

Economics Peer-Review: Problems, Recent Developments, and Reform Proposals*

Christoph Siemroth[†]

First version: March 28, 2023

This version: April 24, 2023

Abstract

This article contributes to the debate in the economics profession on reforming the peer-review process. It examines the current state of peer-review in economics, surveys the relevant literature, and identifies several problems and solutions. Problems to be discussed are referee overreach and excessive revisions, strategic refereeing and conflicts of interest, prestige bias and other discrimination, and the noisy outcome of peer-review. It recommends several solutions for reform. First, enforce referee guidelines that reports must explicitly separate their suggestions into essential and optional, with 3 essential maximum. Second, let authors award the best referee report. Third, adopt conflict of interest policies for referees and punish non-disclosure. Fourth, use double-blind refereeing. Fifth, make better use of prior reports from other journals. Sixth, pay referees for prompt reports. A discussion of the role of editors highlights additional issues that deserve a debate in the profession.

Keywords: Conflicts of Interest, Excessive Revisions, Peer-Review, Prestige Bias, Publication Process, Referee Overreach, Reform, Survey

JEL Classification: A11, A12, A14

*I thank Roy Bailey, Alex Clymo, Dan Friedman, Tim Hatton, Friederike Mengel, Lisa Spantig for helpful comments (these names are not necessarily the ones quoted in the anonymous anecdotes in the article). This work was supported by the Economic and Social Research Council [grant number ES/T015357/1].

[†]University of Essex, Department of Economics, Wivenhoe Park, Colchester, CO4 3SQ, UK. E-mail: christoph.siemroth@essex.ac.uk

Contents

1	Introduction	3
2	Current problems of economics peer-review	5
2.1	Referee overreach and excessive revisions	5
2.1.1	Problem description	5
2.1.2	Potential solutions	7
2.2	Strategic refereeing and conflicts of interest	9
2.2.1	Problem description	9
2.2.2	Potential solutions	10
2.3	Prestige bias and other discrimination	11
2.3.1	Problem description	11
2.3.2	Potential solutions	12
2.4	Noisy outcomes	13
2.4.1	Problem description	13
2.4.2	Potential solutions	14
2.5	The role of editors	15
3	Recommendations for reform	17
4	Concluding remarks	18
	Bibliography	19

1 Introduction

Everybody seems to agree that peer-review in economics is not working well (e.g., Frey, 2003; McAfee, 2014; Rubinstein, 2017; Akerlof, 2020). Most referees and editors do a great job and are underpaid for what they do, nevertheless the process often leads to unsatisfactory outcomes. Yet aside from trying to shorten review times, relatively little has been done about it (Rubinstein, 2017). Among the problems to be discussed here are referees increasingly asking for more and overreaching changes, thus shifting the original focus of the manuscript, making manuscripts longer, reducing readability and increasing publication times. Further problems include how the peer-review system is biased against certain authors, that conflicts of interest affect referee recommendations and ultimately publication decisions, and that the peer-review outcome is noisy and unpredictable. Azar (2006) argued very eloquently why progress on such issues of the peer-review process is hard to achieve:

In academia, there are incentives for individuals to increase their efficiency, but [...] no one has both the power and the incentives to improve the system. While society may benefit from increased productivity in academia, there are no managers with the authority and incentives to make the production process of academic knowledge optimal. It is therefore important that researchers will dedicate time to think seriously about the issues involved in this process.

Indeed, even a well-known economist and former *AER* and *Economic Inquiry* editor, who was very aware of many of these issues (McAfee, 2014) and tried a radical remedy at *Economic Inquiry*, has left university employment in favor of a corporate gig. A single journal or editor cannot fix these issues. Without a coordinated move for change in the profession, these peer-review issues will persist or even worsen, as did the time-to-acceptance in previous decades (Ellison, 2002b).

This paper attempts to make a contribution by formulating the issues, organizing the literature around them, and by discussing possible solutions. The hope is that it helps to bring about positive change in the economics publication process, hence I focus on providing a concise survey and discussion rather than an exhausting discussion of all possible issues. To that end, I synthesize a short list of realistic recommendations for reform, based on the most promising solutions.

First, I recommend that journals enforce standards that referee reports must explicitly separate their suggestions into essential and optional, and that the number of essential changes asked for is capped at 3. This limits the scope of revisions, saves everyone (authors, editors, referees) time, and likely leads to less bloated and more readable papers by focusing on the main issues when revising. Second, I recommend that authors can select one report per first submission for an award. The hope is that these awards make refereeing more attractive and reports more useful. Third, I recommend that conflict of interest policies *for referees* are adopted by all journals, that these include “competing papers”, that failure

to disclose conflicts is punished, and that editors refrain from using referees with conflicts. Fourth, I recommend double-blind peer-review, which is easy to circumvent for referees, but evidence suggests it might still be partially effective in reducing biases in peer-review and the implementation costs are small. Fifth, I recommend that journals encourage the sharing of prior reports from submissions at other journals by authors. If a manuscript was narrowly rejected based on a lack of novelty at a top journal, then it would be a waste to redo the review at a lower ranked journal, and editors should be open to using prior reports for a decision. The good news is that some of these measures are already in use at selected journals. Sharing of prior referee reports is possible within some journal families (e.g., the American Economic Association, AEA), though typically not beyond. And *Econometrica* is now using ambitious new referee guidelines aimed at avoiding excessive revisions. So improvement is not impossible, but still rare.

While slow review and publication times are an issue (e.g., [Ellison, 2002b](#)), I will not discuss it at length, because out of all the problems with peer-review, this one received quite a bit of attention from journals already.¹ Since other problems discussed here, such as referee overreach or low referee agreement, are partial causes of the slow publication process, solving these issues will also speed up the process. In addition, sixth, I recommend that referees are paid for prompt (within 4 weeks) reviews at all journals that charge authors submission fees. Payment as well as short deadlines have been shown to reduce review times in a convincing field experiment ([Chetty et al., 2014](#)), with some evidence that review quality does not suffer.²

I will also not discuss the problems concerning p-hacking, publication bias, or non-reproducibility of empirical research (e.g., [Brodeur et al., 2016](#); [Camerer et al., 2016](#); [Brodeur et al., 2020](#)). These are very deep issues beyond the referee process. For example, the code and data at most journals do not have to be shared by authors with the submission, so there is not much referees can do to check the data themselves.

There have been previous attempts to discuss and research issues with peer-review, which are surveyed below. A special status belongs to the Evaluating Research project (<https://evalresearch.weebly.com/>), which ran a survey among about 1500 publishing and refereeing academic economists. They summarized the literature on peer-review, and discussed all proposals for reform in the literature and from their survey in [Charness et al. \(2022\)](#), sometimes with survey evidence on these proposals' popularity among economists. Theirs is an excellent report, but given over 150 discussed proposals, it might be too exten-

¹E.g., with the introduction of referee payments at *EER*, *AEJs*, most finance journals, etc. Or with expanded desk rejection rates, which at some journals exceed 60% of submissions (e.g., the *QJE* at 64% in 2020, see [Charness et al., 2022](#), Table 4). Desk rejections are not always beneficial if they are more noisy or biased than the referee process, but they are undoubtedly faster than regular peer-review. [Azar \(2006\)](#) discusses further measures to reduce publication delay.

²[Hamermesh \(1994\)](#), based on correlational evidence, also concludes that cash rewards for timely reports reduce referee times.

sive for reform action. In contrast, after discussion of promising candidates, I propose a few very straightforward changes, which are hopefully easier to use in order to change policies.

2 Current problems of economics peer-review

This section aims to clearly articulate the issues and discuss potential solutions. Section 3 selects the most promising solutions as recommendations.

2.1 Referee overreach and excessive revisions

2.1.1 Problem description

What is the task of the referee? Is it:

1. To see if there are errors or omissions in the manuscript, and judge the contribution to the literature (**evaluation**)? Or:
2. To help improve the manuscript to the best state it can get to within 1-3 revisions (**improvement**)?

Peer-review in economics started with evaluation in mind, but has over time added on the improvement task (Ellison, 2002b). On the one hand, why not have referees suggest improvements while they evaluate the manuscript, so both objectives are compatible. On the other hand, this task extension today leads to many more suggestions for changes and more revisions. Too often, these suggested changes are subjective, marginal, or actually make the manuscript worse, so they add a large time cost for referees, authors, and editors without actually improving the manuscript.³ An additional problem is that inter-referee agreement is typically low (see section 2.4), so the suggested changes can be idiosyncratic or even contradictory. A former editor of the *AER* succinctly summarizes some of these problems (McAfee, 2014 with my emphases):

The way that economists operate journals is perhaps the most inefficient operation that I encounter on a regular basis. It is a fabulous irony that a profession obsessed with efficiency operates its core business in an inefficient manner. [...] Many hours are devoted to reviewing papers. This would be socially efficient if the paper was improved in a way that is commensurate with the time spent, but revising papers using blind referees *often makes papers worse*. Referees offer specific advice that *push papers away from the author's intent*. It is one thing for a referee to say, "I do not find this paper compelling because of X," and another

³Due to these delays, the econ public debate often has to rely on working papers rather than peer-reviewed research, as in the Covid-19 crisis 2020, whereas other disciplines are fast enough to contribute to the debate with peer-reviewed publications.

thing entirely to say that the referee would rather see a different paper on the same general topic and try to get the author to write it. *The latter is all too common.* Gradually, [...] we have transformed the business of refereeing from the evaluation of contributions with a little grammatical help into an elaborate system of glacier-paced *anonymous coauthorship*. [...] My sense is that the first revision of papers generally improves them and that it is downhill from there.

Frey (2003) has similar concerns, arguing every referee effectively has veto power, thus forcing authors to follow each of their “suggestions.” This also makes it harder for unconventional ideas or approaches to succeed. Akerlof (2020) echoes these concerns: “As I understand it, a “review” is a journal that takes submissions, and decides which to accept/which to reject. That means that the editors and the referees should be viewing themselves as helpmates, rather than dictators holding authors at ransom before accepting their submissions.” If these problems are so pervasive, their consequences should be measurable and visible in the data, and indeed they are.

Ellison (2002b) finds that submission-to-acceptance times at major economics journals have increased from 6-12 months in 1970 to 18-30 months in 1990. The mean time to acceptance was 6.1 months in 1970 and increased to 17.3 months by 1999 (Ellison, 2002b, Table 1). A large part of this slowdown in economics publishing is attributed to the increase in rounds of revisions. Before 1970, “revise and resubmit” decisions were rare and exceptional, and manuscripts were either accepted or rejected, and sometimes got an “accept and revise” decision. Today, straight acceptances are almost unheard of, and first round decisions are either rejections or R&Rs. The mean number of revision rounds increased from about 0.6 in 1960 to about 2.0 in 1990 (Ellison, 2002b, p. 956).

Manuscripts get longer, with more extensions, robustness checks, and qualifications. Between 1970 and 2010, the mean paper length in top economics journals increased from about 16 to 46 pages (Card and DellaVigna, 2013). Part of this increase is anticipation of the peer-review process, part of it is its product, with referees asking for additions. As Rubinstein (2017) notes, it can make papers less readable.

One cause of the problem might be that we are stuck in a “bad equilibrium” (Rubinstein, 2017; Hadavand et al., 2021) and expectations in this equilibrium call for referees to provide at least 4 pages of comments and suggestions, because that’s how their reports look. So they fill those pages, even though they may be happy with the manuscript except for perhaps two things. But to please the editor (in their mind), they turn those two issues into ten. Multiply this by 2-3 referees, plus maybe an associate editor, and a revision has to address 20-30 suggestions. Similarly, Berk et al. (2017) argue that “referees feel the need to demonstrate their intelligence or industriousness to editors by identifying problems in papers.” Ellison (2002a) also attributes a large part of the increased revision effort to changing norms within the profession. The explanation of the long term trend is that authors see increasingly demanding referee reports, then become increasingly demanding as referees themselves, and

thus the profession produces ever longer papers and ever longer revisions. Hence, to a large extent the excessive revision problems might be due to evolving expectations, not due to a particular institutional feature or rule. If these outcomes are part of a drift in norms, rather than a conscious choice, then perhaps these norms should be revisited.

2.1.2 Potential solutions

There are numerous potential solutions to this problem, but first and foremost, we as a discipline have to decide whether the job of referees is evaluation or improvement.⁴ If evaluation, we need to drastically change the expectations of and guidelines for referees. Because de facto referees are doing the improvement task, and perhaps excessively so, as the “anonymous coauthorship” characterization of McAfee (2014) suggests.

Nudges and appeals. A former editor of the *Scandinavian Journal of Economics* reminds referees of their more limited role during the referee invitation: “Please give me a quick short review that honestly evaluates the contribution rather than spending a long time to list marginal improvements” (Friberg, 2014). The referee guidelines for the journal *Theoretical Economics* were also amended to stress that the referee job is not to expand the paper: “The primary purpose of the report is to help the Editor make a decision. If a revision is requested, it should then provide sufficient conditions for publication. The referee should show restraint when asking for new material” (TE, 2023). However, referees are often doing unpaid and complicated work, so editors are already happy someone is willing to accept the referee request. Thus, single editors might find it hard to lecture referees on how they are supposed to review – at least as long as the norms profession-wide have not changed. Hence, these small nudges are unlikely to change much if it is only single journals or single editors using them.

No revisions. A radical solution is to have a peer-review process that either accepts or rejects a manuscript, but does not grant revisions. This saves referees time, because they do not have to think about improvements, and it obviously saves authors time, thus expediting publication. The *Economic Inquiry* is a journal that offers this referee option to submitting authors. The *AER Insights* uses this policy as a default (AERI, 2023). The new *JoF: Insights and Perspectives* will also be a no R&R journal. Most other journals have not followed the practice. But “there is overwhelming support among our respondents for this type of model, and across all population subgroups”, as Charness et al. (2022, 3.4.2.12)’s survey among economists finds.

If we think the referee job includes improving the manuscript—and this might be reasonable but too excessive currently—then there are still improvements to be made.

Explicitly separate essential and optional. A small but substantial change would

⁴Charness et al. (2022)’s survey among 1500 active publishing economists shows that the majority of authors believes that evaluation is the main goal. But 1 in 6 view “getting useful feedback” as top priority in the peer-review process, so there is a majority for evaluation, but no consensus.

be to require referees to separate “must address for me to recommend acceptance” comments and “suggestion but I won’t be mad if you don’t follow it” comments. One might say this is already a norm, with the separation into “major” comments and “minor” comments, but that’s not enough. The crucial thing is the *explicit* promise not to punish the authors if the suggestion is not implemented. Referees must be explicit about this, so that there is no ambiguity whether a suggestion can be safely disregarded. [Berk et al. \(2017\)](#) also describe this measure in a guide to referees, but really it should be a profession-wide standard enforced by journals. [Econometrica \(2023\)](#) recently released official referee guidelines exactly asking for this separation.

Max n changes per referee. To avoid long lists of minor changes requested by referees, the profession could restrict the number of requested changes to, say, 5. This forces the referee to think about the main issues and presumably saves time, because minor ones do not have to be spelled out. It also sets an expectation in the profession that referee reports do not have to be long and that referees do not have to impress editors by finding every minor issue ([Berk et al., 2017](#)). For some manuscripts, the upper bound of 5 might be binding, but if a manuscript has more than 5 serious issues, then perhaps a rejection recommendation is in order anyway. The [Econometrica \(2023\)](#) guidelines even have a limit of 3 essential changes. More smaller suggestions would be allowed as optional.

Editorial R&R contracts. If the editor gives an R&R, they have to specify which of the referee comments have to be implemented, and if they are, the manuscript must be accepted ([Charness et al., 2022, 3.4.2.15](#)). This proposal reduces the uncertainty for authors, and requires the editors to think upfront which changes are really important, before the authors do all the revision work in vain. Clearly, the commitment has limits, as there are degrees of how satisfactorily the requested changes are implemented. Nevertheless, such an arrangement would be expected to reduce uncertainty and rejection rates after revisions. But, this measure adds workload to editors, so will likely be unpopular with journals.

Decision without referees. [Frey \(2003\)](#) suggests that referees no longer advise editors on whether to publish a manuscript. Instead, editors take that decision without referee input, and only in case of acceptance are referees invited to provide suggestions for improvement, which however do not have to be followed. Presumably, this proposal is hard to implement, because editors need referee advice on whether the manuscript has merit, except in the rare cases where it is exactly in their field (e.g., [Zilibotti, 2014](#)). Plus, ultimately it means that the acceptance decision is made by a single person, rather than a panel of 2-3, which should make decisions even more noisy (see section 2.4). Hence, this might not be the most practical proposal.

Max one round of revisions. Another solution could be to restrict the rounds of revisions to one. Hence, there is acceptance or rejection after at most one revision, which limits revision work. It is not clear if such a change would be welcomed by all, since some referees might not be on board after one revision, but might have come on board after

two or three. If authors prefer an acceptance after three revisions to a rejection after one, we might see authors prefer the longer but slightly safer option in the status quo. Some journals state they aim to have a definite decision after one revision, but this is typically not a commitment.

Referee awards. A new measure is to introduce a new form of a referee award, chosen by authors. A typical decision uses two or three referee reports. After the editorial decision, the authors can, but do not have to, choose one of the reports for the “best report award.” Once a year, journals send winning referees a certificate with the number of awards they received. These awards might make the refereeing job more attractive, while hopefully generating more reasonable and useful reports for authors. The award should only be given for the initial decision, otherwise referees might be tempted to cause more revision rounds. And, the award only makes sense if there is more than one referee. A concern might be that it biases referees to make more positive recommendations to get the award. [Charness et al. \(2022, 3.3.3.3\)](#) suggest similar recognition, except by editors rather than authors. The median economist in their survey is favorable towards this reform.

Referee rating system. Authors like referees who recommend acceptance or R&R, and dislike those who recommend rejection. But *conditional on the recommendation*, there are still better and worse referee reports. Journals could implement a referee rating system, so editors can track referee quality as judged by authors. This would allow editors to avoid the worst referees. These ratings should not be shared with referees, otherwise it might lead authors to inflate ratings to curry favor. It requires a sufficient sample size to make reasonable decisions, hence this might only work if these ratings are shared between journals, which is difficult. Plus, journals already struggle to find referees, so might not be able to disregard willing but poor referees.

2.2 Strategic refereeing and conflicts of interest

2.2.1 Problem description

Editors routinely recruit authors of related papers to referee a manuscript. The rationale is that these authors know the literature and the methods best, hence they make for expert referees. Indeed, if they already know the literature and method well, they should be able to provide a review with the least extra effort, thus making this a seemingly efficient choice. But often referees are also authors of “competing” working papers that try to get published at roughly the same time. They might then exhibit territorial behavior, unduly criticizing the competing manuscripts and recommending rejection for strategic reasons. This for example might be due to a belief that only one paper on that topic can be published in a top journal, and they want that paper to be theirs.

The former editor of the *Review of Economic Studies* writes: “Sometimes authors want to have exclusive rights to some niches, and they block any entrant. [...] There are a number

of renowned authors whom I never ask to write a report for this reason” (Zilibotti, 2014). Of course, it takes time for an editor to identify such strategic referees, and by then a good number of manuscripts might have been unfairly rejected.

Currently, I am not aware of serious empirical research into this issue, so more work is needed. But there are anecdotes. Berk et al. (2017), for example, write that an editor caught a referee with a competing paper deliberately delaying the referee report, thus delaying publication of the manuscript. The referee was caught later when they made their own paper public. The editor contacted other editors to get them to boycott the paper.

A colleague reported the following anecdote:

We submitted our manuscript to a top 5 journal. One referee report summarized our manuscript, then argued it is terrible, has no identification, etc., very harsh criticism. After the rejection, we submitted to a journal below the top 5. One report had the exact same manuscript summary as the report from the previous submission—it was copy and pasted—but now the evaluation was extremely positive. All the criticisms vanished. For the same manuscript! Apparently, someone wanted to make sure we would not get into the top 5.

Such unethical behavior is hard to spot for editors, since it would require the comparison of reports across two different journals, but editors usually have no access to reports from previous submissions.

2.2.2 Potential solutions

Conflict of interest policies for referees. Many journals—though not all—have conflict of interest policies, which state that referees have to disclose if they have been coauthors, former advisors or students, or family members of any of the authors of the manuscript (e.g., Charness et al., 2022, 3.2.2.3). Other conflicts of interest include referees who have competing manuscripts. While conflict of interest policies might describe and rule out such conflicts, they still rely on the potential referee to disclose them to the editor. Strategic referees might not be so transparent. And even if editors know, they might still be open to having the competing researcher serve as referee, given the difficulty of finding willing referees, in which case it is unclear how they can and do use the (more likely biased) recommendation.

Referee non-disclosure punishments. Punishments for not disclosing conflicts of interest might help, such as getting barred from ever submitting to that journal in the future. After all, undisclosed coauthorships or competing papers will be public at some point, so they might be discovered. If journals can coordinate, banning submission at all participating journals would be an even greater deterrent. This could be easily implementable by big publishers (e.g., Elsevier) or families of journals (e.g., *AER*, *AERI*, *AEJs*). Another punish-

ment might be public condemnation after detection, as is sometimes done with plagiarists. The more it hurts, the more effective it should be.

Referee disqualification. Given that conflicts of interest are sometimes hard to observe for the editor, it is not easy to deal with a referee that is eager to provide a biased report. Journals outside of economics sometimes offer authors to disqualify a few names as referees, which might prevent the worst offenders from being recruited. But this is cannot completely solve the issue, and it could be abused by authors to exclude fair referees.

2.3 Prestige bias and other discrimination

2.3.1 Problem description

Do prominent authors or those from top universities get a more favorable peer-review process? Do unknown authors or those from lower ranked institutions get rejected more often purely due to their identity and affiliation? Normally, these questions are hard to assess, since the quality of a manuscript is not independent of author fame and institution: Higher acceptance rates of famous scholars can be fully deserved on the merits and need not result from a bias in their favor. However, there are now at least two field experimental studies which conclude that there is substantial bias against unknown authors or those from lower ranked universities. These studies provide very credible evidence of a bias, because they hold manuscript quality constant and exogenously vary author identity:

Huber et al. (2022) ran a field experiment at a finance journal with a manuscript written by a Nobel prize winner and a mostly unknown early career researcher. This manuscript was peer-reviewed by hundreds of referees with either the author identity hidden, with only the Nobel prize winner’s name, or with only the early career researcher’s name. The rejection recommendation rate by the referees increased from 23% for the Nobel laureate to 48% when the author identity was unknown, and increased further to 65% with the early career researcher’s name on the manuscript. Thus, *the rejection rate almost triples just based on the author name, for the very same manuscript and the very same institutional affiliation* (both authors were at the same institution). The magnitude of the effect is staggering.

Tomkins et al. (2017) ran a field experiment with submissions for a major computer science conference.⁵ Potential referees were randomly divided into two groups, those that see the author names and affiliations, and those that do not. The same manuscript was reviewed by different referees that know the identity of the authors or not. Tomkins et al. (2017) find that famous authors and top universities get a substantially higher acceptance rate when their identity is known, compared to when it is not known, with acceptance recommendation rates increasing by a factor of 1.6.

It is hard to argue with this experimental evidence. Clearly, there is a status bias in

⁵Unlike in economics, a publication in “conference proceedings” is the primary way to publish results in computer science (e.g., Vardi, 2009).

peer-review. The former *JEEA* editor recognizes the issue and argues that journal incentives favor such biases, because the citing profession has these biases: “There are reasons to believe that the peer-review process is biased in favor of established authors. [...] Although individual authors may feel that this is unfair, statistical discrimination may be in the best interests of the journal. The obsession of publishers and associations with objective impact factors reinforces such a bias” (Zilibotti, 2014). In other words, famous authors attract more citations due to their name, so the journal might be willing to accept a manuscript, whereas it would reject the same manuscript from a less famous author.⁶

One might equally suspect that there is a gender bias against female authors, and possibly racial or nationalist biases. In an early experiment, Blank (1991), manuscripts were randomly assigned to be visible or not to referees at the *AER*. Male authors’ acceptance rate significantly dropped with blinded identity, whereas female authors’ did not change with blinded identity.⁷ Hence, there is some evidence that double-blind review removed a male advantage (rather than removing discrimination against women). In a non-experimental study, Card et al. (2020) find that papers published by female authors receive significantly more citations than comparable male-authored papers, suggesting a higher bar for women. But a caveat is that the result relies on heavy parameterization to make studies across genders comparable. In contrast, Forscher et al. (2019) find no or little gender and racial bias in National Institute of Health grant proposals. Aside from the fact that these are not journal manuscripts and the field is not economics, referees also knew that the grant proposals were altered. Hence they might have suspected to be part of an experiment, which invites experimenter demand effects.

Finally, one might suspect that the same manuscript is viewed more favorably if the number of authors is larger: Clearly, with more hands on deck, it must be better!⁸ But I am not aware of a convincing field experiment investigating such a “team size bias.”

2.3.2 Potential solutions

Unfortunately, there are no perfect solutions for this issue.

Double-blind review. The solution prior to the mid 90s would have been double-blind peer-review. If the referees do not know the author identity, then they cannot discriminate. But keeping the author identity secret is almost impossible today with internet search engines and working papers all over the web. Still, double-blind peer-review might at least partly correct such biases, since some referees might not bother uncovering the author identity. Nevertheless, the *AER* abandoned double-blind review in 2011 (Goldberg, 2012),

⁶Card and DellaVigna (2020) confirm that more famous authors, as measured by the number of publications, attract significantly more citations later.

⁷The gap between male and female acceptance rates did not significantly change, however.

⁸Gibbs et al. (2017) have data from reviews of ideas from a corporate idea suggestion system. They observe larger acceptance rates of ideas for male authors and for larger teams of authors. However, these dimensions were not exogenously varied, so these differences might be due to differences in idea quality.

citing higher administrative costs and the limited ability of referees to detect conflicts of interest. Since the *AER* was the only top 5 journal with a double-blind review policy at the time, the decision might also reflect conformism.

Perhaps surprisingly, the median economist has a slightly unfavorable view of double-blind review, except for junior and female economists (who are likely most affected by review biases), for whom the median is neutral (Charness et al., 2022, Figure 15). In finance, double-blind peer-review is standard at the top 3 journals, in economics single-blind peer-review is standard at the top 5 journals, reflecting a surprising difference in otherwise similar disciplines.⁹ The costs of implementing it are small,¹⁰ and it works partly even today in decreasing discrimination (Smirnova et al., 2023). It is hard to see why double-blind reviewing is not the standard in economics, as it is in other disciplines.

Voluntary double-blind review. A voluntary double-blind peer-review, where submitting authors choose between single- and double-blind peer review, is not promising in terms of incentives, because those benefiting from the bias would not exercise this option. Thus, a separating equilibrium might emerge where getting an anonymous manuscript is the same as receiving an unknown author’s manuscript. Despite this concern, in a field experiment, voluntary anonymization helped increase the acceptance rate of unknown authors slightly (Smirnova et al., 2023), though not enough to offset the large biases estimated in the above field experiments. Still, if voluntary anonymization can help, then mandatory anonymization should help even more.

2.4 Noisy outcomes

2.4.1 Problem description

At the best journals in economics, the reason for rejection is often not a factual error or objective flaw, but a subjective assessment by referees or the editor that they do not like the manuscript or do not see a sufficient contribution.¹¹ Since the law of large numbers does not apply with a panel of 2-3 referees, the outcome—i.e., whether the majority recommendation is reject or revise/accept—is extremely noisy. The same paper would get a revision one day while getting rejected the next, because different academics ended up accepting the invitation to referee. Thus, the submission process resembles a lottery. This noise favors topics and methods that are considered general interest or might be a fit at a wide array of journals, and disadvantages less fashionable topics or methods that only have one or two decent journals with the topic within their scope.

But how large is the disagreement among referees? A classical experiment in psychology

⁹But not all economics journals use single-blind peer-review. Exceptions include *Economic Inquiry*, *Environmental and Resource Economics*, *Journal of Economic Surveys* as of March 2023.

¹⁰Authors just have to remove their names plus perhaps replace any mentioning of “my” or “our”. Journals could automate checking the manuscripts for mentions of the author names, etc.

¹¹Which might also be a reason why the biases I discuss in 2.3 can arise.

resubmitted 12 journal articles that had been published in the past 32 months at the same journal (Peters and Ceci, 1982). 3 of these articles were identified by editors or referees as already published, so 9 went through the full referee process. Of the 18 referee reports thus obtained, 16 recommended rejection. So 8 of the 9 papers were rejected. It would seem that getting published was a matter of luck if the same paper could not make it through twice. Of course, by today's standards, the sample size of 9 is awfully small.

More modern studies use referee recommendation statistics to compute referee agreement for the same manuscript. This requires multiple referees per paper, which is usually the case with economics journals, but less so with finance journals. Welch (2014) shows that the correlation coefficient between referee recommendations on the same manuscript is 0.25 among six economics journals, and 0.15 at two of the top finance journals.¹² Agreement tends to be lower at higher ranked journals and at finance journals. Using a structural decomposition model, Welch (2014) argues that a referee-specific signal has a 2/3 weight whereas a common-signal only has a 1/3 weight in explaining referee recommendations. Card and DellaVigna (2020, footnote 20) also find a correlation between referee recommendations of 0.25 in economics journals. Hence, agreement among referees is modest at best.

2.4.2 Potential solutions

Reusing prior reports (mandatory). Recruiting more referees would make the outcome less noisy, but is prohibitively costly. An effective approach might be a centralized system that keeps track of all prior reviews across journals (with the possibility of the authors briefly responding to previous reports).¹³ But that has its own issues and is hard to implement. For example, it would require all publishers to share information, and all referees to agree to share their reviews across journals. The latter could be achieved by only allowing academics as referees that agree, but that would make finding suitable referees even harder. Moreover, the manuscript might change after the first submission, but that could be dealt with via brief author responses (“we solved this concern by doing X in the new version”). Charness et al. (2022, 3.2.4.4) are somewhat negative on the idea, since a single report might punish authors multiple times.¹⁴

Reusing prior reports (voluntary). Some journals, such as the EJ (2023), allow authors to voluntarily hand in past referee reports from other journals with the submission, though that might lead to a selective omission of negative past reports. Other journals, such as the AEJs in the AEA family or QE and TE in the Econometric Society family, allow

¹²It is unclear whether the lower agreement among finance referees is due to their shorter deadlines compared to econ, or other differences. If it is due to the shorter deadlines, then there would be a trade-off between faster review and this dimension of quality.

¹³Azar (2006) describes a similar system.

¹⁴However, if a manuscript is objectively flawed or an author cheated in some way (e.g., designed the experiment for something else, then p-hacked and wrote it up differently), then we want the manuscript to be rejected everywhere. But without a central reports facility, the author can just resubmit repeatedly and hope the flaw is not found next time.

the transfer of past reports from the family’s top journal, which is a limited version of the above centralized solution, where selective omission by the author is not possible. Editors could forgo inviting new referees by using these past reports, and grant revisions in case of marginal rejections at a superior journal or desk reject with more information. Editors can always discard the prior reports, so this is an improvement if it sometimes leads to a decision without new review. Given the voluntary nature on both sides, it is hard to see the harm of allowing the possibility.¹⁵ One reason editors might be reluctant to use prior reports from other journals, without recruiting new referees, is that these prior referees cannot be consulted on a revised manuscript later.

Referee training. A standard way to increase inter-rater agreement is to provide clear guidelines and training (e.g., [Sattler et al., 2015](#)), which is largely missing in economics. Referee guidelines, to the degree they exist, do not actually describe what a good manuscript is, and focus instead on procedural points. Moreover, there is typically no training on writing referee reports in graduate schools. [Charness et al. \(2022, 3.3.3.2\)](#) find that a majority of economists view doctoral training on refereeing favorably. However, voluntary training already exists online, and mandatory training during PhD studies may crowd out more important things – like learning how to do research and then doing it.

Defining standards of quality. Possibly, stronger norms and agreement on what methods and approaches are valid would help referee agreement. For example, experimental economics has strong norms that subjects should be paid by performance, should not be lied to, and failure to do either of these is a very likely cause for rejection. In experimental psychology, there is no such agreement on how to run experiments ([Hertwig and Ortmann, 2001](#)). But such restrictions on valid approaches and methods, let alone topics—if they can be agreed upon at all—might also make economic research more dogmatic and less innovative (e.g., [Akerlof, 2020](#)), and hence do not seem like a good idea.

2.5 The role of editors

The discussion of issues in economics peer-review would not be complete without the role of editors. There is not as much literature or research on it, so the public discussion is not as advanced as it is for the role of referees. Hence, this section gives a first overview of potential issues, but I will not make explicit recommendations based on this early discussion.

Editors have far more power and information (e.g., not blinded even under double-blindness) than referees. Editors are the first line of defense against bad scholarship and decide whether to desk-reject or not, usually without rules, controls, or recourse for authors. While desk-rejections are an essential tool to speed up the review-process, it also invites decision-making based on heuristics, which may generate biases, such as using the author

¹⁵One concern might be how to authenticate prior reports handed in by the author – they could have just been made up.

affiliation or publication history for the decision.¹⁶ If the above field experiments with referees (who often spend hours with the manuscript) find a large prestige bias, then editors who spend only minutes might be even more prone to such biases – especially if a prestige bias is in the interest of an impact-factor maximizing journal. Of course, this does not mean that desk-rejections should be abandoned—there are obvious cases way below the bar that should be rejected quickly—but an argument could be made that double-blindness should apply to editors even more than to referees (e.g., [Hassel, 2021](#)).

If the authors pass the desk-rejection stage, the editor selects referees. Again, the editor has a lot of discretion, and can choose soft or tough referees, those with a competing paper or those that do not, all of which affect the chances of positive reports. As before, without a blinded author identity, editors might subconsciously favor more prestigious authors, male authors, friendly authors, etc. Since even anonymous data on referees is closely guarded, research into referee selection is scarce.

Finally, once referee reports are in, the editor decides whether to grant an R&R or whether to reject. [Welch \(2014\)](#) and [Card and DellaVigna \(2020\)](#) show that editors follow referee recommendations quite closely. But given the previous two stages, editors already had a large impact on the outcome by the time referee reports are in. Another issue at this stage might be missing guidance by the editor in case of conflicting or unreasonable referee comments. Editors could use their authority to limit referee overreach, but rarely do. And at times, editors bring in new referees after the first revision, adding more uncertainty and possibly work for authors. A colleague writes: “You revise the paper but the editor then brings in another referee who is clearly commissioned to trash the paper and then rejects it on that basis.” While we cannot know what is said between editors and referees that are brought in late, there may be good reasons for it (e.g., another referee became unavailable or a new issue arose, requiring new expertise).

There is a large literature on in-group bias by editors, trying to answer whether editors publishing colleagues or former students more often is favoritism, better alignment of topics/methods, or justified based on manuscript quality. [Colussi \(2018\)](#) shows that authors publish significantly more in a top 5 journal if the editor is at the same institution. This effect cannot be explained by closer alignment of research fields, and suggests there is an easier way in for authors with social ties to an editor. [Lutmar and Reingewertz \(2021\)](#) find that Harvard and MIT publications in the house journal *QJE* receive fewer citations than Harvard/MIT publications at other top 5 journals, suggesting a home bias in that journal.¹⁷

¹⁶On a day when I got desk-rejected, someone from the city where the editor is based spent several minutes on my website and looked repeatedly at several pages (based on google analytics). Usually, nobody visits from that city. Apparently, my website was at least as important as the manuscript I submitted in making that decision.

¹⁷Of course, citations depend on more than paper quality, such as journal, topic of study, research method, identity of authors, number of authors (e.g., [Card and DellaVigna, 2020](#)), even the position of the article in the issue (lead articles get more citations even if the order is random, see [Coupé et al., 2010](#)). Hence, citations as measure of quality are far from perfect, and studies relying on controlling for confounding variables may miss some.

Bethmann et al. (2023) have very similar findings. Hackman and Moktan (2020) suspect the long editorial terms at house journals could cause in-group bias. Brogaard et al. (2014) also find that authors with the same affiliation as the editor publish significantly more often in that journal, but these publications also have higher citations, indicating a higher quality. Hence, the evidence is not unanimous, yet the major studies looking at the *QJE* in particular tend to find an in-group bias.

A final concern might be editors publishing in their own journals. They might not be in charge of their own submissions, but their editor-colleagues and possibly friends are. In addition, referees might be worried that their anonymity is threatened, since the submitting editor might nevertheless have editorial access to the review system and might uncover the identities of critical referees. All of these factors suggest a submitting editor might get a more favorable author experience than an author not editing the journal.

These issues suggest there needs to be a discussion in the profession how much independence and discretion editors should have. More checks and balances might increase review times. On the other hand, some targeted changes might also lead to more meritocratic outcomes. Reform candidates include triple-blind review (editors don't know author identities), algorithms selecting referees purely based on topic (thus avoiding favoritism in referee selection), or no drafting of new referees after the first revision.

3 Recommendations for reform

There is no shortage of ideas. Many solutions to improve peer-review have been suggested before (e.g., McAfee, 2014; Berk et al., 2017; Charness et al., 2022), including somewhat radical proposals (e.g., Frey, 2003; Azar, 2006). What is needed for change is coordinated action, especially because part of the problem are expectations of long papers and extensive revisions (Ellison, 2002a), which are not easily changed by a single journal.

The referee guidelines of *Econometrica* (2023) are great: short, to the point, addressing the referee overreach problem, plus setting expectations that reports do not have to be long, thus likely reducing review times. But this is only one journal. This is not enough to reset expectations profession-wide. We need a critical mass of journals to enforce similar referee guidelines, then these might be established as a new standard. The crucial ingredients of the *Econometrica* (2023) guidelines are the separation in essential and optional requested changes, and the limit of n essential changes. Any other guidelines with similar provisions might do as well. We should get to an understanding that a referee is not a coauthor, and it is not the referee's job to rewrite the manuscript.

A system of referee awards for the most useful report is worth a trial. There is an implementation cost by journals, but since they are often struggling to recruit referees, they might be interested in a solution that makes the job more attractive. And if it improves

reports in addition, then all the better. I am not aware that this system is in practice anywhere, so the outcomes are speculation, but it sounds promising in theory.

There is no perfect solution for prestige bias, or gender and racial biases, as double-blind peer-review can be circumvented. Double-blind review can help a little to reduce bias, and given the small costs of implementation, it should be the standard.

In summary, my recommendations (for detailed discussions, see section 2):

1. To prevent excessive referee asks, all journals should adopt the same referee guidelines which require referees to separate essential and optional changes they request, and limit the essential changes to 3.¹⁸
2. To make referee reports more useful and to motivate referees, let authors chose one referee per submission for the “best report” award (excluding resubmissions).
3. To minimize strategic refereeing, all journals should adopt the same conflict of interest policies, which describe conflicts of interest including competing research, and require referees to disclose these conflicts to the editor. Non-disclosure should be punished, for example by banning the referee from submitting to all journals using these same policies. Editors should avoid using referees with conflicts of interest.
4. To minimize review biases, double-blind peer-review should be used.
5. To reduce noise in review outcomes and to reduce review times, journals should accept prior referee reports without commitment to use them or to forgo inviting new referees.¹⁹ If possible, transfer of reports between journals should be facilitated, to avoid selective disclosure of reports by authors.²⁰
6. To speed up review times, 4 week review deadline should be set and referees should be paid for timely reports (at journals that take a submission fee).

4 Concluding remarks

The functioning of the peer-review process is something that affects every publishing economist. Many have debated solutions and proposals to improve peer-review, but ultimately journals need to make changes. These need to be coordinated changes, so that we avoid a confusing state with each journal having their own guidelines and rules (especially important regarding referee guidelines and conflict of interest policies). With sufficient adoption, we should be able to change expectations and norms, so that the issues discussed here are minimized.

¹⁸Already practiced at *Econometrica* (2023).

¹⁹Already practiced at *EJ* (2023).

²⁰Already practiced at *AEJs*, *QE* and *TE*, *JPE-Micro* and *JPE-Macro*.

Open questions for research remain. How pervasive are conflicts of interest in refereeing? What impacts do they have on publication decisions? What measures work against them? And do the new *Econometrica*-style referee guidelines have a measurable impact on referee reports (e.g., length), refereeing times, and extent of revisions? Such research is important to inform a possible reform effort in economics peer-review further. And more research is needed on the role of editors, in particular desk-rejection decisions and referee selection.

Bibliography

- AERI (2023): “Submission Guidelines,” <https://www.aeaweb.org/journals/aeri/submissions/guidelines>, accessed 11/3/2023.
- AKERLOF, G. A. (2020): “Sins of Omission and the Practice of Economics,” *Journal of Economic Literature*, 58, 405–418.
- AZAR, O. H. (2006): “The academic review process: How can we make it more efficient?” *The American Economist*, 50, 37–50.
- BERK, J. B., C. R. HARVEY, AND D. HIRSHLEIFER (2017): “How to write an effective referee report and improve the scientific review process,” *Journal of Economic Perspectives*, 31, 231–244.
- BETHMANN, D., F. BRANSCH, M. KVASNICKA, AND A. SADRIEH (2023): “Home Bias in Top Economics Journals,” .
- BLANK, R. M. (1991): “The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review,” *The American Economic Review*, 1041–1067.
- BRODEUR, A., N. COOK, AND A. HEYES (2020): “Methods matter: P-hacking and publication bias in causal analysis in economics,” *American Economic Review*, 110, 3634–3660.
- BRODEUR, A., M. LÉ, M. SANGNIER, AND Y. ZYLBERBERG (2016): “Star wars: The empirics strike back,” *American Economic Journal: Applied Economics*, 8, 1–32.
- BROGAARD, J., J. ENGELBERG, AND C. A. PARSONS (2014): “Networks and productivity: Causal evidence from editor rotations,” *Journal of Financial Economics*, 111, 251–270.
- CAMERER, C. F., A. DREBER, E. FORSELL, T.-H. HO, J. HUBER, M. JOHANNESSON, M. KIRCHLER, J. ALMENBERG, A. ALTMEJD, T. CHAN, ET AL. (2016): “Evaluating replicability of laboratory experiments in economics,” *Science*, 351, 1433–1436.
- CARD, D. AND S. DELLAVIGNA (2013): “Nine facts about top journals in economics,” *Journal of Economic Literature*, 51, 144–161.
- (2020): “What do editors maximize? Evidence from four economics journals,” *Review of Economics and Statistics*, 102, 195–217.

- CARD, D., S. DELLAVIGNA, P. FUNK, AND N. IRIBERRI (2020): “Are referees and editors in economics gender neutral?” *The Quarterly Journal of Economics*, 135, 269–327.
- CHARNESS, G., A. DREBER, D. EVANS, A. GILL, AND S. TOUSSAERT (2022): “Improving peer review in economics: Stocktaking and proposals,” <https://evalresearch.weebly.com/report.html>.
- CHETTY, R., E. SAEZ, AND L. SÁNDOR (2014): “What policies increase prosocial behavior? An experiment with referees at the Journal of Public Economics,” *Journal of Economic Perspectives*, 28, 169–188.
- COLUSSI, T. (2018): “Social ties in academia: A friend is a treasure,” *Review of Economics and Statistics*, 100, 45–50.
- COUPÉ, T., V. GINSBURGH, AND A. NOURY (2010): “Are leading papers of better quality? Evidence from a natural experiment,” *Oxford Economic Papers*, 62, 1–11.
- ECONOMETRICA (2023): “Guidelines for Referees,” <https://www.econometricsociety.org/publications/econometrica/guidelines-referees>, accessed 11/3/2023.
- EJ (2023): “Economic Journal: Instructions to Authors,” https://academic.oup.com/ej/pages/General_Instructions#submissionpast, accessed 15/3/2023.
- ELLISON, G. (2002a): “Evolving standards for academic publishing: A q-r theory,” *Journal of Political Economy*, 110, 994–1034.
- (2002b): “The slowdown of the economics publishing process,” *Journal of Political Economy*, 110, 947–993.
- FORSCHER, P. S., W. T. COX, M. BRAUER, AND P. G. DEVINE (2019): “Little race or gender bias in an experiment of initial review of NIH R01 grant proposals,” *Nature Human Behaviour*, 3, 257–264.
- FREY, B. S. (2003): “Publishing as prostitution?—Choosing between one’s own ideas and academic success,” *Public Choice*, 116, 205–223.
- FRIBERG, R. (2014): “The Wish List of an Editor: Some Reflections on Editing the Scandinavian Journal of Economics,” in *Secrets of Economics Editors*, ed. by M. Szenberg and L. Ramrattan, MIT Press, 45–52.
- GIBBS, M., S. NECKERMANN, AND C. SIEMROTH (2017): “A field experiment in motivating employee ideas,” *Review of Economics and Statistics*, 99, 577–590.
- GOLDBERG, P. (2012): “Report of the Editor: American Economic Review,” *American Economic Review: Papers & Proceedings*, 102, 653–665.
- HACKMAN, J. AND S. MOKTAN (2020): “Publishing and Promotion in Economics: The Tyranny of the Top Five,” *Journal of Economic Literature*, 58, 419–470.

- HADAVAND, A., D. S. HAMERMESH, AND W. W. WILSON (2021): “Publishing Economics: How Slow? Why Slow? Is Slow Productive? Fixing Slow?” Tech. rep., National Bureau of Economic Research.
- HAMERMESH, D. S. (1994): “Facts and myths about refereeing,” *Journal of Economic Perspectives*, 8, 153–163.
- HASSEL, H. (2021): “Desk Rejects Widen Inequalities within Academia,” *Political Science & Politics*, 54, 699–702.
- HERTWIG, R. AND A. ORTMANN (2001): “Experimental practices in economics: A methodological challenge for psychologists?” *Behavioral and Brain Sciences*, 24, 383–403.
- HUBER, J., S. INOUE, R. KERSCHBAMER, C. KÖNIG-KERSTING, S. PALAN, AND V. L. SMITH (2022): “Nobel and novice: Author prominence affects peer review,” *Proceedings of the National Academy of Sciences*, 119, e2205779119.
- LUTMAR, C. AND Y. REINGEWERTZ (2021): “Academic in-group bias in the top five economics journals,” *Scientometrics*, 126, 9543–9556.
- MCAFEE, P. (2014): “Edifying editing,” in *Secrets of Economics Editors*, ed. by M. Szenberg and L. Ramrattan, MIT Press, 53–66.
- PETERS, D. AND S. CECI (1982): “The fate of published articles, submitted again,” *Behavioral and Brain Sciences*, 5, 187–255.
- RUBINSTEIN, A. (2017): “Comments on economic models, economics, and economists: remarks on economics rules by Dani Rodrik,” *Journal of Economic Literature*, 55, 162–172.
- SATTLER, D. N., P. E. MCKNIGHT, L. NANEY, AND R. MATHIS (2015): “Grant peer review: improving inter-rater reliability with training,” *PloS one*, 10, e0130450.
- SMIRNOVA, I., D. M. ROMERO, AND M. TEPLITSKIY (2023): “Nudging Science Towards Fairer Evaluations: Evidence From Peer Review,” *SSRN 4190623*.
- TE (2023): “Guidelines for referees at Theoretical Economics,” <https://econtheory.org/refereeGuidelines.php>, accessed 15/3/2023.
- TOMKINS, A., M. ZHANG, AND W. D. HEAVLIN (2017): “Reviewer bias in single-versus double-blind peer review,” *Proceedings of the National Academy of Sciences*, 114, 12708–12713.
- VARDI, M. Y. (2009): “Conferences vs. journals in computing research,” *Communications of the ACM*, 52, 5–5.
- WELCH, I. (2014): “Referee recommendations,” *The Review of Financial Studies*, 27, 2773–2804.
- ZILIBOTTI, F. (2014): “Memoirs of an editor,” in *Secrets of Economics Editors*, ed. by M. Szenberg and L. Ramrattan, MIT Press, 97–117.