

Opening the Black Box of Peer Review

More transparency in the peer review process will help researchers to study peer review and improve its quality and fairness.

By **Simine Vazire**

After working for about a decade as a journal editor, I'm concerned that peer review is not the safeguard against bad science that we think it is. I'm convinced that peer review could be a lot better and that increased transparency in the process is key to improving it.

Peer review is an obvious improvement over earlier forms of research evaluation, which involved semiformal groups of senior scholars making decisions about which papers deserved publication. These decisions were partly based on authors' status and reputation—the quintessential old boys' network. It is not hard to imagine how this process might have seemed an

acceptable form of quality control when science as an intellectual pursuit was a privilege for an elite. But as scientific research became more widely accessible, scholars needed more rigorous ways to distinguish good science from bad science. Such a need led to peer review as we know it today—a more formalized process that includes some precautions against bias and corruption [1].

But how well does our modern peer review system work? Does it keep bias, misconduct, and errors at bay? Alas, we don't have enough data to answer these questions. As scientists, we frequently advocate for evidence-based practice, but we don't always practice what we preach when it comes to keeping our own houses in order.

Despite a vibrant and growing field of research on peer review, focused mainly on medical journals, peer review is still mostly a black box. As a journal editor, I often found myself going through the steps of peer review as if going through a ritual—without stopping to ask myself if the review system as a whole achieves its fundamental goal: vetting thoroughly and fairly every paper under scrutiny.

Both scientists and the public have great confidence in peer review, considering it one of the primary mechanisms by which science vets itself [2]. But seeing peer review from the inside has shaken my confidence. In my view, there aren't sufficient guarantees that each paper is thoroughly vetted—often, there is no vetting at all for the most basic qualities you'd expect a quality control system to check, such as computational reproducibility for data-based papers. There is evidence that many errors, even quite obvious ones, make it through peer



Credit: APS/Carin Cain

review [3]. Moreover, several studies have demonstrated that status bias exerts a significant influence on reviewer evaluations [4]. The impact of other types of bias—based on the gender, geographic location, or ethnic origin of a paper’s authors—still needs to be properly assessed [5]. Finally, there have been several proven cases of editorial misconduct, whereby editors were found to have manipulated the content of reviews or changed reviewers’ recommendations. I believe that the mechanisms for holding editors accountable in these cases are generally unsatisfactory [6].

There are thus sufficient reasons to question the validity and limitations of peer review, and it’s clear that we need to gather more evidence. But the opaqueness of peer review—which makes it ripe for error and bias—also makes it hard to study.

There are a few ways to open peer review to scrutiny. First, scientists themselves should push for more transparency. For instance, referee reports and decision letters should be published, which could be done without sacrificing reviewer anonymity—an approach that **some publishers are starting to explore**. By making it clear to editors and reviewers that what they write could be made public, we may deter some of the most problematic practices. To enable this approach, we need to establish clear norms, so that authors do not fear that sharing review material would violate journal policies. Further, journals can partner with peer review researchers, granting them access to some confidential review components and working together on experiments designed to better understand the mechanisms that influence peer review [7]. These experiments will be key to gathering the evidence needed to support large-scale peer review reforms.

We already know enough, however, to implement some changes to peer review now. For instance, journals should immediately act to diversify the pool of reviewers they use, as diversity is vital to robust criticism [8]. Today’s technology, including artificial intelligence, could tremendously facilitate this pool expansion. Moreover, the ability to tap into the specific expertise of multiple referees opens the door to specialized reviewers that could inspect selected aspects of a paper, such as statistical analyses. Reviewers and editors can also harness new

technology to create better platforms and tools for carrying out peer review, including a facilitator-supported discussion method known as structured deliberation [9], journal-club-style discussion groups, or the incorporation of rubrics and checklists. Web-based tools and communication platforms offer innovation opportunities that shouldn’t be ignored.

But to improve peer review, we must be ready to challenge ourselves and recognize that the status quo in peer review can stand in the way of the core scientific values of fairness and accountability.

Simine Vazire: Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia

REFERENCES

1. M. Baldwin, “In referees we trust?” *Phys. Today* **70**, No. 2, 44 (2017).
2. N. Oreskes, *Why trust science?* (Princeton University Press, Princeton, New Jersey, 2019).
3. S. Vazire and A. O. Holcombe, “Where are the self-correcting mechanisms in science?” *PsyArXiv Preprints* (2020).
4. G. Bravo *et al.*, “Hidden connections: Network effects on editorial decisions in four computer science journals,” *J. Informetr.* **12**, 101 (2018); C. Le Goues *et al.*, “Effectiveness of anonymization in double-blind review,” *Commun. ACM* **61**, 30 (2018); A. Tomkins *et al.*, “Reviewer bias in single- versus double-blind peer review,” *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12708 (2017).
5. H. Bastian, *The fractured logic of blinded peer review in journals*.
6. C. O’Grady, “Delete offensive language? Change recommendations? Some editors say it’s OK to alter peer reviews,” *Science* (2020); The Black Goat podcast, “You Took the Words Right Out of My Mouth”.
7. C. K. Soderberg *et al.*, “Initial evidence of research quality of registered reports compared with the standard publishing model,” *Nat. Hum. Behav.* **5**, 990 (2021).
8. H. Longino, *Science as social knowledge* (Princeton University Press, Princeton, New Jersey, 1990).
9. B. Wintle *et al.*, “Predicting and reasoning about replicability using structured groups,” *MetaArXiv Preprints* (2021).