Is peer review ergodic?

The strengths and limitations of peer review have long been documented. The concept of ergodicity from statistical physics may shine a new light on them.

The limitations of peer review have long been the topic of debate among scientists and editors alike¹. Nevertheless, greater awareness of issues such as the reproducibility crisis², the unhealthy pressure (real or perceived) on early career researchers to demonstrate stellar publication records³, and the variety of existing approaches to academic assessment in different disciplines⁴ has latterly appeared to make the subject more urgent. Moreover, as the institutional drive towards open access has shifted the publishing orthodoxy away from its bias to reject submissions⁵, the role of peer review has subtly changed: a consensus seems to be emerging around the notion that there should be less focus on subjective criteria such as immediacy of interest or breadth of appeal, and more emphasis on objective matters such as technical validity and reproducibility.

To many physicists, this picture comes across as somehow incomplete. But for a few notable exceptions, the reproducibility crisis mostly appears to affect the life sciences. Although they are subjected to intense pressure to publish, the sense is that young physicists have so far been spared the worst abuses of impact factor mania. And most importantly, physics has arXiv. For over a quarter of a century now⁶, physicists have become accustomed to freely disseminating their manuscripts on the preprint server, while simultaneously submitting them to traditional journals that oversee their peer review.

This decoupling between the dissemination and evaluation of manuscripts has had far-reaching consequences: first, it satisfies a thirst for the immediate distribution of information — a thirst that has gone unquenched in many other areas of research until relatively recently, in spite (or, some might argue, because) of it being a clear benefit of the Internet that established publishers have been unable to make the most of. Second, it demonstrates that a scientific community can be self-policing: the mere act of signing a manuscript publicly provides the incentive not to compromise on scientific rigour.

Nevertheless, these developments have not led to the disintermediation of journals as one might have expected 20 years ago. While there exist community-specific variations in behaviour, as well as a steady flurry of initiatives to upend the status quo, much of the physics community appears to accept the more ‘conservative’ view that journals help us to separate the signal from the noise, as it were. For the time being at least, journals continue to act as the stamps of approval signalling the technical validity and editorial quality of a physics paper.

Of course, this equilibrium is underpinned by the way in which resources — mostly in the form of funding and academic positions — are allocated. Across most scientifically developed nations, the paradigm to emerge over the past few decades has been to reward academic excellence. In practice, this means exposing the research enterprise to competitive forces: since the resources are finite, only the ‘best’ projects and academics can be funded. Enough ink has been spilled describing the benefits and corrosive shortcomings of this system. The key point is that in order to implement it, academia has in effect outsourced part of its quality control process to journals.

With the benefit of hindsight, it is generally easy to agree on the influence of a particular scientific result or line of work. The problem for funding agencies and tenure committees is that they do not have the benefit of that time: they must make a decision based on a snapshot of criteria, advice and impressions that will

vary by institution in their rigour, speed and transparency.

Journal editors face a similar problem: once they select a submission as being potentially interesting, they typically enlist a small panel of experts to assess the work, and comment on its merits and likely impact. If, on balance, the arguments in favour of publication win the day, the journal effectively bets that the ‘ensemble average’ over the reviewers’ and the editor’s judgement is equivalent to the judgement reached by integrating over time (see figure). Believe it or not, when it comes to scientific assessment, we use journals to save time!

Whether or not this ergodic hypothesis really holds is, of course, a legitimate point of contention. We are not being entirely serious by raising this comparison between peer review and a fundamental tenet of statistical mechanics. Nevertheless, it is an instructive analogy and, we believe, it is one that helps to highlight both the fundamental strengths and limitations of peer review. Be that as it may, from the countless anecdotes we hear of game-changing papers rejected and duds accepted: the ultimate peer reviewer is time. □

Published online: 4 July 2018
<https://doi.org/10.1038/s41567-018-0226-5>

References

1. Relman, A. S. & Agnelli, M. N. *Engl. J. Med.* **321**, 827–829 (1989).
2. Ioannidis, J. P. A. *PLoS Med.* **2**, e124 (2005).
3. Tregoning, J. *Nature* **558**, 345 (2018).
4. Callaway, E. *Nature* **539**, 343 (2016).
5. Horton, R. *Lancet* **385**, 1166 (2015).
6. *Nat. Phys.* **12**, 719 (2016).