PERSPECTIVES IN Publishing

No.2 August 2004

Perspectives in Publishing is an occasional newsletter for journal editors. It aims to address generic issues in scholarly communication.

For general information about Elsevier, or any publishing related matter contact your Publishing Editor or visit the Elsevier website at:http://www.Elsevier.com/

For further information about this article or if you require further copies, please contact:

Adrian Mulligan
Elsevier
The Boulevard
Langford Lane
Kidlington Oxford
OX5 IGB

a.mulligan@elsevier.com

Is peer review in crisis?

A. Mulligan
Elsevier

Peer review is an essential component of scholarly publishing. In recent years it has attracted criticism and its role has been challenged. Based upon the findings of focus groups commissioned by Elsevier, this paper examines the role of peer review and discusses if, and how, it could be improved.

Review by peers has been a method of evaluation since Greek times (Barnes, 1981) and has been a formal part of scientific communication since the first scientific journals appeared over 300 years ago. The *Philosophical Transactions of the Royal Society* is widely accredited as being the first journal to formalise the process (Zuckerman & Merton, 1971 pp 68-69). The referee is now at the heart of scientific publishing and according to Ziman (1968, p111) is '…the lynchpin about which the whole business of Science is pivoted.'

It is testament to the power of peer review that a scientific hypothesis or statement, presented to the world is largely ignored by the scholarly community unless it is first published in a peer-reviewed journal. It is precisely because of this pivotal role as well as some notable incidents of fraud and incompetence that the process has been subjected to a variety of criticisms in recent years.

Possibly the most damaging incidence was in September 2002, when Jan Hendrik Schon, tipped to be a Nobel Prize winner, was discovered to have published a series of fraudulent papers. Subsequently, 16 papers he had published were withdrawn from *Nature, Science, Physical Review* and *Applied Physics Letters* (Lerner, 2003). Similarly fabrication has been found in the life sciences: the German molecular biologists, Fridhelm Herrmann and Marion Brach, were accused of inventing data in forty-seven papers published in a number of prestigious periodicals (Hauptman, 2002).

Unfortunately fraud is not the only issue, a study published on the effects of 'ecstasy' had to be retracted by Dr. George Ricaurte of Johns Hopkins University when it was realised that a more potent drug had been tested by mistake (Ricaurte, 2003).

These events have acted as a catalyst within the scholarly community, with many questioning the role of peer review. Recently, the Royal Society, Britain's most eminent academy, established a working group specifically to examine peer review.

So, after 300 years of scholarly publishing is it time to radically overhaul the peer review process? Qualitative research into peer review conducted by Elsevier, across a range of disciplines, consulting separate groups of authors, referees and editors, suggests not. Altogether 59 respondents, drawn from the USA, UK, and Germany, attended 6 focus groups with each group lasting 2-3 hours. Attendees had a range of experience, some were new to refereeing, editing or authoring, while others had years of experience. Those participating recognised that the process of review could be improved, but none believed that the fundamental precept of peer review should change. Sir Peter Lachmann, President of the Academy of Medical Sciences, and responsible for formulating proposals to deal with fraud agrees:

> Peer review is to science what democracy is to politics. It's not the most efficient mechanism, but it's the least corruptible. (Lachmann, Peter, 2002)

Fraud

What can the reviewer or editor do if they suspect fraud? In the past when suspicions have been strong enough the editor has requested that the research upon which the paper is based be repeated. This was done most famously when Nature published the work of Jacques Benveniste[1] (as quoted by Bhattacharya, Shaoni, 2003). Repeating the experiment is exceptional, and an expensive and impractical suggestion when considering the whole of scholarly publishing. It is also worth noting Nicholas Wade's comments, "There are plenty of honest reasons why two researchers may get different results from the same experiment." (Wade, 2003)

Editors who attended the focus groups agreed that recent developments, if adopted more widely, would make the process of publishing fraudulent work more difficult. Such mechanisms include vouchsafing the article, where the author is asked to sign a declaration stating the work submitted has not been falsified and is their own work. Further, a journal editor may insist that each of the corresponding authors or significant contributors sign a document stating that they agree to publication, including the head of the respective department (if their name is not already present). While unlikely to prevent the truly determined falsifier, formalisation of the institutional review will certainly make it more difficult.

> There's a strong case for viewing the prevention of fraud as the direct responsibility of the lab chief. The lab chief is in the best position to detect fraud. Only he can demand to see the lab notebooks, evidence that is beyond the reach of outsiders. (Wade, 2003).

Arguably, the ultimate recourse would be some form of legal action; however, editors who attended the focus groups agreed that the international nature of scholarly publication makes enforcement difficult and thus unlikely. In the absence of a legally binding framework some medical editors have collaborated to establish the Committee of Public Ethics (COPE)[2]. The group acts as a sounding board for Editors struggling to deal with possible breaches in research and publication ethics. COPE is attempting to put research misconduct firmly on the academic and educational agenda. Similarly, the World Association of Medical Editors (WAME)[3] seeks to provide support. Unfortunately, such groups have limited jurisdiction and ability to enforce any decisions they make.

The simplest course of action for the editor or reviewer who suspects fraud is to reject or recommend rejection of the paper. Unfortunately, the history of science is littered with incidences of original research papers, later proven to be correct, being rejected because they included unbelievable results or equivocal conclusions. The premise of scholarly communication is that research published is new and original. Often, only the author is in a position to truly say a piece of research is false.

1 Beneviste claimed that homeopathic water retained 'memory' of substances once dissolved in it. The peer review process agreed involved the experiments being repeated and observed by independent experts – the results could not be duplicated. Benveniste's reputation was left in ruins; he lost his funding and subsequently his position.

2 http://www.publicationethics.org.uk/

3 http://www.wame.org/

Editors agreed that by its very nature science is self-correcting. Fabricated or misinterpreted data will eventually be detected. Subsequent papers will attempt to build on the reported findings and discover the errors. This is the nature of scholarly publishing. It could be argued that what has changed is not the process of science or an increase in the number of cases of fraud, but the pressures on science, notably the public desire for instant and irrefutable solutions to complex problems.

The purpose of peer review is often misunderstood, not only by some scientists, but the media and public generally. Peer review cannot guarantee the correctness of the results. As Marc H. Brodsky, executive director of the American Institute of Physics, states "Referees cannot determine if data is falsified, nor are they expected to". Nor is it likely a referee will be in a position to spot an error as a consequence of a mislabelled drug, as in the case of the Ecstasy research retraction (Ricaurte, 2003). Moreover, scientists announcing their findings in the press prior to peer review have exacerbated the situation. An unsuspecting public is not always aware of the distinction between peer review and scientific proclamations in the media.

Peer review: purpose and practice

If peer review is not meant to detect fraud, what is its purpose? Unfortunately, there is no one text to which authors, reviewers or editors alike can refer. Peer review is an imprecise term and varies across disciplines: at its widest it is, 'the assessment by an expert of material submitted for publication.' (Carin Olson, 1990, p356-58)

The attendees of the focus groups broadly believe that peer review should prevent an author making egregious claims on minimal results, and where feasible, identify incidences of plagiarism. Furthermore, it should ensure that a consistent and appropriate methodology is used, and that recent reputable work in the area, upon which the article may be based, is correctly referenced and acknowledged.

While the focus groups attendees recognised fraud was likely to continue to capture media attention, they believed other more damaging issues were not. Referees felt that the overall burden of review is increasing; as more papers are being published, fewer reviewers are willing to referee. Authors were recognised by editors and referees alike as becoming increasingly demanding, requiring faster and faster review responses. Authors thought that bias and professional conflict was an issue in reviewing, that unscrupulous reviewers delayed publication or appropriated ideas. Editors felt that it was becoming increasingly difficult to attract and retain good referees; that authors were guilty of multiple submissions, fragmenting studies into 'minimum publishable units', and that plagiarism was increasing.

Improving Peer Review

So how can the process of peer review be improved? The solution in part may lie in changing some review practices, but all attending the focus groups agreed that timely, high quality reviews depended upon attracting, retaining and training good referees.

How best to do this depends upon fathoming the motivations and influences of referees. A number of key motivations emerged during the referee focus groups. All reviewers felt it was their academic duty to review, recognising that they expected someone to do the same for them. Many stated they were driven by a general interest in the area and a desire to be up-to-date with latest developments. Some thought it would help their own research or stimulate new ideas. For younger scientists being asked to review was perceived as an honour and confirmation of their standing in the community. Others hoped that reviewing would encourage the editor to view them more favourably.

Marginal motivations that emerged during the groups included a desire to build a relationship with the editor or being associated with a prestigious journal. A number enjoyed reading research prior to their colleagues. Finally, for some, it served as career development; participation demonstrated a wider interest, indicating initiative and commitment.

In terms of attracting referees, many made it very clear they only wished to review papers relevant to their area of expertise; moreover, a paper perceived as groundbreaking would guarantee review. All agreed that reviewing was time consuming especially if they were unfamiliar with the area; therefore they rejected papers outside their immediate sphere of knowledge.

While the reputation of a journal was a significant factor influencing a referee's decision to review, often his or her own experience on the journal was decisive. Referees are also authors; a poor experience as an author affects a referee's decision to review for a particular journal. The influence of the editor is also important; a referee is more likely to review if they personally know the person contacting them. The mode of communication is also significant, the majority accept e-mail as standard, but a number stated they would reject articles if they did not come as hard copy.

Obstacles identified by referees included difficulties in understanding papers that were badly written or from non-English speakers. Such papers invariably took longer to review and referees felt they should be rejected or improved prior to being sent to them. Editors participating in separate focus groups saw peremptory rejection as an issue. Editors stated they often see their role as mentor to the author, developing and improving the standard of a paper for final publication, a process some believed required a substantial amount of effort on behalf of both the editor and referee.

The role of the editor in evaluating a paper upon receipt is important. A study (Lavelle, 1966, p3-12) of 166 humanities journals revealed that only 9.6% of editors made an acceptance or rejection decision themselves. In contrast editorial staff on the Lancet evaluate all manuscripts. Richard Horton, Editor of The Lancet, said: "All submissions are carefully read and reviewed by a team of physicians and

scientists in-house before a decision is made about external peer review. At that point about 25% of submissions are sent out for further consideration, a vital step in our quality assurance process. No research paper is published in The Lancet which has not gone through expert and statistical review".

By way of a compromise, and in the absence of extensive editoral resources, it was suggested by some editors that poorly constructed papers should be rejected prior to external peer review and the author provided guidelines on how to write a research paper.

Furthermore, the author should be encouraged to resubmit the paper after revision. For papers where language is a substantial issue, it was agreed that the author should be referred to a specialised translation service or encouraged to have the paper moderated by a native English speaker.

The views of referees

According to the referees, encouraging participation in the review process rests upon fundamental criteria: sending only relevant papers of at least a minimum standard in quality, both in terms of content and language. Furthermore, referees indicated that communication is more effective if it is personalised, polite and efficient. In order to ensure such communication, one or two editors mentioned that they keep notes on any pertinent personal details, such as birthdays or recent holidays, which they then use in their communications.

In order to save time, a number of referees stated that they preferred to receive prior notice before receiving a full paper. In fact some stated that they would not review a paper if it arrived unsolicited. Referees indicated that notification, particularly in the form of an abstract would allow them to make a quick and informed decision. It negated the need to read the whole paper and also had the considerable advantage of saving the referee the cost of printing.

Editors recognised that sending advance notice to just two referees allows for a quick response, and importantly, the option for the editor to move onto another referee within a short period of time if an invitation is declined or ignored. Even more importantly, it reduces the overall number of reports produced by up to a third, diminishing the sense of burden among referees.

Retaining the services of the best referees also involves setting limits on how many times a referee will be requested to review, and communicating that limit to them. Many referees felt overburdened with requests and indicated that they invariably take longer to review if they believe there is a risk they will receive more requests.

A number of referees stated that they wished to see other referees' comments and receive feedback on the paper, not only in terms of the overall editorial decision, but also the criteria used. Some reviewers were upset to discover a paper was published when they had advised that it be rejected. In contrast others felt that there was a lack of editorial authority, too many referees' comments were taken at face value. It was suggested that feedback would aid both personal and professional development. Many wanted to see feedback on their own performance, perhaps an end of year review. Individuals who had received an acknowledgement or some form of feedback in the past implied they were more pre-disposed to review for that journal.

Referees did not consider incentives essential, but were nonetheless welcome; suggestions included book discounts, print cartridges and bottles of wine. Recognition emerged as a key covert factor, for many the review process is perceived as a 'chore and not a pleasure'. Reviewers feel this way because they are not rewarded or recognised for their work.

It was felt that many of the perceived problems afflicting the refereeing process could be remedied if the review was formally recognised. Such a system could take the form of accreditation to a journal, society or publisher. It might identify the number of times an individual reviews in different prestigious journals; such an indicator could in part be a measure of that individual's contribution to science. In the absence of such a system, limited recognition could be provided by a framed certificate for younger referees, an end of year acknowledgement, a listing of those who had reviewed published in the journal, or a formal letter to the referee or the head of the referee's department.

Referees raised a number of issues with the concept of referee payment, believing strongly that any financial reward would undermine objectivity. Further, even with a financial incentive some were unlikely to review more thoroughly, as both resources and time were limited.

Increasingly, publishers are becoming more active in the refereeing process through the facilities they offer. The development of online systems is perhaps the most significant contribution, potentially speeding up the process of refereeing. Online reviewing was perceived as advantageous for the author and editor rather than the reviewer. Nearly all referees accepted the inevitability of it, but felt they were shouldering an extra burden. They now had to print the manuscript, which is both time-consuming and expensive. Moreover, returning comments or recommendations electronically was seen as restrictive, some expressed a desire for flexibility of format.

Referees indicated it was more likely, that a specific and reasonable timeline (3 to 4 weeks was appropriate for most), in combination with some of the points already mentioned (e.g. limits on the number of reviews) would help speed up review. In addition, they indicated that they did not mind a polite reminder from the editor. In some incidences a letter from the editor was seen as a useful tool to help focus the referee on their task. Some referees stated that they chose not to commence their review until a reminder arrived.

Authors and referees alike believed that professional conflicts are increasingly common in peer review, particularly in an arena in which the progress of science is becoming more heavily dependent upon funding for its success. It was recognised that this most affected niche areas, where experts reviewing a paper may be competing for

the same funding as the author. A number of attendees recounted incidences of sabotaged papers; where the publication of research had been delayed in order to allow the referee's own research to be published first.

New referee practices

Some referees believed that the different peer review practices, such as double blind refereeing or open refereeing, if adopted globally, might alleviate some of the problems with unscrupulous referees. However, they also recognized that adopting these practices might create more problems than they resolve.

Double blind refereeing, whereby both referee and author remain anonymous, was considered 'ideal' as it served to avoid all potential bias, particularly for those authors who might be working in a developing country or be associated with an institute with an 'ordinary' reputation. It was also seen as relevant in incidences where 'prestigious' authors' papers are reviewed with more regard to the reputation of the author than the content of the research paper. However, it was questioned whether a paper could ever be truly blind, especially in 'niche areas'. Many claimed they could identify the author, through the style or subject matter of the paper, or more often through self-citation. A study by Justice et al (1998) supports the assertions of the referees. The study shows that blinding had little effect on the quality of review, was not successful in 32% of cases, and that well-known authors had been far more difficult to blind.

Open refereeing, where the referee and author are known to one another, was preferred by some respondents. They believed it prevented 'malicious' comments, stopped plagiarism, prevented referees drawing upon their own 'agenda', and encouraged honest open responses. However, the majority of referees believed it would achieve the opposite effect, and promote less open and less honest reports. Many felt that anonymity was key to reviewing in order to avoid 'politics'. A number said they would be less likely to be as forthright in their opinion if their comments were attributed. It was also believed that the problem of referees with an agenda would re-emerge, as a junior researcher reviewing an eminent scientist's work would be less likely to be honest for fear of affecting their own career or funding opportunities. Independent studies tend to support these comments. Open reviewing is practised on the 'British Medical Journal' and in a study conducted by the journal itself it recognised there was no discernable improvement in the quality of reviewing, and importantly it significantly increased the likelihood of reviewers declining to review (Susan Van Rooyen et al, 1999).

A slightly different version of open refereeing exists on the worldwide web; papers are made available for comments by readers, the article is subsequently revised and resubmitted. On balance, the majority of respondents attending the focus groups rejected it, not only for the reasons mentioned previously, but because respondents believed that research had to pass through a formal review process in order to prevent poor quality papers overshadowing those of real scientific value. Some referees questioned whether it would ever work, indicating that as a researcher, it was unlikely that they would look at work on the web that was not peer reviewed, unless it was extremely important to the field, much less comment upon the work themselves.

Editors suggested a simple and common sense solution to the problem of 'political' refereeing: always seek a third opinion when two reports contradict one another or when a report is unnecessarily delayed. Further, it was suggested that circulating referee reports to the other referees would be beneficial: not only would it satisfy the needs of referees who wish to develop professionally: it also has the advantage of eliminating those reviews with an agenda. Moreover, it was acknowledged that referee reports were likely to be tempered, once it was known they were circulated to other reviewers as standard practice.

However, for some editors other concerns were more pressing. They believe that plagiarised, re-published or multiple-submitted works, as well as research fragmented into 'minimum publishable units' are increasingly common. Detecting such practices in part lies with publishers. One approach might be to provide a utility that matches sections of submitted text with the full text in other articles. Another approach might be to provide a database to reviewers of all abstracts to articles in-print and published. However, it is worth noting that Mabe and Amin (2002) believe there is little evidence to support the theory of 'salami publishing'. Contrary, they show that the number of papers per author is dropping.

Improving peer review reports

Editors indicated there was a large degree of variability in the standards of referee reports. Some reports could be as short as one sentence; others ran to several pages, while in a few incidences a referee almost completely reworked the manuscript. Editors recognised that referees were often young, or up for tenure and may only recently have started reviewing. They need advice and like authors themselves, mentoring. It was generally accepted that the mechanism of peer review is largely unknown and hidden. Increasing confidence and understanding in the system, it was suggested, is based upon establishing a transparent framework.

Establishing such a clear framework involves clearly identifying to all the mechanics and expectations of peer review. A number of journals to varying degrees already provide guidelines, in a survey of 139 journals Weller (2002, p25) estimates that about half (51.8%) indicate their journal's position on anonymity of their authors and reviewers. In the same survey just over a third of editors informed the wider community of their rejection-rate, while only 17.3% gave detailed guidelines for reviewers. Transparency benefits readers, authors, reviewers and editors alike, "There appears to be a trend in this direction, with an obvious benefit to both journal readers and scholarly community as a whole" (Weller 2002, p27). Readers know the rigour of the review process, and thus are

able to make a judgement upon the standard of articles within the journal. Authors are able to make a more informed decision as to where to submit a paper. Reviewers will have a clearer idea of what is expected of them. The editor hopefully sees a greater number of submissions, finds referees more willing and will receive fewer queries regarding the refereeing process.

Ideally such transparency would include the following:

- A policy on anonymity.
- A policy statement on whether all articles are sent out for review.
- Information on how reviewers are selected.
- The number of reviewers typically involved.
- The expected length of the review process.
- The protocol a referee is expected to follow. (Criteria for publication evaluation)
- How the final decision is made.
- How the referee's remarks are communicated to the author.
- A statement on conflicts of interest. (e.g. referees who have collaborated with the author recently)
- Rejection rates.
- Feedback to referees.

So, is peer review in crisis? The simple answer is no, but that is not to say it cannot be improved. Enhancing peer review does not necessarily involve supplanting current processes, rather developing and supplementing existing practices. The attraction and retention of the best referees who will improve peer review by providing consistent, timely and quality reviews is key. Facilitating such quality reviews is based upon practices such as asking referees to review only relevant papers, setting limits on the number of times they will be asked to review. Sharing referee reports between referees will encourage constructive criticism. Informing the referee of the final decision, providing a personalised service, one that allows flexibility of response, giving the referee recognition will help develop the referee and make them feel valued.

For researchers peer review is the bedrock upon which scholarly publishing is based. However, for much of the wider public, its function is misunderstood. This is an issue, and the public need to be assured that scientific publishing is self-correcting and ultimately will detect fraud. While steps can be taken to avoid deception, peer review should not be a tool whose primary function is the categorical detection of fraud. Editors, referees and authors alike agree peer review aspires to improve the quality of scientific research published, and where feasible assure the correctness of the findings published. As Weller states:

> Both accidental and deliberate mistakes do happen, but that is not reason to scrap editorial peer review or to underestimate the tremendous importance of editorial peer review to the communication of scholarly and scientific information. (Weller, 2001, p322)

Acknowledgements

DVL Smith – Independent Research Agency who conducted four of the focus groups.

Bibliography

Barnes, Jonathan. (1981) "Proof and the Syllogism". p17-59 in Berti. Discusses the principles of Aristotle's endoxos

Bhattacharya, Shaoni (22nd October 2003). "Homeopathy reduces arsenic poisoning in mice", *New Scientist*

Brodsky, Marc H. (2003) as quoted by Lerner in "Fraud Shows Peer-Review Flaw" Vol 8, Iss 6, p12-17. *The Industrialist Physicist*.

Hauptman, Richard. (2002) "Dishonesty in the Academy", Iss. 2 *Academie,*.

Justice AC, Cho MK,Winker MA. (1998) 'Does masking author identity improve peer review quality?' *JAMA;* 280 pp 240-242

Lachmann, Peter. (2002) "The research integrity initiative: progress report", *Cope Report*, p11,

Lavelle. (1966). *Facts of Journal Publishing*, IV. Publications of the Modern Language Association of America, 81 (6).

Lerner, Eric. (2003) "Fraud Shows Peer-Review Flaw" Vol 8, Iss 6, p12-17. *The Industrialist Physicist*.

Mabe, Michael and Amin, Mayur (2002) "Dr Jekyll and Dr Hyde: author-reader asymmetries in scholarly publishing" *Aslib Proceedings,* Vol 54, iss 33 pp149-157

Olson, Carin M. (1990, July) "Peer Review of the Biomedical Literature," *American Journal of Emergency Medicine* 8 no.4: 356-358.

Smith, Richard. (1999, January 2) "Opening Up BMJ Peer Review," *British Medical Journal* 318 (7175) 4-5.

Ricaurte, George. (2003, Sept 8th) Letter to *Science*.

Van Rooyen, Susan et al. (1999) "Effect of open peer review on quality of reviews and on reviewers' recommendations; a randomised trial", *BMJ*, 318, 23-27 p12.

Wade, Nicholas. (2003, Jan 27)"Fraud Happens: What to Do About It", *The Scientist*, Volume 17, Issue 2, 56,

Weller, A. (2001) *Editorial Peer Review: It's strengths and Weaknesses*, ASIS&T: Information Today Inc Medford, New Jersey.

Ziman, J.M. (1968), *Public Knowledge: an essay concerning the social development of science*. London: Cambridge University Press

Zuckerman, H. & Merton, R.K. (1971, January). *Patterns of evaluation in science: institutionalisation, structure and functions of the referee system*. Minerva, 9(1), 66-100