EMBO reports

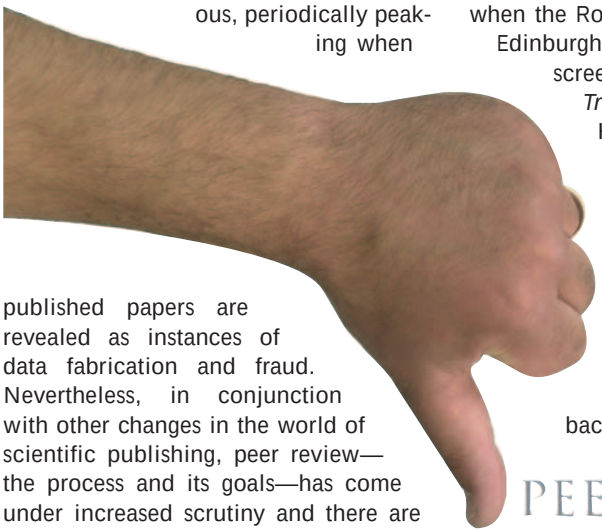
viewpoint

Through a glass darkly

The present and the future of editorial peer review

Les Grivell

Writing in *Nature* in 1977, Thomas Jukes remarked dryly, "it has been said that the warp that holds the complex fabric of science together is peer review, and the woof is the noise made by scientists who complain about it" (Jukes, 1977). In 30 years, it would seem that little has changed, except perhaps that the 'woofs' have become louder and more numerous, periodically peaking when published papers are revealed as instances of data fabrication and fraud. Nevertheless, in conjunction with other changes in the world of scientific publishing, peer review—the process and its goals—has come under increased scrutiny and there are some hints of potential changes.



Peer review is, of course, not restricted to manuscripts submitted for publication. The whole fabric of research funding, academic careers, and the futures of research groups, departments and whole institutes, rests on examination and assessment by 'peers'—experts in a given research area who are expected to give a critical and, despite the fact that they are often competitors, fair judgement of scientific performance. As such, reviews are frequently, and in large measure, based on scientists' track records—as reflected in their published work and its impact on a particular field. Editorial peer review, and its quality and fairness, therefore have a central role in all other assessments.

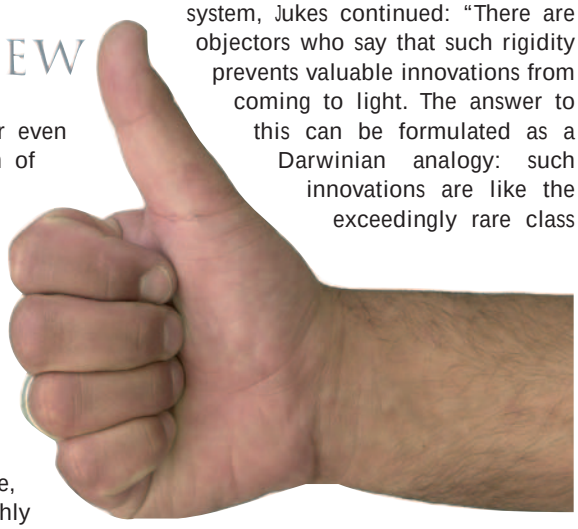
Although modern peer review probably dates back no earlier than the invention of the photocopier, the principle that any piece of work submitted for publication should be subject to prior inspection by a select group of experts probably originated in the mid-eighteenth century, when the Royal Societies of London and Edinburgh adopted this procedure for screening contributions to their *Transactions* (Kronick, 1990). However, as related by Ray Spier in his historical account of the peer-review process (Spier, 2002), the idea that anything potentially challenging the orthodoxy of the day should be subject to review and eventual censorship or suppression goes back further—possibly to the time of the Inquisition or even beyond. Those who complain of unfair treatment at the hands of reviewers would do well to remember that 400 years ago they would have had—at best—the choice between the burning of their manuscripts or of themselves.

PEER REVIEW

So, given that peer review has been accused of being "slow, expensive, profligate of academic time, highly subjective, prone to bias, easily abused, poor at detecting gross defects, and almost useless for detecting fraud" (Smith, 1997), what should it achieve and is it meeting its goals in reality?

> The whole fabric of research funding, academic careers, and the futures of research groups, departments and whole institutes, rests on examination and assessment by 'peers'…

The answer to the first question has generally been that peer review should be a filter to separate the wheat from the chaff, and thus to guarantee and improve the quality of published research. Jukes went further and stated that "science is essentially hierarchical; its progress and its integrity depend on the existence of an 'establishment', and on the rejection of uncontrolled or unrepeatable results" (Jukes, 1977). Evidently anticipating an outburst of protest from victims of the system, Jukes continued: "There are objectors who say that such rigidity prevents valuable innovations from coming to light. The answer to this can be formulated as a Darwinian analogy: such innovations are like the exceedingly rare class



©2006 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

of mutations that are beneficial to a species, and hence overcome all odds against their survival and spread." Alongside these basic criteria of 'correctness', most high-impact journals also ask reviewers to consider other aspects of scientific value, such as whether a manuscript presents a complete story, expands or challenges existing knowledge, creates new insights and/or opens up new research areas.

> …reviewers are increasingly swamped with manuscripts, and the better their reputation for critical and constructive review, the more likely they are to be overworked and to suffer fatigue

Are these goals being achieved? Apparently not, according to a series of independent studies first summarized in a report on peer review in biomedical research in 2003, and updated earlier this year (Jefferson *et al*, 2006). Unfortunately, most of these studies focus only on basic criteria for quality control, and examine the possible effects of blinding reviewers to author identity, of revealing reviewer identity, of reviewer training and feedback, and of peer review on the correctness of presentation and study validity. One such investigation, for example, examined several of these issues for manuscripts submitted to the *British Medical Journal*. Godlee and colleagues (1998) introduced a total of eight weaknesses in design, analysis and interpretation into a paper that had already been peer reviewed and accepted for publication. The modified paper was then sent to 420 reviewers, about 50% of whom responded. The mean number of weaknesses commented on was two, only 10% of the reviewers identified four or more weaknesses and 16% failed to identify any. There was no significant difference in performance between anonymous reviewers who were either aware of or blinded to authors' identities, and the corresponding groups who were required to sign their reviews. Interestingly, the study revealed that younger reviewers performed better than their older colleagues, reviewers who had published frequently tended to identify more errors than those who had published rarely, and reviewers whose past performance had been highly rated by editorial staff tended to perform better.

However, considering all studies as a whole, the report concludes that "There is little systematic, empirical evidence to support the use of editorial peer review as a mechanism to ensure quality of reports of biomedical research in biomedical journals. Practitioners of editorial peer review should recognize the lack of convincing empirical evidence of its effects and bear this in mind when making editorial decisions" (Jefferson *et al*, 2006). The report goes on to recommend a large research programme to investigate the effects of peer review, commenting that much of the research carried out so far has been limited to addressing marginal aspects relating to practitioners' expediency or specific interests. It also points out that attention should be paid to rejected manuscripts and the effects that peer review might have had on them. Given that many articles that are rejected by one journal are eventually published in another—sometimes after several cycles of submission and rejection—it is clear that editorial peer review constitutes an extremely costly process in both human and economic terms.

This last point, in my view, deserves further attention. Writing in his book *Little Science, Big Science…and Beyond*, Price (1986) quipped that: "we can say that 80–90% of all the scientists that have ever lived are alive now," pointing out that scientific activity and the number of scientists have increased exponentially during the past 300 years. As the number of scientists has increased, so has the number of journals and, with them, the number of publications. As a result, reviewers are increasingly swamped with manuscripts, and the better their reputation for critical and constructive review, the more likely they are to be overworked and to suffer fatigue. This phenomenon is not simply the consequence of an increasing number of scientists. It is compounded by the fact that—in Malthusian fashion—science funding is not keeping pace with demand. Shortage of funds and increased competition leads scientists to adopt 'salami-type' tactics when it comes to publishing their research (Lawrence, 2003), spreading their results over several papers rather than publishing one complete account of their work.

In addition, publications alone are not sufficient to guarantee funding or promotion. Although the present scientific generation is well aware that anything published on the web can be found with the click of a mouse, the same scientists will initially send their manuscripts to a high-profile high-impact journal. Their hopes are twofold: first, a 'gilt by association' effect of publishing in such journals; and second, that the exposure generated by any associated press release or other publicity will tip the balance in their favour when it comes to distributing funds. A small minority of researchers succeed in this approach. The remaining manuscripts enter a downward spiral of editorial or peer-review rejection. This ends when, in conformance to a sort of inverted 'Peter principle' or 'Dilbert principle'—according to which, managers are promoted to the position where they can create the least damage (Peter & Hull, 1969; Adams, 1996)—most rejected manuscripts eventually sink to a level that more accurately reflects their value to their respective research communities.

> …most readers … still require an assurance of quality for the material they find on the web, which journal branding provides to some extent

In entering this cycle, researchers display paradoxical behaviour. When acting as authors, they seek recognition from a wide readership; as readers, they are looking for some degree of validity conferred by journal branding. As critical readers, however, they acknowledge that publications in top journals might reflect the 'trendiness' of a particular area and the political skills of the authors, but that they are often oversimplified or compressed to an extent that severely limits their scientific value (Lawrence, 2003). Researchers might also accept that the number of citations received by a paper is essentially uncorrelated with the impact factor of the journal in which it is published (Seglen, 1997). Moreover, for many, if not all, high-impact journals, the distribution of paper citation rates is so highly skewed that they derive most of their citations from a handful of articles. For example, an analysis of papers published in *Nature* in 1999

©2006 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

demonstrated that only 16% of all papers accounted for half of the citations in 2001 (Colquhoun, 2003).

Is editorial peer review required at all, and are there viable and better alternatives to the current system? In short, is it still true that peer review is like Churchill's view of democracy: the worst of its type, apart from all the other forms that have been tried and tested? In the early days of electronic publishing, there were indeed great expectations that this new medium could bring about change. Judson (1994), pursuing ideas put forward by Lederberg (1991), saw the advent of electronic publishing on the web as a structural transformation that would replace the traditional editorial peer-review process by a dialectical mode of publication involving an open dialogue and collaboration among scientists, editors, expert commentators and readers. Harnad (1996), along similar lines, viewed the web as an environment that offers the possibility both to select referees on a broader and more systematic basis—perhaps through calls for review on bulletin boards—and for the post-publication of interactive reviews. Godlee (2000), looking forward to the launch of BioMed Central (BMC), saw electronic publication as free from the constraints enforced by the page budgets of print journals. Working on an understanding that much of what is published can safely be ignored, BMC journals should have a bias towards the publication of submitted manuscripts that pass basic tests of validity, regardless of their originality or relevance.

Neither the relaxed editorial guidelines of BMC nor any form of electronic online open review by the scientific community at large has yet found wide acceptance from authors, at least in biomedical research. As outlined above, most readers—especially those without close knowledge of a given research field—still require an assurance of quality for the material they find on the web, which journal branding provides to some extent. For members of the general public, who are often alerted by the media to instances of scientific fraud, peer review represents an essential step in the validation of research results. It should thus function correctly, and be seen to do so, if the aim is to preserve public confidence in scientists and their work.

Against this background, a few new journals now apply non-conventional models of peer review. The first of these, in the environmental sciences (Pöschl, 2004), uses an interesting two-tier system of editorial peer review: a closed and relatively rapid procedure that screens submissions on the basis of their suitability for publication on the website of *Atmospheric Chemistry and Physics Discussions*, followed by a longer public-review stage that, if successfully completed, allows a final version of the manuscript to be published in the journal *Atmospheric Chemistry and Physics*.

The second, more recent, initiative is *Biology Direct*, which is a new journal that aims to operate exclusively with open peer review. It puts the onus on authors to obtain reviews from members of the journal's editorial board. These reviews are published alongside the accepted articles, with the intention of increasing the responsibility of the reviewers and, at the same time, eliminating any source of abuse of their positions. A manuscript is considered to be rejected if no member of the editorial board agrees to review it, and authors are entitled to withdraw an article if they do not wish to see it published alongside its reviews.

However, there are intriguing questions as to what extent this review system will scale: whether, as in other open-review systems, reviewers will be able to withstand pressure from more senior authors, and whether it will influence authors' submission behaviour. On this last point, given that authors will be dealing directly and openly with their 'peers', they might be less inclined to submit papers that report incremental advances or oversell their work. If so, this could be regarded as a major achievement. If not, the system could end up being swamped and might suffer the same problems of reviewer overload and fatigue as the traditional model.

From the point of view of both authors and readers, some degree of quality control of scientific publication remains necessary if the scientific literature is to fulfil its primary functions as a public registry of claims of intellectual property ownership and as a record of scientific progress (Lederberg, 1991). However, the traditional review procedure is perceived to have shortcomings that include subjectivity and a proclivity for bias and abuse. Positive effects, although likely to be present at the level of individual manuscripts, have turned out to be difficult to demonstrate at an aggregate level. In addition, despite efforts by the editorial staff of many journals (Rørth, 2005) to protect their reviewers from overwork through careful triage and selection of manuscripts sent for in-depth review, fatigue is inevitable when many journals compete for the same reviewers. Change is thus desirable, but is difficult to achieve without corresponding changes in the behaviour of scientists as authors.

Change is thus desirable, but it is difficult to achieve without corresponding changes in the behaviour of scientists as authors

The first, and possibly most important, of these changes will be to discontinue the practice of aiming high in submission and of using peer review as a means to improve manuscripts as they move down the spiral. One strong inducement for authors to do this would be if granting agencies and other review bodies abandoned the use of journal impact factors to assess the scientific performance of individuals. In most cases, an alternative index, recently proposed by Hirsch (2005) and based exclusively on citations to an author's published work, is likely to be a fairer measure of contributions to the field. This simple calculation represents the relationship between the total number of published papers (N_p) and their citations. An individual has an index (h) if h of their published papers have at least h citations each and the other ($N_p - h$) papers have $\leq h$ citations each.

The second change might involve journals adopting models that are similar in spirit to those operated by *Biology Direct* and *Atmospheric Chemistry and Physics*. However, as these publications are relatively young and still have to prove their value, it might take some time before other journals follow suit. Those still dissatisfied with any system might find consolation in the comment from Jukes, which was quoted in the *Oxford English Dictionary* as part of its definition of peer review: "publishing a book is a way of avoiding peer review…"

©2006 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION

REFERENCES
Adams S (1996) *The Dilbert Principle: a Cubicle's-Eye View of Bosses, Meetings, Management Fads & Other Workplace Afflictions*. London, UK: Boxtree
Colquhoun D (2003) Challenging the tyranny of impact factors. *Nature* **423:** 479
Godlee F (2000) How will the peer review process change and improve? Freedom of Information Conference 2000. www.biomedcentral.com/meetings/2000/foi/transcripts/godlee
Godlee F, Gale CR, Martyn CN (1998) Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *JAMA* **280:** 237–240
Harnad S (1996) Implementing peer review on the net: scientific quality control in scholarly electronic journals. In Peek RP, Newby GB (eds) *Scholarly Publishing: The Electronic Frontier*, pp 103–108. Cambridge, MA, USA: MIT Press
Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* **102:** 16569–16572
Jefferson T, Rudin M, Brodney Folse S, Davidoff F (2006) Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database Methodology Rev* **1:** MR000016
Judson HF (1994) Structural transformations of the sciences and the end of peer review. *JAMA* **272:** 92–94
Jukes TH (1977) Peer review. *Nature* **265:** 203
Kronick DA (1990) Peer review in 18th-century scientific journalism. *JAMA* **263:** 1321–1322
Lawrence PA (2003) The politics of publication. *Nature* **422:** 259–261
Lederberg J (1991) Communication as the root of scientific progress. Paper presented at Science Editing in the Age of Global Communication, the Sixth Conference of the International Federation of Science Editors, Woods Hole, MA, USA, 15 Oct 1991
Peter LJ, Hull R (1969) *The Peter Principle*. London, UK: Souvenir
Pöschl U (2004) Interactive journal concept for improved scientific publishing and quality assurance. *Learned Pub* **17:** 105–113
Price DJ de Solla (1986) *Little Science, Big Science…and Beyond*. New York, NY, USA: Columbia University Press
Rørth P (2005) Authors, reviewers and editors at the *EMBO Journal*. *EMBO J* **24:** 3831–3883
Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. *BMJ* **314:** 498–502
Smith R (1997) Peer review: reform or revolution? *BMJ* **315:** 759–760
Spier R (2002) The history of the peer-review process. *Trends Biotechnol* **20:** 357–358

Les Grivell is the Manager of the Electronic Information Programme at the European Molecular Biology Organization in Heidelberg, Germany.
E-mail: grivell@embo.org

doi:10.1038/sj.embor.7400718

 ©2006 EUROPEAN MOLECULAR BIOLOGY ORGANIZATION