

“Challenges to Quality Assurance: The Problem of Bias”

David Shatz, Yeshiva University

(December 2011)

Of the various criticisms that have been leveled at peer review, perhaps the most prevalent is that the process is biased. Usually, when we hear the term “biased reviewing,” we think of a reviewer bias for or against specific people, groups, institutions, ideas, or methods. In an expanded sense, however, any psychological factor that can skew an evaluation of merit can be called a bias (e.g., a general tendency of a reviewer toward leniency or stringency, or the recently studied phenomenon of decision fatigue—people’s decisions about similar cases depend upon when the decisions are made). In another sense, that connoted by “biased sample,” no psychological factor need be present; my concern here is with psychological factors.

My presentation discussed some literature on psychological bias and the issues that emerge from those studies. Some of the issues are discussed in my book on peer review (Shatz 2004), but the book deals with journal and book submissions, whereas my focus here is evaluating grant proposals.

The following were the main points of the presentation.

1. There may be no procedures for conducting peer reviews that do not require *some* tradeoffs. It has often been said of peer review what Churchill said of democracy—that it is the worst system for evaluation except for all the others. Still, there is a general feeling that while problems exist, peer review works well-- and can also be improved.
2. Although studies of journal peer review a few decades ago documented the operation of biases, more recent studies suggest that key alleged biases have been greatly reduced if not virtually eliminated. This may in part be a result of heightened sensitivity and awareness produced by the studies. A study of the Australian Research Council (Marsh et. al 2008), stated that “the only major source of systematic bias that we found” was in the use of author-nominated reviewers. The study’s authors therefore suggested not using a system with author nominations. In some cases there could be a tradeoff in excluding author-nominated reviewers—the latter may know a given specialized area of research better than reviewers not nominated by the author. In addition, even when author-nominated reviewers are not used, the external reviewers may turn out to be the very people whom the author would have nominated were there to have been nominations; in which case, if the Australian study is valid, we would have to conclude that bias would appear anyway (unless a reviewer writes differently about an author when the reviewer knows that he or she has not been nominated by the author). Still, the recommendation makes sense if bias is truly likely to be present. The Marsh study also found that North American reviewers tended to be more lenient. (I emphasize that in this paragraph I have cited only one study and that others may have conflicting results.)
3. Discussions by reviewers at meetings, combined with applicants’ responses to reviewer assessments, should serve as correctives to bias, except in cases where an evaluator

changes a score only to conform or to defer to other panelists. Interestingly, it has sometimes been proposed that *journals* should not provide referees with the results sections of submissions because the results might be judged in a prejudicial fashion. If this is true, then since grant applications do not provide results, they would be more fair than journal reviews that do provide reviewers with results. (This is not to say there *could* not be biases about, say, the significance of a project). Finally, as Marsh et. al note, reviewers render more meaningful judgments when they evaluate groups of proposals, as in panels, rather than assessing one or two proposals in isolation; it gives them a context. For this reason, sending external reviewers more than one proposal is probably a good idea; and to correct for differences in scoring methods, their rankings should be used rather than scores. In sum, vis-à-vis certain problems of bias, peer review of grant applications may have advantages over peer review of journal submissions.

4. In the Australian study, the authors found some effect of institutional affiliation but observed that “it is unclear whether this institutional affiliation effect on grant proposal ratings represents a source of validity (researchers from more prestigious universities are stronger researchers) or a source of bias.” This raises an interesting issue. In their responses to a 1982 study that strongly suggested institutional biases, some maintained that favoring authors from prestigious institutions is not a bias but a rational procedure. This is a controversial viewpoint because it produces the Matthew effect: the rich get richer and the poor poorer. It also may marginalize certain excellent researchers, and could stifle potentially impactful research. Clearly this is a difficult issue, and some sort of balance is a desideratum. (For details of the various arguments on the point, see Shatz, 63-68.)
5. Given that track records are taken into account in grant evaluations, it is more difficult for young and new investigators to receive grants, publish, and ultimately be selected as reviewers themselves. The use of special categories for young and new investigators (as in NIH) addresses this problem.
6. Many claim (and not only in the sciences) that peer reviewers are biased against new theories, notwithstanding that innovativeness is supposed to be a criterion for funding and publication. Some of the most widely cited scientific papers, including Nobel work, were originally rejected and received only delayed recognition. Reviewers, continues the criticism, tend to be those who are established and experienced in their fields and are wedded to particular methods and theories, which further entrenches existing paradigms. In response, some argue that new ideas, especially paradigm-busters, *ought* to encounter resistance, because in the nature of the case the existing paradigm will have more supporting evidence; also, they note, conservatism is a rational epistemological principle. (See Shatz, 83-107.). Medical editor David Horrobin argues, however, that in medicine, when we ask “What is peer review for?” the usual answer is “quality control.” But a further question must be asked, “What is quality control for?”, and the answer is: curing and relieving illness. For this reason, says Horrobin, “quality control must be only part of the equation.” How to assess risky proposals that might yield great benefits if funded and accepted for publication is an extremely difficult question. (Obviously, those who charge reviewers with conservatism must not understate the occurrence of innovation-- of breakthroughs and the revising or replacing of paradigms.) (For arguments on both sides of the conservatism debate, see Shatz 83-107.)

7. It is often thought that low interreferee agreement (reliability) is a problem for peer review, and that ways must be found to increase reliability. Yet, while disagreements make funding and publication decisions more difficult, they do not necessarily reflect biases or a lamentable subjectivity. In fact, there is a contrary viewpoint: “Too much agreement is in fact a sign that the review process is not working well, that reviewers are not properly selected for diversity, and that some are redundant” (Bailar). Langfeldt writes: “Low inter-reviewer agreement on a peer panel is no indication of low validity or low legitimacy of the assessments. In fact, it may indicate that the panel is highly competent because it represents a wide sample of the various views on what is good and valuable research” (p. 821; see also Bornmann and Daniel). The same is true of scientific disagreements themselves—they need not reflect bias but rather objective judgments based on reasoning, argument, and exchange of ideas. Why unbiased reviewers disagree and why some of their disagreements are irresolvable—rock bottom-- is an interesting question, as is the question of whether the disagreement of a peer should lower each peer’s confidence in his or her belief (see Feldman and Warfield). Be those questions as they may, disagreement is not a sign that judgments are not arrived at objectively.
8. The ultimate test of a particular funding procedure is the impact of the work that it funds. While this data can be traced, it would be good to know how much highly cited and impactful work had difficulty earning funding.
9. By way of conclusion: Awareness of potential bias in one’s judgments is a key to reducing bias. (Of course, “bending over backwards” to avoid being influenced by bias in one direction may sometimes produce a bias in the other direction, and reviewers should be alert to that potential bias too.) We can sensitize reviewers, and even in the extreme case where a particular like or dislike cannot be easily dislodged, it can be rendered causally inoperative by conscious awareness and effort. Jurors, professors, and judges often are able to set their biases aside—and so are scientific reviewers.

REFERENCES

- Bailar, J. C. 1991. Reliability, fairness, objectivity and other inappropriate goals in peer review. *Behavioral and Brain Sciences* 14(1): 137-138.
- Bornmann, L. and Daniel, H.-D. 2010. Reliability of reviewers’ ratings when using peer review: a case study. *Learned Publishing* 23 (2): 124-131.
- Feldman, Richard and Ted A. Warfield (eds.) 2010. *Disagreement*. Oxford: Oxford University Press.
- Horrobin, David 1991.. “The philosophical basis of peer review. *JAMA* 263 (10) 1438-41.
- Langfeldt, L. 2001. The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science* 31(6): 820-841.

Marsh, Herbert, Upali W. Jayasinghe, Nigel W. Bond 2008. Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist* 63 (3): 16-168.

Shatz, David 2004. *Peer review: a critical inquiry*. Lanham, MD: Rowman& Littlefield.