGPT to predict the credibility and policy impact of manuscripts (Van Noorden, 2022; Schulz et al., 2022). A final proposal worth highlighting is for journals to adopt policies or establish tracks similar to *AER: Insights*. This would not be a simple endeavour, but it would address many of our respondents' concerns around the revision process, and it was among the most popular ideas that we have survey data on.

3. **Transparency and confidentiality.** One of the main challenges to implementing proposals from Section 5 is that many require the explicit consent of reviewers and authors (COPE Council, 2017b). To get the ball rolling, journals could start with publishing manuscript metadata, including the submission date, the number of revision rounds, the identity of the handling editor, and the number of reviewers contacted. These steps are likely already within the power of journals and would promote accountability. Moreover, some journals already publish a subset of this information. In order to enhance our understanding of the review process, journals could also publish short summaries of how a manuscript changed from original submission to publication.

These ideas are general enough that journals and institutions across many subfields of economics could adopt them. We strongly encourage them to proceed with reform initiatives and deliberate experimentation as they see fit. However, we also acknowledge that stakeholders might find many of the proposals in this article to be infeasible or undesirable. We hope, in any case, that economists see our efforts as an invitation to share their perspectives and as a call to action for efforts to improve our peer review system.

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.

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.

Armstrong, J. Scott, "Peer review for journals: Evidence on quality control, fairness, and innovation," *Science and Engineering Ethics*, 1997, *3* (1), 63–84.

Arpinon, Thibaut and Romain Espinosa, "A practical guide to Registered Reports for economists," *Journal of the Economic Science Association*, 2023, pp. 1–33.

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, "How to Write an Effective Referee Report and Improve the Scientific Review Process," *Journal of Economic Perspectives*, 2017, *31*, 231–244.

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 Aluaș, 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 Iriberri, "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.

Charness, Gary, Anna Dreber, Daniel Evans, Adam Gill, and Séverine Toussaert, "Improving Peer Review in Economics: Stocktaking and Proposals," Technical Report 04 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.

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.

Conley, John P., Mario J. Crucini, Robert A. Driskill, and Ali Sina Önder, "The Effects Of Publicataion Lags On Life-Cycle Research Productivity In Economics," *Economic Inquiry*, 2013, *51* (2), 1251–1276.

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.

da Silva, Jaime Teixeira and Aceil Alkhatib, "Should Authors be Requested to Suggest Peer Reviewers?," *Science and Engineering Ethics*, 2018, *24*, 275–285.

Davis, Phil, "PNAS: Tighter Editorial Policy Improves NAS Papers," *Scholarly Kitchen*, January 2016.

Delios, Andrew, Elena Giulia Clemente, Tao Wu, Hongbin Tan, Yong Wang, Michael Gordon, Domenico Viganola, Zhaowei Chen, Anna Dreber, Magnus Johannesson et al., "Examining the generalizability of research findings from archival data," *Proceedings of the National Academy of Sciences*, 2022, *119* (30), e2120377119.

DellaVigna, Stefano and Devin Pope, "Predicting Experimental Results: Who Knows What?," *Journal of Political Economy*, 2018, *126* (6), 2410–2456.

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.

Dufwenberg, Martin and Peter Martinsson, "Sealed Envelope Submissions Foster Research Integrity," *Revue économique*, 2019, *70* (6), 919–926.

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*, 2011, *49* (3), 635–657.

Ersoy, Fulya and Jennifer Pate, "Invisible Hurdles: Gender and Institutional Bias in the Publication Process in Economics," Working Paper, 2021.

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.

Foster, Andrew, Dean Karlan, Edward Miguel, and Aleksandar Bogdanoski, "Pre-results Review at the Journal of Development Economics: Lessons learned so far," *World Bank Blogs*, 2019.

Frey, Bruno S, "Publishing as prostitution?–Choosing between one's own ideas and academic success," *Public choice*, 2003, *116* (1-2), 205–223.

Goldberg, Pinelopi Koujianou, "Report of the Editor: American Economic Review," *American Economic Review*, 2012, *102* (3), 653–65.

Gonçalves, Duarte, Jonathan Libgober, and Jack Willis, "Learning versus Unlearning: An Experiment on Retractions," Working Paper 29512, National Bureau of Economic Research November 2021.

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.

Gordon, Michael, Michael Bishop, Yiling Chen, Anna Dreber, Brandon Goldfedder, Felix Holzmeister, Magnus Johannesson, Yang Liu, Louisa Tran, Charles Twardy et al., "Forecasting the publication and citation outcomes of COVID-19 preprints," *Royal Society open science*, 2022, *9* (9), 220440.

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?," *Journal of Economic Literature*, 2022, (Forthcoming).

Hamermesh, Daniel S., "Facts and Myths about Refereeing," *Journal of Economic Perspectives*, 1994, *8* (1), 153–163.

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.

Huber, Juergen, Sabiou Inoua, Rudolf Kerschbamer, Christian König-Kersting, Stefan Palan, and Vernon L. Smith, "Nobel and novice: Author prominence affects peer review," Working Paper Series, Social and Economic Sciences 2022-01, Faculty of Social and Economic Sciences, Karl-Franzens-University Graz August 2022.

Huisman, Janine and Jeroen Smits, "Duration and quality of the peer review process: the author's perspective," *Scientometrics*, 2017, *113*, 633–650.

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.

Kolev, Julian, Yuly Fuentes-Medel, and Fiona Murray, "Is Blinded Review Enough? How Gendered Outcomes Arise Even Under Anonymous Evaluation," Working Paper 25759, National Bureau of Economic Research April 2019.

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.

_ **and Michael J. Piette**, "Does the "Blindness" of Peer Review Influence Manuscript Selection Efficiency?," *Southern Economic Journal*, 1994, *60*, 896.

_ **and** _ , "Favoritism versus Search for Good Papers: Empirical Evidence Regarding the Behavior of Journal Editors," *Journal of Political Economy*, 1994, *102* (1), 194–203.

List, John A, "Editor's Introduction to JPE Micro," *Journal of Political Economy Microeconomics*, 2023, *1* (1), 1–6.

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.

McAfee, R, "Edifying Editing," *American Economist*, 03 2010, *55*.

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.

Nature, "Three-year trial shows support for recognizing peer reviewers," *Nature*, 2019, *568*, 275.

_ , "Nature is trialling transparent peer review — the early results are encouraging," *Nature*, 2022, *603*, 8.

Nature Communications, "Transparent peer review at Nature Communications," *Nature Communications*, 2015, *6* (1), 10277.

_ , "Transparent peer review one year on," *Nature Communications*, 2016, *7* (1), 13626.

_ , "Transparent peer review for all," *Nature Communications*, 2022, *13* (1), 6173.

Noorden, Richard Van, "The researchers using AI to analyse peer review.," *Nature*, 2022.

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.

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

Otis, Nicholas, "Policy Choice and the Wisdom of Crowds," *Available at SSRN 4200841*, 2022.

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

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

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.

Rigby, John, Deborah Cox, and Kigongo Julian, "Journal peer review: a bar or bridge? An analysis of a paper's revision history and turnaround time, and the effect on citation," *Scientometrics*, 01 2018, *114*.

Scheel, Anne M, Mitchell RMJ Schijen, and Daniël Lakens, "An excess of positive results: Comparing the standard Psychology literature with Registered Reports," *Advances in Methods and Practices in Psychological Science*, 2021, *4* (2), 25152459211007467.

Schulz, Robert, Adrian Barnett, René Bernard, Nicholas JL Brown, Jennifer A Byrne, Peter Eckmann, Małgorzata A Gazda, Halil Kilicoglu, Eric M Prager, Maia Salholz-Hillel et al., "Is the future of peer review automated?," *BMC Research Notes*, 2022, *15* (1), 1–5.

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

Siemroth, Christoph, "Economics Peer-Review: Problems, Recent Developments, and Reform Proposals," *Recent Developments, and Reform Proposals (March 28, 2023)*, 2023.

Silbiger, Nyssa J and Amber D Stubler, "Unprofessional peer reviews disproportionately harm underrepresented groups in STEM," *PeerJ*, 2019, *7*, e8247.

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

Squazzoni, Flaminio, Petra Ahrweiler, Tiago Barros, Federico Bianchi, Aliaksandr Birukou, Jan Blom, Giangiacomo Bravo, Stephen Cowley, Virginia Dignum, Pierpaolo Dondio, Francisco Grimaldo, Lynsey Haire, Jason Hoyt, Phil Hurst, Rachael Lammey, Catriona MacCallum, Ana Marusic, Bahar Mehmani, Hollydawn Murray, and Michael Willis, "Unlock ways to share data on peer review," *Nature*, 02 2020, *578*, 512–514.

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.

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.

Wolfram, Dietmar, Peiling Wang, Adam Hembree, and Hyoungjoo Park, "Open peer review: promoting transparency in open science," *Scientometrics*, 05 2020, *125*, 1–19.

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

Xu, Chiwei, Jun Xu, Hong-Wen Tang, Maria Ericsson, Jui-Hsia Weng, Jonathan DiRusso, Yanhui Hu, Wenzhe Ma, John M Asara, and Norbert Perrimon, "A phosphate-sensing organelle regulates phosphate and tissue homeostasis," *Nature*, 2023, pp. 1–9.

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

A Data discussions

This appendix is devoted to explaining certain aspects of our data collection in greater detail and presenting additional figures of importance.

A.1 Survey Data

Table A.1: Descriptive Statistics of Survey Population

		Andre & Falk (2021)		
	Peer review survey	Study population	Unweighted sample	Weighted sample
Demographics				
Female	23.5%	26.0%	23.1%	25.8%
Age:				
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%
Location:				
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) (AF) or derived for us by Peter Andre.

[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.

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:** From the Andre and Falk (2021) (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:** 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:** 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:** The number of publications is capped at 200.

A.2 Construction of the interdisciplinary comparison table

We made use of a wide range of data sources in the construction of Table 1. Furthermore, we made several subjective judgments about the best way to code the available information and data. As such, in this section we provide more information about sample selection, data sourcing, and variable definitions:

- Our first judgment relates to which journals to include in the sample for each discipline. Given their importance to the careers of economists (Heckman and Moktan, 2020), we include each of the traditional "top 5" journals in our economics sample. For the others, we chose the top 3 journals in each discipline, according to their Scimago Journal Rank h-index. Disciplines are sometimes paired together in these rankings (e.g. Sociology and Political Science) and interdisciplinary journals are often included as well (e.g. *Trends in Cognitive Sciences* within Psychology). In these cases, we made a judgment call about which journals can be classified as belonging to the relevant discipline. The journals sampled for each discipline are as follows:
    - **Economics.** We include the *American Economic Review* (*AER*), *Econometrica*, the *Journal of Political Economy* (*JPE*), the *Quarterly Journal of Economics* (*QJE*), and *The Review of Economic Studies* (*REStud*).
    - **General Science.** We include *Nature*, the *Proceedings of the National Academy of Sciences* (*PNAS*), and *Science*.
    - **Medicine.** We include *The Journal of the American Medical Association* (*JAMA*), *The Lancet*, and *The New England Journal of Medicine* (*NEJM*).
    - **Political Science.** We include the *American Journal of Political Science* (*AJPS*), the *American Political Science Review* (*APSR*), and *The Journal of Politics* (*JOP*).
    - **Psychology.** We include the *Journal of Applied Psychology* (*JAP*), the *Journal of Personality and Social Psychology* (*JPSP*), and *Psychological Bulletin* (*PB*).
- The statistics in the **Reach** section of the table were sourced entirely from the Scimago Journal Rank as of early May 2023. Journals differ greatly in their reporting practices, so the common source helps to ensure the comparability of metrics between journals. The *Articles Published (2019-2021)* variable corresponds to the *Citable Docs. (3 years)* column on Scimago. *Impact Factor (two-year)* comes from the *Cites/Doc (2 years)* column. And *h-index* comes from the column of the same name. Each of these statistics is averaged across journals within each sample in order to calculate a discipline-specific metric.
- The statistics in the **Administration and Policies** section come from a wide range of sources.
    - The *Professional Editorial Board* variable records whether manuscript decisions are (primarily) made by part-time editors (= 0), who are largely academics hosted at universities or medical professionals, or by editors who are employed full-time by the journal (= 1). For each journal, this information was obtained by inspection of editorial board pages. Across disciplines, a great majority of journal editors have part-time editorial roles (AER-1, AJPS-1, APSR-1, ECMA-1, JAMA-1, JAP-1, JOP-1, JPE-1, JPSP-1, NEJM-1, PB-1, PNAS-1, QJE-1, RES-1). There are three exceptions. The editorial boards of *Nature* and *The Lancet* are composed of full-time editors (LAN-1, NAT-1). The *Science* editorial board employs a substantial mix of part-time and full-time editors; we code it as having a professional editorial board since the deputy and senior editors are full-time staff (SCI-1). In the table, this statistic is reported as the number of journals in each sample that have a professional editorial board (e.g. 1 of 3).
    - The *Restriction on Manuscript Word Count* variable reports the journal's stated or suggested restriction on the maximum length of manuscripts that it will consider and/or publish. The restriction is sometimes given as a page count (with 1-1.5 line spacing) rather than a word count. Page counts are converted into word counts at a rate of 525 words per page.[17] Each restriction is

[17] This is an approximate midpoint between two estimates of the number of words per page in journal articles. *Econometrica* states that an article should have no more 45 pages and 17,000 words (∼378 words per page), while *PNAS* estimates 4,000 for 6 pages (∼667 words per page). The number of words per page will obviously vary based on formatting requirements, but the ranking of disciplines by word count does not change for any reasonable estimates of words per page e.g. 300 to 800 words per page.

taken from the journal page providing submission instructions or formatting guidelines to authors (AER-2, AJPS-2, APSR-2, ECMA-2, JAMA-2-a, JOP-2-a, LAN-2, NAT-2, NEJM-2, PNAS-2, RES-2, SCI-2). *PNAS* states that a typical article is 6 pages and 4,000 words, while a 12 page maximum is enforced; we therefore use 8,000 words as the approximate word count restriction. Five journals do not give any explicit guidelines around length (i.e. the *JPE* and *QJE* in Economics and all of the Psychology journals). In the table, this statistic is averaged across journals within each sample.

- The *Authors Can Suggest/Oppose Referees* variable is an indicator for whether a journal formally allows authors to suggest or oppose specific referees when submitting their manuscript (= 1) or not (= 0). Suggesting and opposing referees are pooled together because journals in the sample either allowed both or neither. In most cases, this was verified by clicking through the journal's manuscript submission portal. In some cases, this possibility is explicitly mentioned by the journal in their submission guidelines or on another page (AJPS-3, APSR-3, JOP-2-b, NAT-3, NEJM-3, SCI-2). In the table, this statistic is reported as the number of journals in each sample that allow authors to do this.
- The *Referee Report Guidelines Available* variable indicates whether a journal makes referee report-writing guidelines publicly available (= 1) or not (= 0). These guidelines were found on journal websites (AER-3, AJPS-4, APSR-4, ECMA-3, NAT-4, PNAS-3-a, SCI-3). In the table, this statistic is reported as the number of journals in each sample that make guidelines available.
- The *Referees Paid for Reports* variable indicates whether the journal offers cash compensation to reviewers for timely reports (= 1) or not (= 0). Within our sample. this practice is exclusive to the *AER* and the *JPE* (AER-3). We confirmed with editors that the *JPE* offers referees payment. In the table, this statistic is reported as the number of journals in each sample that do this.
- The *Desk Rejection Rate* variable indicates the average desk rejection rate among journals in the sample, if they report one. Many journals make this information available on their websites or in periodic reports (AER-4-a, AJPS-5, APSR-5-a, ECMA-4-a, JOP-3, QJE-2, SCI-4-a). *Econometrica*'s rate was not made directly available but was recalculated from submission and decision volumes in their report. We obtained rates directly from journal editors for *JPE* and *REStud*. The most recent desk rejection rate available from PNAS is from 2014 (54%, PNAS-4). We prefer using newer data when available, but this rate appears not to have changed much in recent years, as *PNAS* currently states that they desk reject "more than 50% of submissions" (PNAS-5). In the table, this statistic is averaged across journals within each sample.

- The statistics in the **Journal Metrics** section come from a smaller range of sources. Many journals do not make this information publicly available.
  - The *Average Days to First Post-review Decision* variable records the mean number of days from submission to when a journal communicates a first decision, for papers that are not desk rejected. For the economics journals, we obtained this information from a uniform source to ensure comparability. *PNAS* is the only other journal in the sample that reports this exact statistic (PNAS-6). In the table, this statistic is averaged across journals within each sample.
  - The *Median Days to First Post-review Decision* variable records the median number of days between manuscript submission to when a journal communicates a first decision, for papers that are not desk rejected. We again obtained this information from the same uniform source for the economics journals. Other journals provide data on this statistic through their reports and other pages on their website (NAT-5-a, SCI-4-b). The *APSR* does not directly provide this statistic; rather, they provide the median number of days to rejection after a full review (70) and the median number of days to an R&R decision (87, APSR-5-b). Using these statistics and the number of submissions that receive each decision (1,542 and 434, respectively), we use a weighted average of these medians as the overall median (74). In the table, this statistic is averaged across journals within each sample.
  - The *Median Days to First Post-review Decision* variable records the median number of days between manuscript submission to when a journal communicates a first decision, for papers that

are not desk rejected. We again obtained this information from the same uniform source for the economics journals. Other journals provide data on this statistic through their reports and other pages on their website (NAT-5, SCI-4-b). The *APSR* does not directly provide this statistic; rather, they provide the median number of days to rejection after a full review (70) and the median number of days to an R&R decision (87, APSR-5-b). Using these statistics and the number of submissions that receive each decision (1,542 and 434, respectively), we use a weighted average of these medians as the overall median (74). In the table, this statistic is averaged across journals within each sample.

- The *Average Days from Submission to Acceptance* variable records the average number of days between manuscript submission to when a journal communicates a final acceptance decision, for papers that are eventually accepted. For the economics journals, this came from editors' reports (AER-4-b, ECMA-4-b). In the table, this statistic is averaged across journals within the economics sample. The other statistics come from Hadavand et al. (2022); they present statistics that pool the *APSR*, *JAP*, and *JPSP* together. In 2020, these journals took an average of 14.37 weeks from submission to acceptance, or approximately 437 days. This is why an average of 437 days is reported for both disciplines.
- The *Median Days from Submission to Acceptance* variable records the median number of days between manuscript submission and the communication of a final acceptance decision. For the general science journals, this came from their journal metrics pages (NAT-5-b, SCI-4-c). In the table, this statistic is averaged across journals within the sample.

- The statistics in the **Confidentiality and Transparency** section come from reading journal editorial and review policies.
  - The *Double-Blind Refereeing* variable indicates whether manuscripts sent to referees are anonymized in a double-blind fashion (= 1) or are single-blind (= 0). The top 5 journals are uniform in their application of single-blind refereeing, which we (the authors) can confirm through personal experience. Most journals mention their policy in their guidelines or information pages (AJPS-6, APSR-6, JAMA-2-b, JAP-2, JOP-2-c, JPSP-2, LAN-3, PB-2). *Nature* gives authors the option of submitting their manuscripts double-blind (NAT-4); we code this as = 1 for double-blind refereeing. *NEJM* and *Science* do not explicitly state their policies, but we can infer that they use single-blind reviewing from their requests that referees consider any potential conflicts they have with the authors (NEJM-4, SCI-5). Finally, we found a statement of *PNAS*' single-blind policy in Davis (2016). In the table, this statistic is reported as the number of journals in each sample that offer double-blind refereeing.
  - The *Public Referee Identity* variable indicates whether the journal sometimes publishes the names of the referee(s), subject to consent. *Nature* allows this; see Xu et al. (2023) for a recent example. This is also possible at *PNAS*, specifically when NAS members make contributions (PNAS-3-b). In the table, this statistic is reported as the number of journals that sometimes publish referee identities.
  - The *Public Referee Reports* variable indicates whether the journal sometimes publishes the content of referee reports, subject to consent. *Nature* is the only journal in the sample that does this, which is visible in the same example as above (Xu et al., 2023). In the table, this statistic is reported as the number of journals that sometimes publish the contents of referee.

A.3 Journal sourcing table

We present Table A.2 in this section as an exhaustive list of sources about journal statistics and policies that we cited throughout the article. We elected not to formally 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.2: Sources for journal statistics and editorial policies

Journal Name	Source ID(s)	Source Name	Source Year	Source Link
AER	AER-1	Editors of the AER	2023	AER-1
	AER-2	AER: Submission Guidelines	2023	AER-2
	AER-3	AER: Information for Reviewers	2023	AER-3
	AER-4-a, AER-4-b	Report of the Editor	2022	AER-4
AJPS	AJPS-1	Editorial board	2023	AJPS-1
	AJPS-2	Manuscript Preparation	2023	AJPS-2
	AJPS-3	Submit Manuscript	2023	AJPS-3
	AJPS-4	*AJPS* Reviewer Instructions	2023	AJPS-4
	AJPS-5	Annual Report	2022	AJPS-5
	AJPS-6	Guidelines for Manuscripts	2023	AJPS-6
APSR	APSR-1	Editorial board	2023	APSR-1
	APSR-2	Preparing your materials	2023	APSR-2
	APSR-3	Submitting your materials	2023	APSR-3
	APSR-4	Instructions for peer reviewers	2023	APSR-4
	APSR-5-a, APSR-5-b	*APSR* Editorial Report: Executive Summary	2022	APSR-5
	APSR-6	Notes from the Editors	2019	APSR-6
Econometrica	ECMA-1	Editorial Board	2023	ECMA-1
	ECMA-2	Instructions for Submitting Articles	2023	ECMA-2
	ECMA-3	Guidelines for Referees	2023	ECMA-3
	ECMA-4-a, ECMA-4-b	Report of the Editors	2023	ECMA-4
Economic Inquiry	EI-1	Journal Policies	2023	EI-1
F1000Research	F1000-1-a, F1000-1-b	Reviewer Guidelines	2023	F1000-1
JAMA	JAMA-1	Editorial Staff	2023	JAMA-1
	JAMA-2-a, JAMA-2-b	Instructions for Authors	2023	JAMA-2
JAP	JAP-1	Editorial Staff	2023	JAP-1
	JAP-2	Submission Guidelines	2023	JAP-2
JDE	JDE-1-a, JDE-1-b	Pre Results Review Papers	2023	JDE-1
JESA	JESA-1	Submission guidelines	2023	JESA-1
JOP	JOP-1	Editorial Board	2023	JOP-1
	JOP-2-a, JOP-2-b, JOP-2-c	Instructions to Authors	2023	JOP-2
	JOP-3	Interim Report to the Editorial Board	2023	JOP-3
JPE	JPE-1	Editorial Board	2023	JPE-1
	JPE-2	ETHICS	2023	JPE-2
JPSP	JPSP-1	Editorial Board	2023	JPSP-1
	JPSP-2	Submission Guidelines	2023	JPSP-2
JUE	JUE-1	Call for Shorter Papers on COVID-19	2023	JUE-1
The Lancet	LAN-1	People at The Lancet	2023	LAN-1
	LAN-2	Information for Authors	2023	LAN-2
	LAN-3	Peer review	2023	LAN-3
Nature	NAT-1	About the Editors	2023	NAT-1
	NAT-2	Formatting guide	2023	NAT-2
	NAT-3	Editorial criteria and processes	2023	NAT-3
	NAT-4	Peer Review	2023	NAT-4
	NAT-5-a, NAT-5-b	Journal Metrics	2023	NAT-5
NEJM	NEJM-1	Editors and Publishers	2023	NEJM-1
	NEJM-2	Article Types	2023	NEJM-2
	NEJM-3	Frequently Asked Questions	2023	NEJM-3
	NEJM-4	Publication Process	2023	NEJM-4

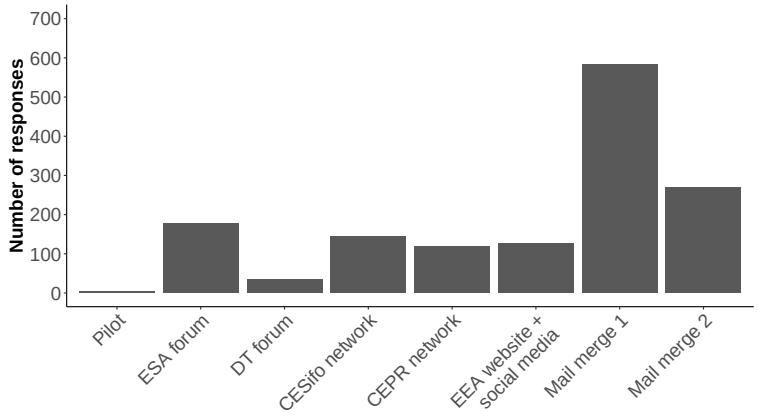
Continued on next page

Nature Human Behaviour	NHB-1-a, NHB-1-b	Peer Review	2023	NHB-1
Oxford Open Economics	OOE-1	Registered Reports Guidelines	2023	OOE-1
PNAS	PNAS-1	Editorial Board	2023	PNAS-1
	PNAS-2	Submitting Your Manuscript	2023	PNAS-2
	PNAS-3-a, PNAS-3-b	For Reviewers	2023	PNAS-3
	PNAS-4	Simplifying the Direct Submission process	2014	PNAS-4
	PNAS-5	Editorial and Journal Policies	2023	PNAS-5
	PNAS-6	Article and Journal Metrics	2023	PNAS-6
PB	PB-1	Editorial Board	2023	PB-1
	PB-2	Submission Guidelines	2023	PB-2
QJE	QJE-1	Editorial Board	2023	QJE-1
	QJE-2	Tweet	2023	QJE-2
Science	SCI-1	Editorial Board	2023	SCI-1
	SCI-2	Information for authors	2023	SCI-2
	SCI-3	Guidelines for Reviewers	2023	SCI-3
	SCI-4-a, SCI-4-b, SCI-4-c	Journal metrics	2023	SCI-4
	SCI-5	Peer Review at Science Journals	2023	SCI-5
REStud	RES-1	Editorial Board	2023	RES-1
	RES-2	Submission Guidelines and Policy	2023	RES-2

Journal acronyms: *AER = American Economic Review*; *AJPS = American Journal of Political Science*; *APSR = American Political Science Review*; *JAMA = The Journal of the American Medical Association*; *JAP = Journal of Applied Psychology*; *JDE = Journal of Development Economics*; *JESA = Journal of the Economic Science Association*; *JOP = Journal of Politics*; *JPSP = Journal of Personality and Social Psychology*; *JPE = Journal of Political Economy*; *NEJM = The New England Journal of Medicine*; *NHB = Nature Human Behaviour*; *PNAS = Proceedings of the National Academy of Sciences*; *PB = Psychological Bulletin*; *QJE = Quarterly Journal of Economics*; *REStud = Review of Economic Studies*.

B Additional Figures

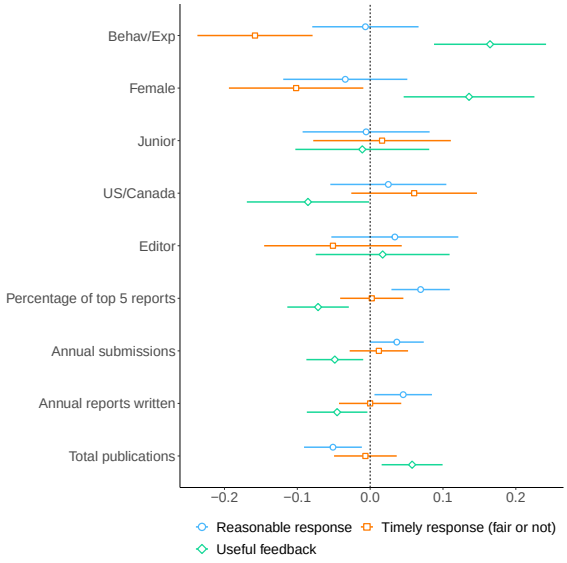
Figure B.1: Distribution of responses across recruitment channels



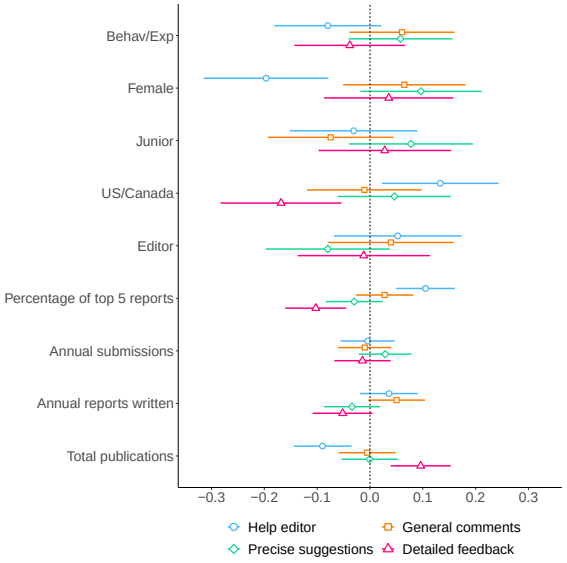
Notes: This figure only includes respondents who fully completed the survey. $N = 1,459$.

Figure B.2: Regressions of rank assigned to stated peer review outcomes on respondent characteristics

(a) Expected outcomes

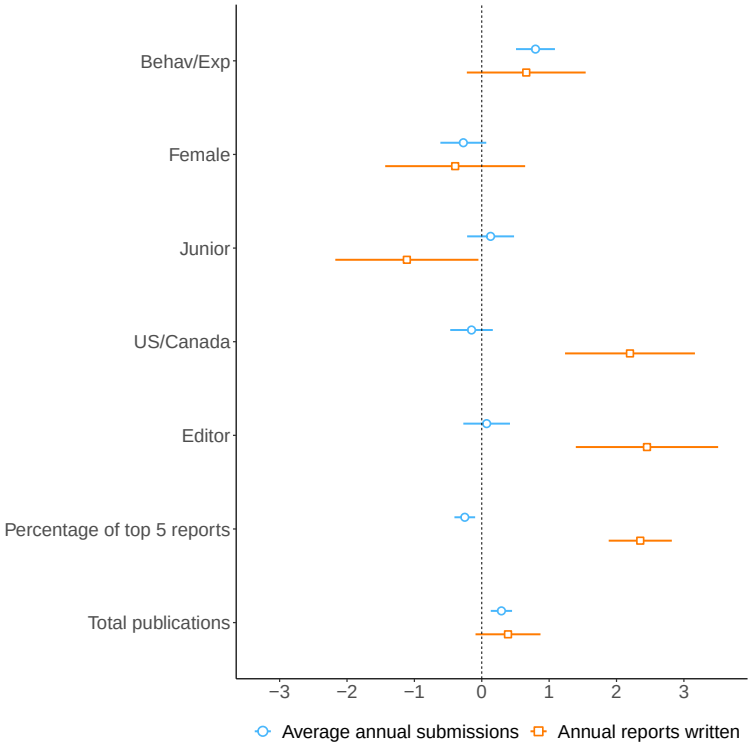


(b) Objectives of referee reports



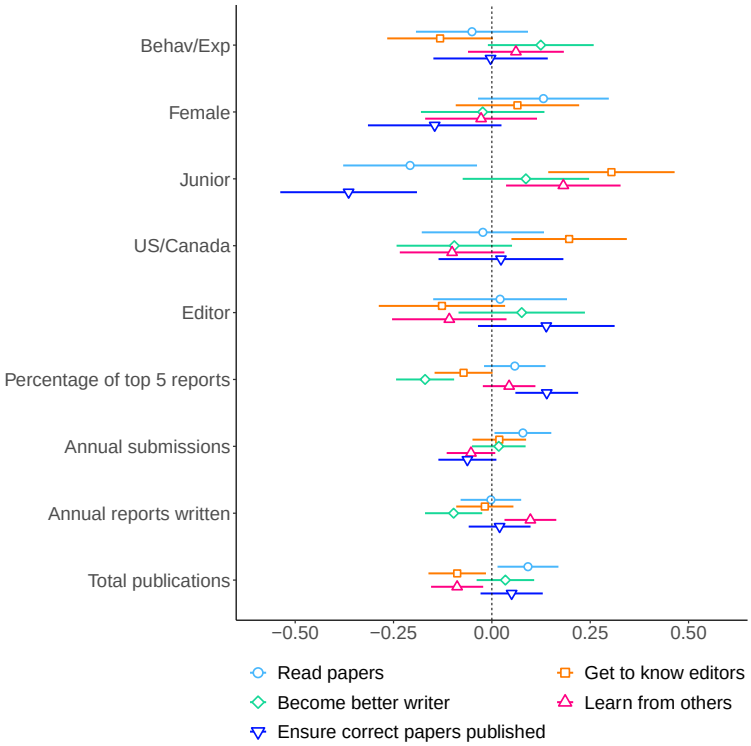
Notes: For both panels and each outcome, the negative numeric value of the rank assigned by respondents is regressed, so that positive coefficients indicate a better rank on average. N = 1,459 for both panels.

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



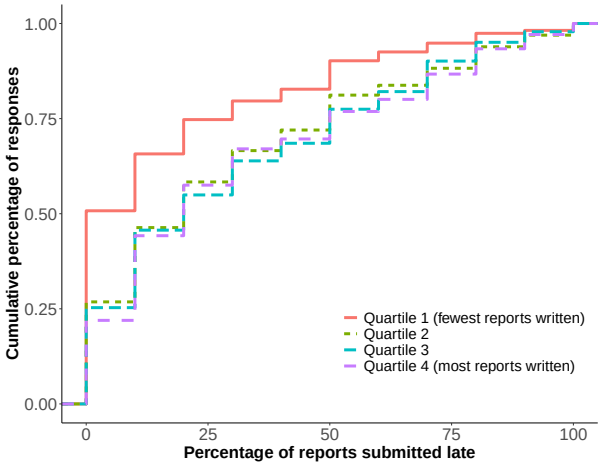
Notes: For both submissions made and reports written, the annual amount is regressed. $N = 1,459$.

Figure B.4: Regression of rank assigned to stated benefits of refereeing on respondent characteristics



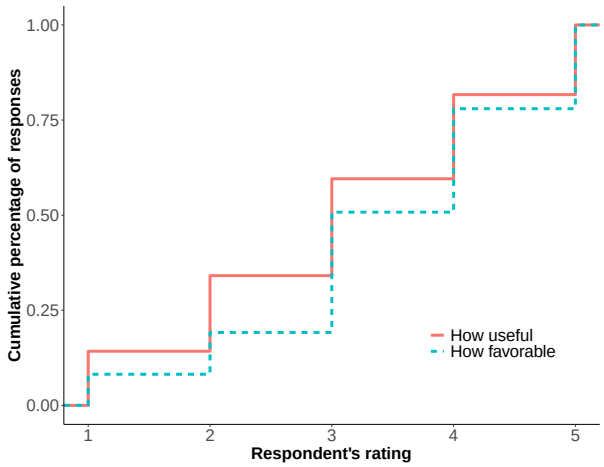
Notes: For each potential benefit, the negative numeric value of the rank assigned by respondents is regressed, so that positive coefficients indicate a better rank on average. N = 1,459.

Figure B.5: Percentage of late reviews as a function of the number of reports written



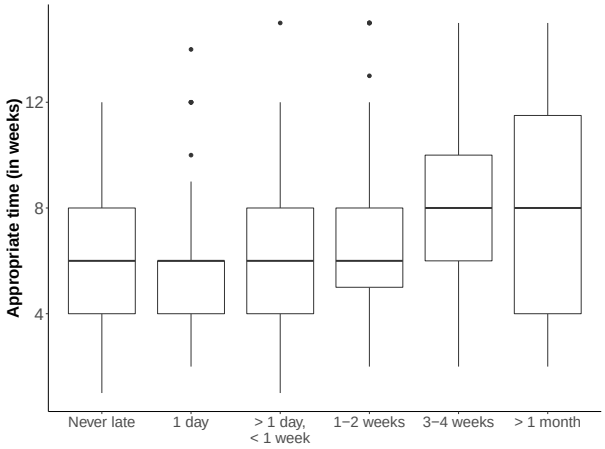
Notes: $N = 1,483$.

Figure B.6: Overall support and perceived usefulness for a centralized platform



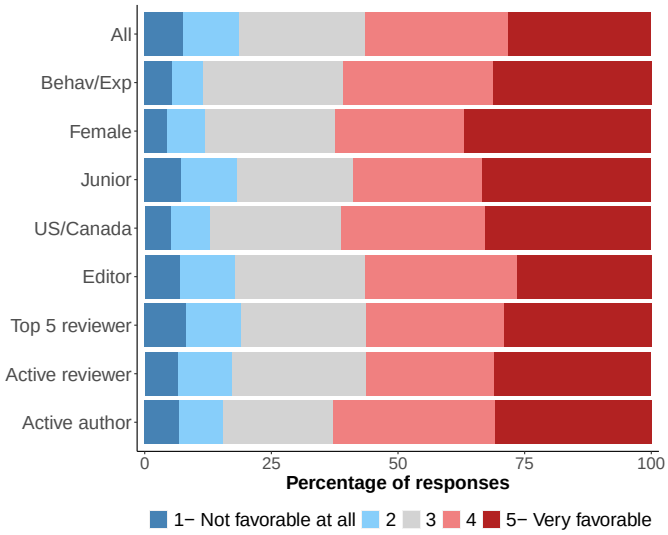
Notes: $N = 1,467$.

Figure B.7: Appropriate deadline by own delays



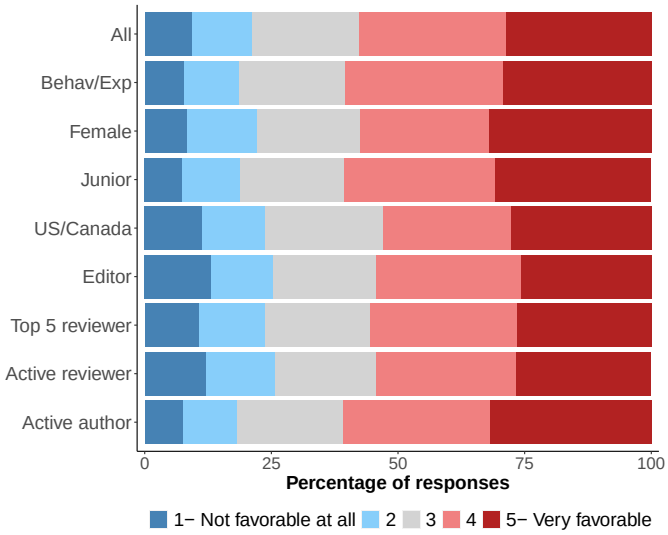
Notes: $N = 1,455$.

Figure B.8: Support for allowing authors to disqualify certain reviewers



Notes: $N = 684$.

Figure B.9: Attitudes towards a formal appeal procedure



Notes: $N = 1,459$.

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. 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 B.1).[18] 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).

After completing our primary analysis of the survey data, we sent out a follow-up survey in February 2022 ($N = 117$) to clarify our interpretation of the initial results and gather further evidence. Our additional inquiries included questions on what respondents consider to be reasonable report-writing activity, whether they feel pressured to write more referee reports due to publication concerns around their own manuscripts,

[18] A 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.

and the percentage of their submissions from 2020-2021 that were desk rejected (27.8%). These results were included in our original report, but most are excluded from the present article. The follow-up survey sampled relatively fewer full (33.6%) and assistant (19.0%) professors, and relatively more postdoc/PhD candidates (19.0%). 33.3% of respondents reported that they are currently editors.

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."

Table C.2: Characteristics across recruitment channels[a]

	ESA forum	CESifo	CEPR	EEA / social media	Mail merge 1	Mail merge 2
Demographics						
Female	27.5%	17.9%	28.4%	28.5%	23.8%	25.9%
Age:						
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%
Location:						
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.

D Survey questions

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]

[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]

[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]

[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

[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

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.

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

[Q8]: Do you think journals or associations should provide a template for referee reports? [YES/NO]

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]

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

[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 Communcations. 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]

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]

Recognition

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

[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]

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]

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

[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]

[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]

[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]

[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]

[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]

[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]

[Q24]: Have you occupied or are you currently occupying an editorial position? [YES/NO]

[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]

[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]

[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]

[Q28]: What do you think is a reasonable number of reports to be assigned per year? [Dropdown with numbers]

[Q29]: Did you reject a request to referee over these past two years? [YES/NO]

[Q30]: How many times did you reject a request to referee? [Dropdown with numbers]

[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]

[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]

[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.

[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.

[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.

[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]

[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]

[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]

[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]

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