

Is Peer Review a Good Idea?

Remco Heesen and Liam Kofi Bright*

Abstract

Prepublication peer review should be abolished. We consider the effects that such a change will have on the social structure of science, paying particular attention to the changed incentive structure and the likely effects on the behaviour of individual scientists. We evaluate these changes from the perspective of epistemic consequentialism. We find that where the effects of abolishing prepublication peer review can be evaluated with a reasonable level of confidence based on presently available evidence, they are either positive or neutral. We conclude that on present evidence abolishing peer review weakly dominates the status quo.

- 1 *Introduction*
- 2 *Setting the Stage*
- 3 *Benefits of Abolishing Peer Review*
 - 3.1 *Sharing scientific results*
 - 3.2 *Time allocation*
 - 3.3 *Gender skew in publications*
 - 3.4 *Library resources*
 - 3.5 *Scientific careers*
 - 3.6 *The power of gatekeepers*

*Both authors contributed equally.

4 *Where Peer Review Makes No Difference*

4.1 *Epistemic sorting*

4.2 *Malpractice detection*

4.3 *Herding behaviour*

4.4 *Long-run credit*

5 *Difficulties for Our Proposal*

5.1 *A guarantee for outsiders*

5.2 *A runaway Matthew effect*

6 *Conclusion*

1 Introduction

Peer review plays a central role in contemporary academic life. It sits at the critical juncture where scientific work is accepted for publication or rejected. This is particularly clear when the results of scientific work are communicated to non-scientists. The question ‘Has this been peer reviewed?’ is commonly asked, especially by journalists, and a positive answer is frequently taken to be a necessary and sufficient condition for the results to be considered serious science.

Given these circumstances, one might expect peer review to be an important topic in the philosophy of science as well. Peer review should arguably play a more prominent role in the debate about demarcation criteria (what separates science from other human pursuits?), as it seems to be used in practice exactly to differentiate scientific knowledge from other claims to knowledge, at least by journalists. Yet social-procedural accounts of science, like the one found in (Longino [1990]), remain in the minority and usually do not place great emphasis on peer review in particular. Aside from this particular debate, there are normative questions about the proper epistemic role of peer review and more practical questions about the extent to which it manages to fulfil them, all of which should interest philosophers of science.

Here we bring together the work of philosophers of science (especially social epistemologists of science) who have written about the strengths and weaknesses of various aspects of the social structure of science and empirical work about the effects of peer review. We argue that where

philosophers of science have claimed the social structure of science works well, their arguments tend to rely on things other than peer review, and that where specific benefits have been claimed for peer review, empirical research has so far failed to bear these out. Comparing this to the downsides of peer review, most prominently the massive amount of time and resources tied up in it, we conclude that we might be better off abolishing peer review.

Whereas philosophers have investigated various aspects of the social structure of science, there has been surprisingly little reflection on peer review in particular. Most of what exists has focused on the role of biases in peer review; see, for example, (Saul [2013], Section 2.1; Lee [2012]; Lee *et al.* [2013]; Jukola [2017]; Katzav and Vaesen [2017]; Heesen [2018a]). Only very occasionally have philosophers turned to discussing the strengths and weaknesses of peer review as such. Among our allies here we count Katzav and Vaesen ([2017]), who explicitly wish to minimize the role of peer review, and Atkinson ([1994]), who gives a fairly fundamental critique of peer review as such. A philosophically sophisticated book-length treatment of peer review and its problems can be found in (Shatz [2004]), although it deliberately (see p. 12) omits a detailed discussion of technical issues in epistemology that are a significant part of our focus here. Further, there is some work critical of prepublication peer review outside of philosophy, including letters and opinion pieces (Gibson [2007]; Gowers [2017]), a special issue collecting scientists' views (Kriegeskorte *et al.* [2012]), and some full-length articles (Smith [2006]); in this category Nosek and Bar-Anan ([2012]) and Teixeira da Silva and Dobránszki ([2015]) are particularly close to our proposal. We have benefited a great deal from such work, but it tends to be vague about the normative standard against which peer review or its alternatives are to be evaluated. This is something we aim to remedy in Section 2.

Some brief clarifications. Our target is prepublication peer review, that is the review of a manuscript intended for publication, where publication is withheld until one or more editors deem the manuscript to have successfully passed peer review. We set aside other uses of peer review (of grant proposals or conference abstracts) and we explicitly leave room for postpublication peer review, where manuscripts are published before review. The key change we are arguing for, then, is to move the 'date of publication' such that an article is considered published before it enters peer review. Because of this last point, some readers may think that our

terminology ('abolishing prepublication peer review') suggests a more dramatic change than what we actually advocate. We invite such readers to substitute in their preferred terminology. We should also clarify that we use 'science' in a broad sense to include the natural sciences, the social sciences, and the humanities.

The overall structure of our argument is as follows: We think there are a number of clear benefits to abolishing prepublication peer review. In contrast, while various benefits of the existing system (downsides of abolishing peer review) have been suggested, we do not think there exist any that have clear empirical support. Insofar as empirical research exists, it is ambiguous in some cases, and speaks relatively clearly against the claimed benefit of the existing system in others. While we admit to a number of cases where the evidence is ambiguous or simply lacking (see especially Section 5), we claim that the present state of the evidence suggests that abolishing prepublication peer review would lead to a peculiar sort of Pareto improvement: each factor considered is either neutral or favours our proposal.

We say that this is a 'peculiar sort of Pareto improvement' because it is not a typical dominance argument. We do not mean that for every possible state of the world our proposal comes out preferable to the status quo, so that we do not need to know which state of the world we are in to know that our proposal is preferable. Nor that all individual scientists are at least as well off under our proposal compared to the status quo, so that we can bypass interpersonal comparisons of utility. We shall instead be comparing things across different standards of evaluation rather than different possible states of the world. Hence, rather than not needing to know the probability distribution over states of the world, what we are saying is that one does not need to know precisely how to weigh the significance of the different factors we consider.

This comparison across evaluative standards rather than states of the world has consequences for the epistemic strength of our argument. For each evaluative criterion we consider we say either that presently available evidence supports a preference for our alternative over the status quo, present evidence suggests that the status quo and our alternative are equally good per this criteria, or present evidence is insufficient to conclude one way or another. So where evidence is sufficient to make definite conclusions, our claim is that depending on how one weighs the different factors one will never be able to say that the status quo is preferable to our proposal,

and will sometimes be able to say that our proposal is preferable to the status quo. However, for some criteria of evaluation present evidence does not warrant definite conclusions. Hence there is considerable inductive uncertainty about our conclusion. This is something we aim to be upfront about, and will repeatedly draw to the reader's attention to prevent our article being misread.

Regarding the alternative to the status quo we shall be considering, note that our primary aim is to evaluate the current system. However, we believe that is only really possible by comparing it to an alternative. We are not claiming that the proposal we put forward is the best of all possible alternatives. It has been constructed to be a system that could constitute the aforementioned type of Pareto improvement over the current system. Given that it has not been implemented yet to the full extent that we favour, we cannot guarantee it would work as advertised or what empirical properties it would have. But in offering a relatively specific alternative, we hope to get people thinking about real change, which pointing out problems with the present system has so far failed to do.

Even despite these provisos, we realize that ours is a strong claim, and our proposal a large change to the social structure of science. It is therefore important to repeatedly highlight that our central claim concerns the balance of presently available evidence. We are not further claiming that the matter is so conclusively settled as to render further research superfluous or wasteful. On the contrary, we think there are a number of points in our argument where present evidence is severely limited, and we take the calls for further empirical research we make in those places to be just as important a part of the upshot of our article as our positive proposal. We hope, therefore, that even a sceptical reader will read on; if not to be convinced of the need of abolishing prepublication peer review, then at least to see where in our view their future research efforts should concentrate if they are to shore up prepublication peer review's claims to good epistemic standing.

2 Setting the Stage

The purpose of peer review is usually construed in terms of quality control. For example, Katzav and Vaesen ([2017], p. 6) write, 'The epistemic role of peer review is assessing the

quality of research’, and this seems to be a common sentiment per Eisenhart ([2002], p. 241) and Jukola ([2017], p. 125). But how well does peer review succeed in its purpose of quality control? The empirical evidence (reviewed below) is mixed at best. As one prominent critic puts it, ‘we have little evidence on the effectiveness of peer review, but we have considerable evidence on its defects’ (Smith [2006], p. 179).

Peer review’s limited effectiveness would perhaps not be a problem if it required little time and effort from scientists. But in fact the opposite is true. Going from a manuscript to a published article involves many hours of reviewing work by the assigned peer reviewers and a significant time investment from the editor handling the submission. The editor and reviewers are all scientists themselves, so the epistemic opportunity cost of their reviewing work is significant: instead of reviewing, they could be doing more science.

Given these two facts—high (epistemic) costs and unclear benefits—we raise the question whether it might be better to abolish prepublication peer review. In the following, we provide our own survey and assessment of the evidence that bears on this question. Our conclusions are not sympathetic to peer review. However, we encourage any proponents of peer review to give their own assessment. We only ask that any benefits claimed for peer review are backed up by empirical research, and that they are epistemic benefits. That is, we ask for empirical evidence that peer review makes for better science on science’s own terms.

We take the status quo to be as follows: The vast majority of scientific work is shared through journal publications, and the vast majority of journals uses some form of prepublication peer review. Ordinarily this means that an editor assigns one to three peers (scientists whose expertise intersects the topic of the submission), who provide a report and/or verdict on the submission’s suitability for publication. The peer reviews feed into the final judgement: the submission is accepted or rejected with or without revisions.

Our proposal is to abolish prepublication peer review. Scientists themselves will decide when their work is ready for sharing. When this happens, they publish their work online on something that looks like a preprint archive (think arXiv, bioRxiv, or PhilSci-archive, although the term ‘preprint’ would not be appropriate under our proposal). Authors can subsequently publish updated versions that reply to questions and comments from other scientists, which

may have been provided publicly or privately. The business of journals will be to create curated collections of previously published articles. Their process for creating these collections will involve (postpublication) peer review, insofar as they currently use prepublication peer review.

We are not the first to make such suggestions; see, for example, (Gibson [2007]; Nosek and Bar-Anan [2012]; Teixeira da Silva and Dobránszki [2015]; Katzav and Vaesen [2017]). Moreover, our proposal is in line with how certain parts of mathematics and physics already work: uploading an article to arXiv is considered publishing it for most purposes, with journal peer review and publication happening almost as an afterthought (Gowers [2017]). It seems that journal publication can function as something like a prize, accruing glory to the scientist who achieves it but doing little to actually help spread the idea beyond calling attention to something that has already been made public elsewhere. Further, various disciplines and journals are experimenting along the lines we propose, sometimes responding to just the sort of concerns we outline in this article. For instances of such experimentation see the practices of F1000, the Winnower, peerJ, and PLoS. Journals associated with the European Geoscience Union are run in a way very conducive to our proposal and their website has a clear description of their procedure and its rationale.¹ We are not aware of any detailed comparative studies of the effects these changes have had in those fields, so we will not rest any significant part of our argument on them. But for those who worry that science will immediately and irrevocably fall apart without peer review, we point out that this does not appear to have happened in the relevant parts of mathematics, physics, or geoscience. Indeed, we take ourselves to be providing argumentative support for a shift in scientific norms that is already occurring and changing practice in some areas.

In the remainder of this article we break down the consequences of our proposal. Our strategy here is to focus on a large number (hopefully all) aspects of the social structure of science that will be affected. In particular, the reader may already have a particular objection against our proposal in mind. We encourage such a reader to skip ahead to the section where this objection is discussed before reading the rest of the article.

For example, one reader may think that peer review as currently practised is important be-

¹See <www.egu.eu/publications/>.

cause it forces scientists to read and review each other's work, and without peer review they will spend less time on such tasks. This is discussed in Section 3.2. Another reader may worry that without peer review and the journal publications that go with them it will be more difficult to evaluate scientists for hiring or promotion (Section 3.5). Yet another reader may be concerned about losing peer review's ability to prevent work of little merit from being published, or at least to sort articles into journals by epistemic merit so scientists can easily find good work (Section 4.1). A fourth reader might think peer review plays an important role in detecting fraud or other scientific malpractice (Section 4.2). A fifth reader may think the guarantee provided to outsiders when something has been peer reviewed is an important reason to preserve the status quo (Section 5.1). And a sixth reader may want to point out that anonymized peer review gives relatively unknown scientists a chance at an audience by publishing in a prestigious journal, whereas on our proposal perhaps only antecedently prominent scientists will have their work read and engaged with (Section 5.2).

Other aspects of the social structure of science that will be considered: whether and when scientists share their work (Section 3.1), how many articles are published by women or men (Section 3.3), library resources (Section 3.4), the power of editors as gatekeepers (Section 3.6), science's susceptibility to fads and fashions (Section 4.3), and ways to get credit for scientific work other than through journal publications (Section 4.4). In each case we evaluate whether the net effects of our proposal on that aspect can be expected to be positive. To tip our hand: aspects where we will claim a benefit are gathered in Section 3, aspects where we expect little or no change are in Section 4, and aspects that we consider neutral due to a present lack of evidence are in Section 5.

In making these evaluations, we commit to a kind of epistemic consequentialism (see Goldman [1999]). One may think of what we are doing as roughly analogous to the utilitarian principle, where for each issue our yardstick is whether prepublication peer review shall generate the greatest amount of knowledge produced in the least amount of time. More specifically, we consider changes in the incentive structure and expected behaviour of scientists, as well as other changes that would result from abolishing prepublication peer review. We evaluate these changes in terms of their expected effect on the ability of the scientific community to produce

scientific knowledge in an efficient manner. Working out in detail what such an epistemic consequentialism would look like would be very complicated, and we do not attempt the task here. For most of the issues we consider, we think that the calculus is sufficiently clear that fine details do not matter. Where it is unclear (the issues discussed in Section 5) we think this results from ignorance of empirical facts about the likely effect of policies, rather than conceptual unclarity in the evaluative metric. So we do not need to use our consequentialist yardstick to settle any difficult tradeoffs. All we need for our purposes is to make it clear that we are evaluating the peer review system by how well it does in incentivising efficient knowledge production. Notice that by committing to this epistemic yardstick we set aside other things one might care about regarding the production and evaluation of scientific knowledge. For example, we will occasionally mention fairness considerations where we think they affect epistemic considerations (see Sections 3.3, 3.5, and 5.2), but we will not attempt an overall evaluation of our proposal in terms of fairness.

What do we mean by the incentive structure of science, mentioned in the previous paragraph? This addresses the motivations of scientists. Scientists are rewarded for their contributions with credit, that is, with recognition from their peers as expressed through such things as awards, citations, and prestigious publications (Merton [1973b]; Hull [1988]; Zollman [2018]). Scientific careers are largely built on the reputations scientists acquire in this way (Latour and Woolgar [1986], Chapter 5). As a result, scientists engage in behaviours that improve their chances of credit (Merton [1973c]; Dasgupta and David [1994]; Zollman [2018]).

While individual scientists may be motivated by credit to different degrees (curiosity, the thrill of discovery, and philanthropic goals are important motivations for many as well), the effect on careers means that credit-maximizing behaviour is to some extent selected for. Thus we think it important to ensure that our proposal does not negatively affect the incentives currently in place for scientists to work effectively and efficiently.

3 Benefits of Abolishing Peer Review

3.1 Sharing scientific results

An important feature of (academic) science is that there is a norm of sharing one's findings with the scientific community. This has been referred to as the communist norm (Merton [1973a]). In recent surveys, scientists by and large confirm both the normative force of the communist norm and their actual compliance (Louis *et al.* [2002]; Macfarlane and Cheng [2008]; Anderson *et al.* [2010]). This norm is epistemically beneficial to the scientific community, as it prevents scientists from needlessly duplicating each other's work.

Will abolishing peer review affect this practice? In order to answer this question, we need to know what motivates scientists to comply with the communist norm, that is to share their work. On the one hand there is the feeling that they ought to share generated by the existence of the norm itself. There is no reason to expect this to be changed by abolishing peer review.

On the other hand there is the motivation generated by the desire for credit. According to the priority rule, the first scientist to publish a particular discovery gets the credit for it (Merton [1973b]; Dasgupta and David [1994]; Strevens [2003]). So a scientist who wants to get credit for her discoveries has an incentive to publish them as quickly as possible, in order to maximize her chances of being first. Recent work suggests that this applies even in the case of smaller, intermediate discoveries (Boyer [2014]; Strevens [2017]; Heesen [2017b]). All of this helps motivate scientists to share their work.

If peer review were to be abolished, the communist norm and the priority rule would still be in effect, so scientists would still be motivated to share their work as quickly as possible. However, the following change would occur.

In the absence of prepublication peer review, scientists would be able to share their discoveries more quickly. In the current system, peer review can hold up publication for significant amounts of time, especially in the case of fields with high rejection rates or long turnaround times. During this time, other scientists cannot build on the work and may spend their time needlessly duplicating the work. Cutting out this lag by letting scientists publish their own work when they think it is ready will speed up scientific progress. While being faster is not

always better (it may increase the risk of error, see Heesen [2018b]), in this case delays in publication are reduced without any reduction in the time spent on the scientific work itself.

To some extent this already occurs. Scientists, especially well-connected scientists, already share preprints that make the community aware of their work in advance of publication. For people who regularly do this, practically speaking little would change upon adopting the system we advocate. However, our proposal turns pre-journal-publication dispersal of work from a privilege of a well-connected few into the norm for everyone.

On this point, then, abolishing peer review is a net positive, as scientists will still be incentivized to share their work as soon as possible, but the delays associated with prepublication peer review are removed.

3.2 Time allocation

The current system restricts the way scientists are allowed to spend their time. For each article submitted to a journal, a number of scientists are conscripted into reviewing it, and at least one editor has to spend time on that article as well.

On our proposal, scientists would be free to choose how much of their time to spend reading and reviewing others' work as compared to other scientific activities. Some scientists would spend less time reviewing, some scientists would spend more, and some would spend exactly as much as under the current system.

For scientists in the latter category our proposal makes no difference, while for scientists in the other two categories our proposal represents a net improvement of how they spend their time, at least in their own judgements. We think people are the best judges of how to use their own time and labour. We thus trust scientists' decisions in these regards, and welcome changes that would render many scientists' choices about how to allocate their own labour independent of the preferences of the relatively small number of editorial gatekeepers.

So we assume that scientists are well placed to judge how best to use their own abilities to meet the community's epistemic needs. We claim, moreover, that the reward structure of science is set up so as to make it in their interest to do so: the credit economy incentivises scientists to spend their time on pursuits the epistemic value of which will be recognized by the

community (Zollman [2018]). Hence freeing up the way scientists allocate their time leads to net epistemic benefits to the scientific community.

One might object that journals perform a useful epistemic sorting role, telling scientists what is worth spending their time on. We will address these concerns in Section 4.1.

One might think that this would lead scientists to spend significantly less time reading and reviewing others' work. If this is right, we still think it would be an overall improvement for the reasons mentioned above. But we also want to point out that this is not as obvious a consequence as it may seem. Here are two reasons to expect scientists to spend as much time or more reading and reviewing on our proposal. First, for many scientists reading and reviewing are intrinsically valuable and can help their own research. Second, the current system provides no particular incentive to read and review either: scientists agree to review only because they independently want to or because they feel an obligation to the research community. While no one scientist is conscripted, at the group level editors are going to keep going until they find someone. This can amount to picking whomever is most weak-willed or under some extra-epistemic social pressure. It is not obvious that this way of deciding who does the reviewing has much to recommend it. Any rewards that exist for reviewing will still exist on our proposal, and may be amplified by the possibility of making postpublication reviews public.

3.3 Gender skew in publications

Male scientists publish more, on average, than female scientists, a phenomenon known as the productivity puzzle or productivity gap (Zuckerman and Cole [1975]; Valian [1999]; Prpić [2002]; Etzkowitz *et al.* [2008]). Several explanations have been suggested, none of which are entirely satisfactory; see especially (Etzkowitz *et al.* [2008], pp. 409–12). Two of these explanations that are relevant to our concerns here are the direct effects of gender bias and the indirect effects of the expectation of gender bias.

There is some evidence of gender bias in peer review, although this is not unambiguous, see (Lee *et al.* [2013], pp. 7–8; Lee [2016], and references therein). Insofar as there is gender bias—in the sense of women's work being judged more negatively by peer reviewers—abolishing peer review will remove this and help level the playing field for men and women.

We expect positive epistemic consequences from the removal of these arbitrarily different standards.

While the evidence of gender bias in peer review is not entirely clear-cut, there is good evidence that women expect to face gender bias in peer review, see (Lee [2016]; Bright [2017a]; Hengel [unpublished], and references therein). In an effort to overcome this perceived bias, women tend to hold their own work to higher standards. Hengel ([unpublished]) provides evidence that women spend more time correcting stylistic aspects of their article during peer review, presumably due to higher expectations of scrutiny on such apparently superficial elements of their work. On the plausible assumption that if women have higher standards for each article they will produce fewer articles overall, this means that the mere expectation of gender bias can contribute to the productivity gap.

After abolishing peer review both women and men will hold their work primarily to their own individual standards of quality, and secondarily to their expectations of the response of the entire scientific community, but not to their expectations of the opinion of a small arbitrary group of gatekeepers. We do not know whether this will lead the women to behave more like the men (producing more articles) or the men to behave more like the women (holding individual articles to a higher standard of quality). However, in line with our view above that scientists are well-placed to judge how best to spend their own time, we take it that any resulting change in behaviour will be a net epistemic positive.

3.4 Library resources

Journal subscription fees currently take up a large amount of library resources (RIN [2008]; Van Noorden [2013]). To summarize some key figures from the 2008 report: Research libraries in the UK spent between £208,000 and £1,386,000 on journal subscriptions annually (and that was a decade ago, with subscriptions having risen substantially since). The cost for publishing and distributing an article was estimated to be about £4000, or about £6.4 billion per year in total. Savings from moving to author-paid open access were estimated at £561 million, about half of which would directly benefit libraries.

On our proposal, publication happens at the initiative of authors on one or more online

archives of scientific publications. Insofar as journals continue to exist, they will create issues consisting of previously published articles. Since these are already publicly available, they will not be in a position to charge for access, so all journals will be open access (we expect these to be run mainly by academic societies rather than for-profit publishers, and to be fewer in number).

The costs of scientific publishing will then primarily be in maintaining the archive(s) of publications. The example of existing large preprint archives such as arXiv, bioRxiv, and PhilSci-archive suggests that this can be done at a fraction of the cost currently spent on journal subscription fees. As a rough guideline, Van Noorden ([2013]) estimates maintenance costs of arXiv at just \$10 per article. So our proposal involves significant savings on library resources, which could be used to expand collections, retain more or better-trained staff, or other purposes that would be of epistemic benefit to the scientific community.

Two additional effects should be considered in relation to this. First, the fact that the online archive will be open access means that scientific publications will be available to everyone, not just to those with a library subscription or some other form of access to for-profit scientific journals. This will yield epistemic benefits by improving the research capabilities of independent scholars and those at universities with fewer resources.

Second, the fact that any value added by for-profit journals would be taken away. The two tasks currently carried out by journals that could plausibly be supposed to add value to scientific publications are peer review and copy-editing (Van Noorden [2013]). It is the purpose of all other sections of this article to argue that peer review does not in fact (provably) add value, so we set that aside. This leaves copy-editing. We propose that libraries use some of the funds freed up from journal subscriptions to employ some copy-editors. Each university library would make copy-editing services available to the scientists employed at that university. We contend that after paying for the maintenance of an online archive and a team of copy-editors, under our proposal libraries would still end up with more resources for other pursuits than under the current system.

We note that this particular advantage of our proposal is a bit more historically contingent than the others. There seems to be no particular reason why prepublication peer review has to

be implemented through for-profit journals, and if the open access movement has its way we might be able to free up most of these library resources without abolishing prepublication peer review. But our proposal also achieves this goal, and so we count it as an advantage relative to the system as it is currently actually implemented.

3.5 Scientific careers

The ‘publish or perish’ culture in science has been widely noted (Fanelli [2010]). Universities judge the research productivity of scientists through their publications in (peer reviewed) journals, with some focusing more on ‘quantity’ (counting publications) and others on ‘quality’ (publishing in prestigious journals). Scientific journals and the system of prepublication peer review thus play an important role in shaping scientific careers. What will become of this if peer review is abolished?

We note first that the ‘publish or perish’ culture is a subset of a larger system that we discussed above: the credit economy. Publishing in a journal is one way to receive credit for one’s work, but there are others, most prominently citations and awards. Scientific careers depend on all of these, with different institutions weighing quantity of publications, quality of publications, citation metrics, and awards and other honours differently.

Any of these types of credit represents some kind of recognition of the scholarly contributions of the scientist by her peers. But here we distinguish two types of credit, which we will call short-run credit and long-run credit. Getting an article through peer review yields a certain amount of credit: more for more prestigious journals, less for less prestigious ones. But this is short-run credit in the following sense: The editor and peer reviewers judge the technical adequacy and the potential impact of the article, as well as various other factors, shortly after it is written. Their judgement is (among other things) a prediction of how much uptake the article is likely to receive in the scientific community.

In contrast, citations (as well as awards, prizes, and inclusion in anthologies or textbooks) represent long-run credit. They actually constitute the uptake the article receives in the scientific community. Long-run credit is both a more considered opinion of the scientific importance of the article and a more democratic one (citations can be made by anyone, and awards usually

reflect a consensus in the scientific community, whereas peer review is normally done by up to three individuals). So long-run credit reflects a more direct and better estimate of the real epistemic value of a contribution to science.

So what would the effect of our proposal be? For better or worse, our proposal does not make it impossible for universities to use metrics to judge research productivity. While journal rankings and impact factors would disappear, citation metrics for individual scientists and articles would still be available. This may mean that universities stop judging their scientists based on the impact factors of the journals they publish in and start judging them on the actual citation impact of their articles. More generally, our proposal will decrease or remove the role of short-run credit in shaping career outcomes and increase the role of long-run credit, which we take to be a better measure of scientific importance. So we think this is an improvement on the status quo.

One may want to criticize citation metrics as a measure of quality (Lindsey [1989]; Gruber [2014]; Belter [2015]; Heesen [2017a]). We are sympathetic, and to be clear our aim here is not to defend citation metrics. Our general point here is about preferring long-run to short-run credit. Our more specific point is that conditional on citation metrics being used at all, it is better to look at an individual scientist's actual citations rather than the average number of citations to the journals she has published in (which is what impact factors are).

What about junior hires and related career decisions, where long-run credit may be absent or minimal? If abolishing peer review means completely getting rid of journals and the associated prestige rankings, this robs hiring departments of some information regarding the scientific importance of candidates' work. If this means those on the hiring side need to read and form an opinion of candidates' work for themselves, we do not think that is a bad thing. This would of course take time, but if journals and peer review are completely abolished, that just means the time spent reviewing the article is transferred to the people considering hiring the scientist, which again, we do not think is a bad thing. In fact, since very few academics are on a hiring committee year after year, whereas referee requests are a constant feature while one is in the community, we think that even this added burden when hiring might still be a net time-saver for academics.

But it does not have to be that way. We never said journals and peer review have to be completely abolished—our proposal in Section 2 explicitly suggests journal issues may still appear, but as curated collections of articles based on postpublication peer review. So short-run credit based on journal prestige need not disappear. It need not even be slower as there is no particular reason postpublication peer review needs to take longer than prepublication peer review. But there is the added advantage that the article is already published while it undergoes peer review, so the wider community outside the assigned reviewers also has a chance to respond before it is included in a journal.

3.6 The power of gatekeepers

The discussion immediately above touched on another effect, one that we think is worth bringing out as a benefit of our proposal in its own right. As mentioned our proposal suggests that in evaluating the importance of scientific work we decrease our reliance on short-run credit (journal prestige), with a corresponding increase in long-run credit (citations, among other things). This means that the overall credit associated with a particular article depends less on the judgments made by an editor and a small number of reviewers, and more on its actual uptake in the larger scientific community.

Editors in particular currently play a large role in determining which scientific work is worthy of attention, as they are a relatively small group of people with a deciding vote in the peer review process of a large number of articles. They are often referred to as gatekeepers for this reason (Crane [1967]). Our proposal entails significantly decreasing both the prevalence and importance of this role. By replacing some of this importance with long-run credit, which comes from the scientific community as a whole, it makes the evaluation of scientific work a more democratic process. Not only is there some reason to think that democratic evaluation of scientific claims is more in line with general communal norms accepted within science (Bright *et al.* [2018]), but general arguments from democratic theory and social epistemology of science give epistemic reason to welcome the increased independence of judgement and evaluation this would introduce (List and Goodin [2001]; Heesen *et al.* [forthcoming]; Perović *et al.* [2016], pp. 103–4).

4 Where Peer Review Makes No Difference

In this section we consider a number of aspects of the scientific incentive structure for which we think a case can be made that abolishing peer review will leave them basically unaffected. This serves partially to forestall objections to our proposal that we anticipate from defenders of the peer review system, and partially to avoid overstating our case—in some of what follows we argue that abolishing peer review will likely have no effect in cases where one might have expected it to be beneficial.

4.1 Epistemic sorting

Given the stated purpose of peer review mentioned in Section 2 the first and most apparent disadvantage of our proposal is that it would remove the epistemic filter on what enters into the scientific literature. One might worry that the scientific community would lose the ability to maintain its own epistemic standards, and thus the general quality of scientific research would be reduced. We argue here that despite the intuitive support this idea might have, the present state of the literature on scientific peer review does not support it.

Separate out two kind of epistemic standards one may hope that the peer review system maintains. First, that peer review allows us to identify especially meritorious work and place it in high profile journals, while ensuring that especially shoddy work is kept from being published. Call this the ‘epistemic sorting’ function of peer review. Second, that peer review allows for the early detection of fraudulent work or work that otherwise involves research misconduct. Call this the ‘malpractice detection’ function of peer review. We deal with each of these in turn.

Let us step back and ask why, from the point of view of epistemic consequentialism, one would want peer review to do any sort of epistemic sorting. We take the answer to be that epistemic sorting helps scientists fruitfully direct their time and energy by selecting the best work and bringing it to scientists’ attention through publication in journals. They read and respond to that which is most likely to help them advance knowledge in their field.

How could peer review achieve this? One might hope that peer review functions by keeping bad manuscripts out of the published literature and letting good manuscripts in. This, however, is a non-starter. There are far too many journals publishing far too many things, with standards

of publication varying far too wildly between them, for the sheer fact of having passed peer review somewhere to be all that informative as to the quality of a manuscript.

Instead, if peer review is to serve anything like this purpose it must be because reviewers are able (even if imperfectly) to discern the relative degree of scientific merit of a work, and sort it into an appropriately prestigious journal. Epistemic sorting happens not via the binary act of granting or withholding publication, but rather through sorting manuscripts into journals located on a prestige hierarchy that tracks scientific merit (in a related article we argue that postpublication peer review would do this better than prepublication peer review, see Arvan *et al.* [unpublished]).

Our first critique is that the available evidence does not support the idea that the prestige hierarchy of journals tracks some underlying quality of articles. Articles published in high-ranking journals are retracted more, exaggerate effect sizes to a greater extent, and do not appear to be more methodologically sound than articles in low-ranking journals (Brembs *et al.* [2013]; Brembs [2018]). In short, an empirically measurable notion of epistemic merit that can support the idea that journals effectively sort articles by merit is yet to be found.

Our second critique challenges a necessary condition for epistemic sorting to work as advertised, namely that reviewers be reliable guides to the merit of the scientific work they review. Investigation into reviewing practices has not generally found much inter-reviewer reliability in their evaluations (Peters and Ceci [1982]; Ernst *et al.* [1993]; Lee *et al.* [2013], pp. 5–6). What this means is that one generally cannot predict what one reviewer will think of a manuscript by seeing what another reviewer thought. If there was some underlying epistemic merit scientists were accurately (even if falteringly) discerning by means of their reviews, one would expect there to be correlations in reviewers evaluations. However, this is not what we find. Indeed, one study of a top medical journal even found that ‘reviewers [...] agreed on the disposition of manuscripts at a rate barely exceeding what would be expected by chance’ (Kravitz *et al.* [2010], p. 3). Findings like these are typical in the literature that looks at inter-reviewer reliability (for a review of the literature see Bornmann [2011], p. 207).

The best that can be said about this is that it is possible such persistent disagreement represents reasonable disagreement among reviewers about how to weigh or apply multiple stan-

dards of scientific evaluation (Lee [2012]). But while this is possible, we know of no direct evidence that low inter-reviewer reliability is in fact caused by this. And in any case the existence of such widespread disagreement in the scientific community would further undermine the claim that epistemic sorting is a desirable feature of the present system of prepublication peer review. After all, if it turns out we tend to disagree with each other about how to apply or weigh scientific virtues, why should the fact that an article is agreed to be good by the standards of a small group of reviewers be good evidence that it is worth spending time on according to the standards of a potential reader?

Our third critique of the epistemic sorting idea speaks more directly to the ideal it tracks. We are not persuaded that the best way to direct scientists' attention is to continually alert them to the best pieces of individual work, and have them proportion their attention according to position on a prestige hierarchy. We take it the intuition behind this is a broadly meritocratic one. This intuition has been challenged by some modelling work (Zollman [2009]). While Zollman retained some role for peer review, his model still found that striving to select the best work for publication is not necessarily best from the perspective of an epistemic community; his model favoured a greater degree of randomisation.

We do not wish to rest our case on the results of one model that in any case does not fully align with our argument, but it highlights that the ideal of meritocracy stands in need of more defence than it is typically given. We take it that scientists most fruitfully direct their attention to that package of previous work and results that, when combined, provides them with the sort of information and perspectives they need to best advance their own epistemically valuable projects. It is a presently undefended assumption that this package of work should be composed of works that are themselves individually the most meritorious work, or that paying attention to the prestige hierarchy of journals and proportioning one's attention accordingly will be useful in constructing such a package. Hence, even if it did turn out that the peer review system could sort according to scientific merit, it is an under-appreciated but important fact that this is not the end of the argument. Further defence of the purpose of this kind of epistemic sorting is needed from the point of view of epistemic consequentialism.

Before moving on we note a potential objection. Even if one did not think that peer review

was detecting some underlying quality or interestingness, one might think that the process of feedback and revision that forms part of the peer review system would be beneficial to the epistemic quality of the scientific literature. In this way epistemic sorting may have a positive epistemic effect even if it fails in its primary task.

However, this returns us to the points regarding gatekeepers and time allocation from Section 3. We are not opposed to scientists reading each other's work, offering feedback, and updating their work in light of that. This can indeed lead to improvements (Bornmann [2011], p. 203), though in this context it is worth noting the results of an experiment in the biomedical sciences, which found that attempting to attach the allure of greater prestige to more epistemically high calibre work did little to actually improve the quality of published literature (Lee [2013]). Fully interpreting these results would require discussion of the measures of quality used in such literature. We will not do that here, since we will not dispute the point that it is desirable for scientists to give feedback and respond to it.

We would expect this sort of peer-to-peer feedback to continue under a system without prepublication peer review. Curiosity, informal networking, collegial responsibilities, and the credit incentives to engage with others' work and make use of new knowledge before others do; these would all be retained even without prepublication peer review. What would be eliminated is the assignment of reviewing duties to articles that scientists did not independently decide were worth their time and attention, and the necessity of giving uptake to criticism (in order to publish) independently of an author's own assessment of the value of that feedback.

We thus conclude that, from the point of view of epistemic consequentialism, there is presently little reason to believe that a loss of the epistemic sorting function of prepublication peer review would be a loss to science. Inclusion in the literature does not do much to vouch for the quality of an article; the evidence does not favour the hypothesis that reviewers are selecting for some latent epistemic quality in order to sort into appropriate journals; and the ideal underlying the claimed benefits of epistemic sorting is dubious. While peer reviewers do give potentially valuable feedback, there is no particular reason to think that changes in how scientists decide to spend their time would make things worse in this regard, and (per our arguments in Section 3) some reason to think that they would make things better.

4.2 Malpractice detection

The other way peer review might uphold epistemic standards is through malpractice detection. However, once again, the literature does not support this. A number of prominent cases of fraudulent research managed to sail through peer review. Upon investigation into the behaviour of those involved it was found there was no reason to think that peer reviewers or editors were especially negligent in their duties (Grant [2002], p. 3). Peer reviewers report unwillingness to challenge something as fraudulent even where they have some suspicion, and avoid the charge (Francis [1989], pp. 11–2). A criminologist who looked into fraudulent behaviour in science reported that ‘virtually no fraudulent procedures have been detected by referees because reading a paper is neither a replication nor a lie-detecting device’ (Ben-Yehuda [1986], p. 6). A more recent survey of the evidence found, at the least, no consistent pattern in journals’ self-reported ability to detect and weed out fraudulent results (Anderson *et al.* [2013], p. 235).

Even if the prospect of peer review puts some people off committing fraud, the fact that it is so unreliable at detecting fraud suggests that this is a very fragile deterrence system indeed. Even this psychological deterrence would be rapidly undermined by more adventurous souls, or those pushed by desperation, since many would quickly learn that prepublication peer review is a paper tiger.

Conversely, there are various ways for malpractice detection to operate in the absence of peer review. These include motive modification (Nosek *et al.* [2012]; Bright [2017b]), encouraging postpublication replication and scrutiny (Bruner [2013]; Romero [2017]), and the sterner inculcation of the norms of science coupled with greater expectation of oversight among coworkers (Braxton [1990]). All of these methods of deterring fraud or meliorating its effects would still be available under our proposal.

What evidence we now have gives little reason to suppose that abolishing prepublication peer review is any great loss to malpractice detection. Thus in this regard our proposal would make no great difference to the epistemic health of science. Combining this with the discussion of epistemic sorting, we conclude there is presently no reason to believe prepublication peer review is adding much value to science by upholding epistemic standards.

4.3 Herding behaviour

Where above we argued that prepublication peer review is not making a positive difference often claimed for it, in this section we downplay a potential benefit of our proposal. A consistent worry about scientific behaviour is that it is subject to fads or, in any case, some sort of undesirable herding behaviour, see (Chargaff [1976]; Abrahamson [2009]; Strevens [2013]). A natural thought is that prepublication peer review encourages this, since by its nature it means that to get new ideas out there one must convince one's peers that the work is impressive and interesting (at least with typical journals, PLoS ONE being a notable exception in asking reviewers to check only for methodological soundness). It has thus been claimed that prepublication peer review encourages unambitious within-paradigm work that unduly limits the range of scientific activity (Francis [1989], p. 12). Reducing the incentive to herd might thus be claimed as a potential benefit of our proposal. However, we are not convinced that it is prepublication peer review that is doing the harmful work here.

As mentioned above, our proposal eliminates or significantly reduces the importance of short-run credit, the credit that accrues to one in virtue of publishing in a (more or less prestigious) scientific journal. Long-run credit, on the other hand, is left untouched. Under any sort of credit system, a scientist needs to do work that the community will pay attention to, build upon, and recognize her for. The mere fact that (she believes that) her peers are interested in a topic and liable to respond to it is thus still positive reason to adopt a topic. This is true even if the scientist would not judge that topic to be the best use of her time if she were (hypothetically) free from the social pressures and constraints of the scientific credit system.

The best that could be said about our proposal in this regard is that scientists would not specifically have to pass a jury of peers before getting their work out there. But given that we anticipate continued competition for the attention of scientific coworkers, it is hard to say what the net effect in encouraging more experimental or less conformist scientific work would be.

Whatever conformist effects the credit incentive has (see also the discussion immediately below) do not depend on whether it is short- or long-run credit one seeks. The conformism comes from the fact that credit incentives focus scientists' attention on the predicted reaction of their fellow scientists to their work. Prepublication peer review might make this fact especially

salient by bringing manuscripts before a jury of peers before they may be entered into the literature. But even without prepublication peer review the credit-seeking scientist must be focused on her peers' opinions. So there is no particular reason to think that removing the prepublication scrutiny of manuscripts will free scientists from their own anticipations of the fads and fashions of their day.

4.4 Long-run credit

We end this section by noting that many of the effects of the credit economy of science studied by social epistemologists really concern long-run credit rather than the short-run credit affected by retaining or eliminating prepublication peer review. This point is not restricted to herding behaviour.

For instance, social epistemologists have studied both the incentive to collaborate, and various iniquities that can arise when scientists do not start with equal power when deciding who shall do what work and how they shall be credited (Harding [1995]; Boyer-Kassem and Imbert [2015]; Bruner and O'Connor [2017]; O'Connor and Bruner [2019]; Rubin and O'Connor [2018]). Whether or not manuscripts would have to pass prepublication peer review in order to enter the scientific literature, there would still be benefits in the long run to collaboration, and (alas) there would still be social inequalities that allow for iniquities to manifest in the scientific prestige hierarchy.

For another example, social epistemologists have studied the ways that the credit incentive encourages different strategies for developing a research profile or molding one's scientific personality to be more or less risk-taking (Weisberg and Muldoon [2009]; Alexander *et al.* [2015]; Thoma [2015]; Heesen [forthcoming]). Once again, prepublication peer review plays no particular role in the analysis. The incentives to differentiate oneself from one's peers (without straying too far from the beaten path) and to mould one's personality accordingly exist independently of prepublication peer review.

Two especially influential streams of work in the social epistemology of science have been the study of the division of cognitive labour (Kitcher [1990]; Strevens [2003]), and the role of credit in providing a spur to work in situations with a risk of under-production (Dasgupta and

David [1994]; Stephan [1996]). These two streams have directed the focus of the field, and have formed some of the chief defences of the credit economy of science as it now stands, but see (Zollman [2018]) for a more critical take.

We mention them here because prepublication peer review or short-run credit again plays no particular role in the analyses offered by these articles. What drives their results is scientists' expectation that genuine scientific achievement will be recognized with credit. As we have argued above, it is long-run credit that best tracks genuine scientific achievement, and so it is long-run rather than short-run credit that grounds scientists' expectation in this regard. So in social epistemologists' most prominent defences of the credit economy of science, long-run credit (while not named such) is the mechanism underlying the claimed epistemic benefits of the credit economy.

5 Difficulties for Our Proposal

We have discussed some benefits that would predictably accrue from abolishing peer review and some ways that its apparent benefits are either under-evidenced or better attributed to the effects of long-run credit, which our proposal leaves untouched. We now discuss some cases that we take to be more problematic for our proposal—but by this point we hope to have at least convinced the reader that prepublication peer review rests on shakier theoretical grounds than its widespread acceptance may lead one to suppose.

5.1 A guarantee for outsiders

One purpose prepublication peer review serves is providing a guarantee to interested but non-expert parties. Science journalists, policy makers, scientists from outside the field the manuscript is aimed at, or interested non-scientists can take the fact that something has passed peer review as a stamp of approval from the field. At a minimum, peer review guarantees that outsiders are focusing on work that has convinced at least one relatively disinterested expert that the manuscript is worthy of public viewing. Given that there are real dangers to irresponsible science journalism or public action that is seen to be based on science that is not itself trustworthy (Bright [2018], Section 4), and that it is hard for non-experts to make the relevant

judgement calls themselves, having a social mechanism to provide this kind of guarantee for outsiders is useful.

It is difficult to predict in advance what norms would come to exist for science journalists in the absence of prepublication peer review. We thus first and foremost call for empirical research on this issue, possibly by studying what has happened in parts of mathematics and physics that already operate broadly along the lines we suggest (Gowers [2017]).

However, against the presumption that things would be worse, we have two points to make. As the recent replication crisis has made clear, the value of peer review as a stamp of approval should not be overstated. There are reasons to doubt that peer review reliably succeeds in filtering out false results. We give three of them. First, peer reviewers face difficulties in actually assessing manuscripts—and just about anything can pass peer review eventually—as discussed under the heading of ‘epistemic sorting’ in Section 4.1. Second, there are problems with the standards we presently use to evaluate manuscripts, in particular with the infamous threshold for statistical significance used in many fields (Ioannidis [2005]; Benjamin *et al.* [2018]). And third, deeper features of the incentive structure of science make replicability problems endemic (Smaldino and McElreath [2016]; Heesen [2018b]). Using peer review as a stamp of approval may just be generating expert overconfidence (Angner [2006]), without the epistemic benefits of greater reliability that would back this confidence up. In fact, known strategies for manipulating the public into believing false claims exploit the prestige given to (potentially false or misleading) claims that have passed through peer review (Oreskes and Conway [2010]; Holman and Bruner [2017]; Weatherall *et al.* [forthcoming]).

For the second part of our reply, recall that it is only prepublication peer review that we seek to eliminate. We do not object to postpublication peer review resulting in articles being selected for inclusion in journals that mark the community’s approval of such work, ideally after due and broad-based evaluation. If some such system were implemented then outsiders could use inclusion in such a journal as their marker of whether work is soundly grounded in the relevant science.

If such a stamp of approval from a journal or other communally recognized institution only comes a number of months or years after something is first published then we would expect

it to represent a more well-considered judgement. Note that this would not necessarily slow the diffusion of knowledge as under the present system the same article would have spent time hidden from view going through prepublication peer review. The end result might not even be all that different from what happens in the present system, except that postpublication peer review would take into account more of the response or uptake from the wider scientific community. Thus it would more closely approximate the considered judgement of the community, as ultimately reflected in the long-run credit accorded to the article.

5.2 A runaway Matthew effect

The second problem we are less confident we can deal with is that of exacerbating the Matthew effect. This is the phenomenon, first identified by Merton ([1973d]), of antecedently more famous authors being credited more for work done simultaneously or collaboratively, even if the relative size or skill of their contribution does not warrant a larger share of the reward.

Arguably the present system helps put a damper on the Matthew effect, allowing a junior or less prestigious author to secure attention for their work by publishing in a high profile journal. Without such a mechanism to grab the attention of the field, perhaps scientists would just decide what to pay attention to based on their prior knowledge of the author or recommendation from others. This would strengthen the effects of networks of patronage and prestige bias favouring fancy universities, thus squandering valuable opportunities to learn from those who were not initially lucky in securing a prestigious position or patronage from the establishment.

A related worry notes that our proposal gives greater weight to long-run credit (and less to short-run credit) than prepublication peer review. Insofar as long-run credit is particularly vulnerable to the Matthew effect (due to citation cascades resulting at least partially from prestige, for example), it again looks like our proposal makes the Matthew effect worse. Moreover, on this version of the worry the Matthew effect threatens to undermine the advantage of long-run over short-run credit we argued for in Section 3.5.

While some have defended the Matthew effect (Strevens [2006]), we will not go that route in defending our proposal for two reasons. First, the Matthew effect can perpetuate iniquities that themselves harm the generation and dissemination of knowledge (Bruner and O'Connor

[2017]). Second, even if it could be justified at the level of individual publications, its long-term effects are epistemically harmful. The scientific community allocates the resources necessary for future work on the basis of its recognition of past performance. So if there is excess reward for some and unfair passing over of others at the present stage of inquiry, this will ramify through to future rounds of inquiry, misallocating resources to people whose accomplishments do not fully justify their renown (Heesen [2017a]). Hence on grounds of epistemic consequentialism we take seriously the problem of a runaway Matthew effect.

As mentioned, due to the pressures of credit-seeking and their own curiosity, scientists would still have incentive to read others' work and adapt it to suit their own projects. There is always a chance that valuable knowledge may be gathered from the work of one who has been ignored, which could provide an innovative edge. To some extent this creates opportunities for arbitrage: if the Matthew effect ever became especially severe there would be a credit incentive to specialize in seeking out the work of scientists who are not getting much attention. The lesson here is that the Matthew effect can only ever be so severe, before the credit incentive starts providing counter-veiling motivations.

However, this does not fully solve our problem. Moreover, so long as resource allocation is tied to recognition of past performance the differences in recognition generated by the Matthew effect can and often do become self-fulfilling prophecies, as those with more gain the resources to do better in the future, and those without are starved of the resources necessary to show their worth.

It is not clear where to go from here. From the above it may seem like a solution would be to pair our proposal with a call to loosen the connection between recognition of a scientist's greatness based on their past performance and resource allocation. Indeed, this may well be independently motivated (Avin [forthcoming]; Heesen [2017a], Section 6). However, even short of this far-reaching change, we feel at present that this matter deserves more study rather than any definitive course of action.

Another suggestion would be to implement reforms specifically aimed at counteracting the Matthew effect, that is, at giving more attention to the work of junior scholars and those at lesser-known institutions. We might have a role similar to current journal editors, which con-

sists in actively soliciting reviews for newly published articles. Like prepublication peer review, this system would guarantee a minimum number of readers for each article. This would serve to address that part of the worry that suggests that some articles would not be read at all under our proposal.

This leaves the worry that attention will still be more skewed towards antecedently famous scholars overall. Our present thought is that this is a very speculative objection, and there is no empirical evidence to back up the claim that eliminating prepublication peer review will have dire consequences in this regard. In particular, while the present system may (rarely) allow a relative outsider to make a big splash, the common accusation of prestige bias in peer review (Lee *et al.* [2013], p. 7) suggests that on the whole prepublication peer review may contribute to the Matthew effect rather than curtailing it. If this gives famous scientists more opportunities to publish articles, then our alternative system may provide welcome relief, since it allows more people to get their articles out there.

Here we simply acknowledge our present paucity of evidence regarding different institutional arrangements' effects on long and short-run credit. There is probably prestige bias affecting both what gets published and what receives long-run uptake through discussion, citations, and follow-up research. But we know of no evidence that addresses whether having a prestigious affiliation (say) provides a greater relative benefit during peer review or after. Thus whether shifting focus to long-run credit would exacerbate the Matthew effect is an open empirical question. So, while we grant that a runaway Matthew effect may occur under our system, we prefer to stress that at this point it is just not known whether the Matthew effect will be worse with or without prepublication peer review.

What we propose is a large change, involving freeing up a lot of time and opening it up to more self-direction on the part of scientists, and it is not clear what sort of institutional changes it would be paired with. With more study of epistemic mechanisms designed especially to promote the work of junior or less prestigious scientists there might be found some way of surmounting the problem of a runaway Matthew effect, should it arise. Ultimately, only empirical evidence can settle these questions. Given the clear benefits and the unclear downsides of our proposal, we hope at minimum to inspire a more experimental attitude towards peer review.

6 Conclusion

Prepublication peer review is an enormous sink of scientists' time, effort, and resources. Adopting the perspective of epistemic consequentialism and reviewing the literature on the philosophy, sociology, and social epistemology of science, we have argued that we can be confident that there would be benefits from eliminating this system, but have no strong reasons to think there will be disadvantages. There is hence a kind of weak dominance or Pareto argument in favour of our proposal.

To simplify things, imagine forming a decision matrix, with rows corresponding to 'Keeping prepublication peer review' and 'Eliminating prepublication peer review'. The columns would each be labelled with an issue studied by science scholars that we have surveyed here: gender bias in the literature, speed of dissemination of knowledge, efficient allocation of scientists' time and attention, and so on. For each column, if there is a clear reason to think that either keeping or eliminating prepublication scientific peer review does better according to the standards of epistemic consequentialism, place a one in the row of that option, and a zero in the other. If there is no reason to favour either according to present evidence, put a zero in both rows.

Our present argument could then be summarized with: As it stands, the only ones in such a table would appear in the row for eliminating prepublication peer review. We thus advocate eliminating prepublication peer review. Journals could still exist as a forum for recognizing and promoting work that the community as a whole perceives as especially meritorious and wishes to recommend to outsiders. Scientists would still have every reason to read, respond to, and consider the work of their peers; prepublication peer review is not the primary drive behind either the intellect's curiosity or the will's desire for recognition, and either of those suffice to motivate such behaviours.

The overall moral to be drawn mirrors that of our invocation of the importance of long-run over short-run credit. The best guarantor of the long run epistemic health of science is science: the organic engagement with each others' ideas and work that arises from scientists deciding for themselves how to allocate their cognitive labour, and doing the hard work of replicating and considering from new angles those ideas that have been opened up to the scrutiny of the

community. All this would continue without prepublication peer review, and the best you can say for the system that currently uses up so much of our time and resources is that it often fails to get in the way.

Acknowledgements

Both authors contributed equally. We would like to thank Justin Bruner, Adrian Currie, Cailin O'Connor, Jan-Willem Romeijn, Kevin Zollman, two reviewers for this journal, and our audience at the Philosophy of Science Association meeting in Seattle for valuable comments. This research was supported by the Netherlands Organisation for Scientific Research (016.Veni.195.141 to Remco Heesen); the National Science Foundation (SES 1254291 to Liam Kofi Bright); the Leverhulme Trust (Early Career Fellowship to Remco Heesen); the Isaac Newton Trust (Early Career Fellowship matching funding to Remco Heesen).

Remco Heesen

Department of Philosophy

University of Western Australia

Crawley, WA, Australia

and

Faculty of Philosophy

University of Groningen

Groningen, The Netherlands

remco.heesen@uwa.edu.au

Liam Kofi Bright

Department of Philosophy, Logic and Scientific Method

London School of Economics and Political Science

London, UK

liamkbright@gmail.com

References

- Abrahamson, E. [2009]: ‘Necessary Conditions for the Study of Fads and Fashions in Science’, *Scandinavian Journal of Management*, **25**, pp. 235–9.
- Alexander, J. M., Himmelreich, J. and Thompson, C. [2015]: ‘Epistemic Landscapes, Optimal Search, and the Division of Cognitive Labor’, *Philosophy of Science*, **82**, pp. 424–53.
- Anderson, M. S., Ronning, E. A., De Vries, R. and Martinson, B. C. [2010]: ‘Extending the Mertonian Norms: Scientists’ Subscription to Norms of Research’, *The Journal of Higher Education*, **81**, pp. 366–93.
- Anderson, M. S., Shaw, M. A., Steneck, N. H., Konkle, E. and Kamata, T. [2013]: ‘Research Integrity and Misconduct in the Academic Profession’, in M. B. Paulsen (ed.), *Higher Education: Handbook of Theory and Research*, Vol. 28, Dordrecht: Springer, pp. 217–61.
- Angner, E. [2006]: ‘Economists as Experts: Overconfidence in Theory and Practice’, *Journal of Economic Methodology*, **13**, pp. 1–24.
- Arvan, M., Bright, L. K. and Heesen, R. [unpublished]: ‘Jury Theorems for Peer Review’, available at <philsci-archive.pitt.edu/15931/>.
- Atkinson, M. [1994]: ‘Regulation of Science by “Peer Review”’, *Studies in History and Philosophy of Science, Part A*, **25**, pp. 147–58.
- Avin, S. [forthcoming]: ‘Centralised Funding and Epistemic Exploration’, *British Journal for the Philosophy of Science*, available at <dx.doi.org/10.1093/bjps/axx059>.
- Belter, C. W. [2015]: ‘Bibliometric Indicators: Opportunities and Limits’, *Journal of the Medical Library Association*, **103**, pp. 219–21.
- Ben-Yehuda, N. [1986]: ‘Deviance in Science: Towards the Criminology of Science’, *British Journal of Criminology*, **26**, pp. 1–27.

- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C. *et al.* [2018]: ‘Redefine Statistical Significance’, *Nature Human Behaviour*, **2**, pp. 6–10.
- Bornmann, L. [2011]: ‘Scientific Peer Review’, *Annual Review of Information Science and Technology*, **45**, pp. 197–245.
- Boyer, T. [2014]: ‘Is a Bird in the Hand Worth Two in the Bush? Or, Whether Scientists Should Publish Intermediate Results’, *Synthese*, **191**, pp. 17–35.
- Boyer-Kassem, T. and Imbert, C. [2015]: ‘Scientific Collaboration: Do Two Heads Need to Be More than Twice Better than One?’, *Philosophy of Science*, **82**, pp. 667–88.
- Braxton, J. M. [1990]: ‘Deviance from the Norms of Science: A Test of Control Theory’, *Research in Higher Education*, **31**, pp. 461–76.
- Brembs, B. [2018]: ‘Prestigious Science Journals Struggle to Reach Even Average Reliability’, *Frontiers in Human Neuroscience*, **12**, pp. 1–7.
- Brembs, B., Button, K. and Munafò, M. [2013]: ‘Deep Impact: Unintended Consequences of Journal Rank’, *Frontiers in Human Neuroscience*, **7**, pp. 1–12.
- Bright, L. K. [2017a]: ‘Decision Theoretic Model of the Productivity Gap’, *Erkenntnis*, **82**, pp. 421–42.
- Bright, L. K. [2017b]: ‘On Fraud’, *Philosophical Studies*, **174**, pp. 291–310.
- Bright, L. K. [2018]: ‘Du Bois’ Democratic Defence of the Value Free Ideal’, *Synthese*, **195**, pp. 2227–45.
- Bright, L. K., Dang, H. and Heesen, R. [2018]: ‘A Role for Judgment Aggregation in Coauthoring Scientific Papers’, *Erkenntnis*, **83**, pp. 231–52.
- Bruner, J. and O’Connor, C. [2017]: ‘Power, Bargaining, and Collaboration’, in T. Boyer-Kassem, C. Mayo-Wilson and M. Weisberg (eds), *Scientific Collaboration and Collective Knowledge*, Oxford: Oxford University Press, pp. 135–57.

- Bruner, J. P. [2013]: 'Policing Epistemic Communities', *Episteme*, **10**, pp. 403–16.
- Chargaff, E. [1976]: 'Triviality in Science: A Brief Meditation on Fashions', *Perspectives in Biology and Medicine*, **19**, pp. 324–33.
- Crane, D. [1967]: 'The Gatekeepers of Science: Some Factors Affecting the Selection of Articles for Scientific Journals', *The American Sociologist*, **2**, pp. 195–201.
- Dasgupta, P. and David, P. A. [1994]: 'Toward a New Economics of Science', *Research Policy*, **23**, pp. 487–521.
- Eisenhart, M. [2002]: 'The Paradox of Peer Review: Admitting too Much or Allowing too Little?', *Research in Science Education*, **32**, pp. 241–55.
- Ernst, E., Saradeth, T. and Resch, K. L. [1993]: 'Drawbacks of Peer Review', *Nature*, **363**, p. 296.
- Etzkowitz, H., Fuchs, S., Gupta, N., Kemelgor, C. and Ranga, M. [2008]: 'The Coming Gender Revolution in Science', in E. J. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman (eds), *The Handbook of Science and Technology Studies*, Cambridge, MA: MIT Press, pp. 403–28.
- Fanelli, D. [2010]: 'Do Pressures to Publish Increase Scientists' Bias? An Empirical Support from US States Data', *PLoS ONE*, **5**, p. e10271.
- Francis, J. R. [1989]: 'The Credibility and Legitimation of Science: A Loss of Faith in the Scientific Narrative', *Accountability in Research: Policies and Quality Assurance*, **1**, pp. 5–22.
- Gibson, T. A. [2007]: 'Post-publication Review Could Aid Skills and Quality', *Nature*, **448**, p. 408.
- Goldman, A. I. [1999]: *Knowledge in a Social World*, Oxford: Oxford University Press.
- Gowers, T. [2017]: 'The End of an Error?', *The Times Literary Supplement*, available at <www.the-tls.co.uk/articles/public/the-end-of-an-error-peer-review/>.
- Grant, P. M. [2002]: 'Scientific Credit and Credibility', *Nature Materials*, **1**, pp. 139–41.

- Gruber, T. [2014]: ‘Academic Sell-out: How an Obsession with Metrics and Rankings Is Damaging Academia’, *Journal of Marketing for Higher Education*, **24**, pp. 165–77.
- Harding, S. [1995]: ‘“Strong Objectivity”: A Response to the New Objectivity Question’, *Synthese*, **104**, pp. 331–49.
- Heesen, R. [2017a]: ‘Academic Superstars: Competent or Lucky?’, *Synthese*, **194**, pp. 4499–518.
- Heesen, R. [2017b]: ‘Communism and the Incentive to Share in Science’, *Philosophy of Science*, **84**, pp. 698–716.
- Heesen, R. [2018a]: ‘When Journal Editors Play Favorites’, *Philosophical Studies*, **175**, pp. 831–58.
- Heesen, R. [2018b]: ‘Why the Reward Structure of Science Makes Reproducibility Problems Inevitable’, *The Journal of Philosophy*, **115**, pp. 661–74.
- Heesen, R. [forthcoming]: ‘The Credit Incentive to Be a Maverick’, *Studies in History and Philosophy of Science*, available at <[dx.doi.org/10.1016/j.shpsa.2018.11.007](https://doi.org/10.1016/j.shpsa.2018.11.007)>.
- Heesen, R., Bright, L. K. and Zucker, A. [forthcoming]: ‘Vindicating Methodological Triangulation’, *Synthese*, available at <[dx.doi.org/10.1007/s11229-016-1294-7](https://doi.org/10.1007/s11229-016-1294-7)>.
- Hengel, E. [unpublished]: ‘Publishing While Female: Are Women Held to Higher Standards? Evidence from Peer Review’, available at <www.erinhengel.com/research/publishing_female.pdf>.
- Holman, B. and Bruner, J. [2017]: ‘Experimentation by Industrial Selection’, *Philosophy of Science*, **84**, pp. 1008–19.
- Hull, D. L. [1988]: *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*, Chicago, IL: University of Chicago Press.
- Ioannidis, J. P. A. [2005]: ‘Why Most Published Research Findings Are False’, *PLoS Medicine*, **2**, p. e124.

- Jukola, S. [2017]: 'A Social Epistemological Inquiry into Biases in Journal Peer Review', *Perspectives on Science*, **25**, pp. 124–48.
- Katzav, J. and Vaesen, K. [2017]: 'Pluralism and Peer Review in Philosophy', *Philosophers' Imprint*, **17**, pp. 1–20.
- Kitcher, P. [1990]: 'The Division of Cognitive Labor', *The Journal of Philosophy*, **87**, pp. 5–22.
- Kravitz, R. L., Franks, P., Feldman, M. D., Gerrity, M., Byrne, C. and Tierney, W. M. [2010]: 'Editorial Peer Reviewers' Recommendations at a General Medical Journal: Are They Reliable and Do Editors Care?', *PLoS ONE*, **5**, p. e10072.
- Kriegeskorte, N., Walther, A. and Deca, D. [2012]: 'An Emerging Consensus for Open Evaluation: 18 Visions for the Future of Scientific Publishing', *Frontiers in Computational Neuroscience*, **6**, pp. 1–5.
- Latour, B. and Woolgar, S. [1986]: *Laboratory Life: The Construction of Scientific Facts*, Princeton, NJ: Princeton University Press.
- Lee, C. J. [2012]: 'A Kuhnian Critique of Psychometric Research on Peer Review', *Philosophy of Science*, **79**, pp. 859–70.
- Lee, C. J. [2013]: 'The Limited Effectiveness of Prestige as an Intervention on the Health of Medical Journal Publications', *Episteme*, **10**, pp. 387–402.
- Lee, C. J. [2016]: 'Revisiting Current Causes of Women's Underrepresentation in Science', in J. Saul and M. Brownstein (eds), *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*, Oxford: Oxford University Press, pp. 265–82.
- Lee, C. J., Sugimoto, C. R., Zhang, G. and Cronin, B. [2013]: 'Bias in Peer Review', *Journal of the American Society for Information Science and Technology*, **64**, pp. 2–17.
- Lindsey, D. [1989]: 'Using Citation Counts as a Measure of Quality in Science: Measuring What's Measurable Rather than What's Valid', *Scientometrics*, **15**, pp. 189–203.

- List, C. and Goodin, R. E. [2001]: ‘Epistemic Democracy: Generalizing the Condorcet Jury Theorem’, *Journal of Political Philosophy*, **9**, pp. 277–306.
- Longino, H. E. [1990]: *Science as Social Knowledge*, Princeton, NJ: Princeton University Press.
- Louis, K. S., Jones, L. M. and Campbell, E. G. [2002]: ‘Macroscope: Sharing in Science’, *American Scientist*, **90**, pp. 304–7.
- Macfarlane, B. and Cheng, M. [2008]: ‘Communism, Universalism and Disinterestedness: Re-examining Contemporary Support among Academics for Merton’s Scientific Norms’, *Journal of Academic Ethics*, **6**, pp. 67–78.
- Merton, R. K. [1973a]: ‘A Note on Science and Democracy’, in his *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago, IL: University of Chicago Press, Chapter 13.
- Merton, R. K. [1973b]: ‘Priorities in Scientific Discovery: A Chapter in the Sociology of Science’, in his *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago, IL: University of Chicago Press, Chapter 14.
- Merton, R. K. [1973c]: ‘Behavior Patterns of Scientists’, in his *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago, IL: University of Chicago Press, Chapter 15.
- Merton, R. K. [1973d]: ‘The Matthew Effect in Science’, in his *The Sociology of Science: Theoretical and Empirical Investigations*, Chicago, IL: University of Chicago Press, Chapter 20.
- Nosek, B. A. and Bar-Anan, Y. [2012]: ‘Scientific Utopia: I. Opening Scientific Communication’, *Psychological Inquiry*, **23**, pp. 217–43.
- Nosek, B. A., Spies, J. R. and Motyl, M. [2012]: ‘Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability’, *Perspectives on Psychological Science*, **7**, pp. 615–31.

- O'Connor, C. and Bruner, J. [2019]: 'Dynamics and Diversity in Epistemic Communities', *Erkenntnis*, **84**, pp. 101–19.
- Oreskes, N. and Conway, E. M. [2010]: *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*, New York: Bloomsbury Press.
- Perović, S., Radovanović, S., Sikimić, V. and Berber, A. [2016]: 'Optimal Research Team Composition: Data Envelopment Analysis of Fermilab Experiments', *Scientometrics*, **108**, pp. 83–111.
- Peters, D. P. and Ceci, S. J. [1982]: 'Peer-Review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again', *Behavioral and Brain Sciences*, **5**, pp. 187–95.
- Prpić, K. [2002]: 'Gender and Productivity Differentials in Science', *Scientometrics*, **55**, pp. 27–58.
- RIN [2008]: 'Activities, Costs and Funding Flows in the Scholarly Communications System in the UK', Technical Report for Cambridge Economic Policy Associates on behalf of the Research Information Network, available at <www.rin.ac.uk/system/files/attachments/Activites-costs-flows-report.pdf>.
- Romero, F. [2017]: 'Novelty versus Replicability: Virtues and Vices in the Reward System of Science', *Philosophy of Science*, **84**, pp. 1031–43.
- Rubin, H. and O'Connor, C. [2018]: 'Discrimination and Collaboration in Science', *Philosophy of Science*, **85**, pp. 380–402.
- Saul, J. [2013]: 'Implicit Bias, Stereotype Threat, and Women in Philosophy', in K. Hutchison and F. Jenkins (eds), *Women in Philosophy: What Needs to Change?*, Oxford: Oxford University Press, pp. 39–60.
- Shatz, D. [2004]: *Peer Review: A Critical Inquiry*, Lanham, MD: Rowman and Littlefield.
- Smaldino, P. E. and McElreath, R. [2016]: 'The Natural Selection of Bad Science', *Royal Society Open Science*, **3**, p. 160384.

- Smith, R. [2006]: ‘Peer Review: A Flawed Process at the Heart of Science and Journals’, *Journal of the Royal Society of Medicine*, **99**, pp. 178–82.
- Stephan, P. E. [1996]: ‘The Economics of Science’, *Journal of Economic Literature*, **34**, pp. 1199–235.
- Strevens, M. [2003]: ‘The Role of the Priority Rule in Science’, *The Journal of Philosophy*, **100**, pp. 55–79.
- Strevens, M. [2006]: ‘The Role of the Matthew Effect in Science’, *Studies in History and Philosophy of Science, Part A*, **37**, pp. 159–70.
- Strevens, M. [2013]: ‘Herding and the Quest for Credit’, *Journal of Economic Methodology*, **20**, pp. 19–34.
- Strevens, M. [2017]: ‘Scientific Sharing: Communism and the Social Contract’, in T. Boyer-Kassem, C. Mayo-Wilson and M. Weisberg (eds), *Scientific Collaboration and Collective Knowledge*, Oxford: Oxford University Press, pp. 3–33.
- Teixeira da Silva, J. A. and Dobránszki, J. [2015]: ‘Problems with Traditional Science Publishing and Finding a Wider Niche for Post-Publication Peer Review’, *Accountability in Research*, **22**, pp. 22–40.
- Thoma, J. [2015]: ‘The Epistemic Division of Labor Revisited’, *Philosophy of Science*, **82**, pp. 454–72.
- Valian, V. [1999]: *Why So Slow? The Advancement of Women*, Cambridge, MA: MIT Press.
- Van Noorden, R. [2013]: ‘The True Cost of Science Publishing’, *Nature*, **495**, pp. 426–9.
- Weatherall, J. O., O’Connor, C. and Bruner, J. P. [forthcoming]: ‘How to Beat Science and Influence People: Policy Makers and Propaganda in Epistemic Networks’, *British Journal for the Philosophy of Science*, available at <dx.doi.org/10.1093/bjps/axy062>.
- Weisberg, M. and Muldoon, R. [2009]: ‘Epistemic Landscapes and the Division of Cognitive Labor’, *Philosophy of Science*, **76**, pp. 225–52.

Zollman, K. J. S. [2009]: 'Optimal Publishing Strategies', *Episteme*, **6**, pp. 185–99.

Zollman, K. J. S. [2018]: 'The Credit Economy and the Economic Rationality of Science', *The Journal of Philosophy*, **115**, pp. 5–33.

Zuckerman, H. and Cole, J. R. [1975]: 'Women in American Science', *Minerva*, **13**, pp. 82–102.

BJPS Final Draft