

Comment

R. J. Carroll

While the Editor of the *JASA Theory and Methods* section (1988–90), I was particularly interested in the issues of bias as they related to new researchers and senior researchers of outstanding reputation. Lacking hard numbers, I can only fall back on my remembered impressions. I have, however, discussed this with a number of current *JASA* Associate Editors, who have all assured me that I am not too badly miscalibrated.

The issue of double-blind reviewing has some similarities to the bias-variance tradeoff in nonparametric curve estimation. The Reid Report and the proposed pilot study make clear that adding variance to an already noisy process is a concern if a double-blind review system were to be implemented. If increased variance is observed, and I have no idea whether this will be the case, the question is whether the current biases can and should be removed.

Bias means that there are changes in response probabilities on the basis of factors external to the science (age, gender, institution, etc.). As an Editor, I observed a strong positive bias in favor of new researchers, both by Associate Editors and by referees. The New Researchers' report feared that the bias was to accept weak papers simply on the basis of new researcher status, but I found that this was not the case. Overwhelmingly, the form of the bias was in helping the new researcher turn a publishable idea into a publishable paper. Most often, this involved cutting the length of the paper and focusing on main points.

In effect, in my experience, new researchers had a higher probability of getting a helpful report and a

friendlier reception for their work. There were many cases of papers which in their published form are quite good, but whose original version would have received a quick and curt rejection except for the fact that the author was a new researcher. The pilot study, if done, should modify its stature variable to account for new researcher status, the categories "unknown or not well regarded" not being necessarily the same!

Bias in reviewing papers by senior authors is more complex than one might think. I followed a policy of having only senior Associate Editors handle such papers, to lessen the potential for intimidation to drive the system. There were three striking characteristics of reviews in this context:

- New researchers tended to give honest and fair reviews of papers, generally higher quality reviews than one would expect for papers from the average author. Interestingly, the recommendations tended to be more positive than the reviews. Perhaps simplistically, I attributed this to a reluctance to criticize superstars (halo effect).
- Senior referees and Associate Editors not personally connected to the author were often extremely, even shockingly harsh. There was a tendency for the peers of senior authors to hold the authors to a higher standard. If I had \$1,000 for every report that said something like "the author is capable of much better work," I would have a down payment for 100 acres bordering a trout stream in Montana.
- Senior authors had a lower probability of being hassled about the correctness of technical details.

My no-data impression is that a move to an implementable double-blind review system would have minor effects on senior authors and somewhat negative effects on new researchers not readily identifiable as such.

R. J. Carroll is Professor of Statistics and Member of the Faculties of Nutrition and Toxicology, Department of Statistics, Texas A&M University, College Station, Texas 77843-3143.

Comment

Christian Genest

Sometimes described as the gatekeeper of science, the peer review system provides credibility to the mod-

Christian Genest is Professor, Département de mathématiques et de statistique, Université Laval, Sainte-Foy, Québec, Canada G1K 7P4.

ern academic publishing industry. Its chief function is to ensure that minimum standards of objectivity and rigor are met in communicating the results of a scientific investigation. Quality, originality and pertinence of the material are some of the other issues that may be dealt with through refereeing, notably in highly

specialized or technical areas such as statistics. But the assessment of impact on the discipline and relevance of material to the readership of a particular medium often involve judgment calls that are best left to members of a journal's editorial board, whose responsibility it is to weigh the collected views of referees and give character to their publication.

There is a long tradition attached to the peer review system. As *users* of science, we all depend on it: our professional realizations are based upon the work of others, and we count on journal (and book) editors to separate the wheat from the tares. Although there is no such thing as perfection, it would be a disservice to the profession if too many scientific writings addressed irrelevant issues or contained gross factual errors. As *producers* of science, it is also in our interest that the system be fair: favoritism, discrimination and condescension bring discredit on the entire operation and ultimately work against the discipline, even if individual benefits occasionally may accrue in the short term.

At times, we have all heard of, if not encountered, cases of abuse or seemingly unethical behavior on the part of authors, referees and editors in the processing of scientific manuscripts: fraud, plagiarism, intellectual dishonesty, temporization, discrimination, arbitrariness in editorial decisions and other inequities of various kinds. Though such incidents hopefully remain isolated, they occur. We should thus fight our natural tendency to be conservative and skeptical when criticisms leveled against the current safety mechanisms for dissemination of knowledge suggest the need for change. As scientists, we should rather welcome all occasions to reflect on the act of writing, evaluating, editing and publishing research findings. The issue of double-blind refereeing, which recurs periodically in scientific circles, provides us with such an opportunity.

The Reid Committee Report does a reasonable job, I think, of summarizing the issues and findings pertinent to the question of double-blind refereeing. A substantial portion of the literature that they review points to potential sources of intentional and unintentional biases involved in the appraisal of scientific manuscripts. Because of obvious limitations inherent in studies of such a highly varying, subjectively based and irreproducible process, it is possible that evidence of bias in peer review may never be totally convincing. Nevertheless, a fact remains. Refereeing is *perceived* by many writers as being subject to various kinds of biases: biases in favor of male or female, young or established, national or foreign researchers, working at small or large institutions, in well-developed or developing countries and so on. Whether such biases are sufficiently strong and widespread to distort the whole review process is beyond the point. So long as the *potential* for abuse is there, we should guard against it, and double-blind refereeing is but one means of

ensuring such protection. It may not be the cure for all evils, nor the only remedy, but I believe it is a step in the right direction.

Because of their familiarity with the need to avoid bias in data collection and their knowledge of randomization and double-blind experimentation, I would have expected most statisticians to be spontaneously favorable to anonymous refereeing. I am surprised that the idea should meet with strong resistance in some segments of our scientific community. In fact, it is almost ironic that a leading statistical organization such as the IMS should be reduced to asking investigators who conduct double-blind clinical trials year round to plan an experiment to evaluate the merit of that very procedure in a refereeing context! Frankly, I do not understand what is still preventing most statistics journals from having their manuscripts reviewed anonymously in this time and age. Two and one-half years ago, *The Canadian Journal of Statistics* became one of the few exceptions, and I am proud to have been associated with that change, but one can hardly speak of innovation, considering that *Psychometrika* has used double-blind refereeing since its inception in 1936!

Though somewhat limited, my experience with double-blind refereeing as Associate Editor for *The Canadian Journal of Statistics* has been very positive, ever since the procedure was implemented in January 1991. Concealment of authors' identities has not made it any more difficult for me to find referees, no reviewer has yet complained to me or raised objections about anonymous refereeing, and more importantly, I have not suffered any increase in my editorial workload because of that policy. The blinding of submissions is done at the editorial office at minimal cost, it being understood that the onus is on writers to suppress all evidence internal to their manuscript (including the bibliography) that might tip off their identity to potential reviewers; there is a statement to that effect in the journal's cover pages. Being aware of authors' identities and their institutional affiliations, I have no difficulty avoiding flagrant conflicts of interest (colleagues and the like) and count on people's honesty and integrity to spot other difficulties of this nature. From that angle, things are not any different today than they were when authors' identities were known to the referees. I also continue to peruse the submissions that fall under my purview, just as in the old system. I find this to be an essential prerequisite to the selection of reviewers, and it provides protection against the (small) risk that some unethical contributors could take advantage of their anonymity to create the illusion that their manuscripts come from the pen of more prestigious authors. Finally, in cases where the reports I obtain are unusually abrasive or conflicting, it remains my responsibility, as always, to prepare an Associate Editor report that puts the reviewers' comments in per-

spective and provides guidance to the authors as to what would need to be done to make their submission suitable for publication in our journal.

Given my 2½ years of positive association with anonymous refereeing at *The Canadian Journal of Statistics*, I find it very difficult to understand why there should be so much reticence within the IMS institutions to such a harmless, low-cost measure. Double-blind refereeing truly requires minimal logistics, involves basically no additional work for editorial boards and has the great virtue of explicitly recognizing the irrelevance of an author's name for assessing a manuscript's suitability for publication while guarding the peer review process against undesirable potential biases. Yet, even in its cautiously receptive view of double-blind refereeing, the Reid Committee appears uncertain that the benefits of that procedure outweigh the five potential disadvantages they were able to identify. I would like to argue that these disadvantages are not real, and that there is no serious reason to further delay the introduction of anonymous refereeing in IMS publications. In the following, italicized headings are excerpted from the Reid Committee Report.

Potential Disadvantage 1: *Most editors and many referees feel that they provide extra advice to new researchers, particularly over presentation; this extra assistance will not be available under double-blind refereeing.*

There are three implicit assertions here. The first is that manuscripts submitted by new researchers should be appraised differently than those of experienced authors, not only as regards presentation but presumably with respect to contents also. The second is that preferential treatment of this sort is applied under the current system. The third is that targeted groups of researchers could not be given special attention if manuscripts were refereed anonymously.

In connection with the first point, I would begin by questioning the wisdom of giving favored treatment to young Ph.D.'s in terms of the quality or scope of their publications. In my opinion, this is a double-edged sword. For one thing, new researchers are being misled when they are sent inappropriate signals about what constitutes a significant contribution to the discipline. Though such leniency may provide initial encouragement, my impression is that it does more harm than good in the long run. After meeting with early success, many contributors will be puzzled, frustrated and often even discouraged from conducting further research when they realize that their contributions suddenly need to improve substantially in quality and scope to remain suitable for publication. As for those who persevere, it may come to them as a shock if they ever discover that they *may* once have benefited from preferential treatment; for, they will then have to live forever in lingering doubt that their early work *may* not

have been quite up to the standards but was published nonetheless, as an act of sympathy or encouragement. Is it not simpler and fairer to apply the same standards of excellence to all? Otherwise, where will one draw the line between the young and the old, the beginner and the experienced and so on?

This brings me to the second point. If indeed the editorial boards of some statistics journals currently have an implicit "affirmative action" policy favoring new researchers or any other classes of article contributors, I would urge them to make such practices public and explicit without delay. On the one hand, it would avoid misleading new researchers and would spare them anguish and disillusionment as they grow older. On the other hand, it would guarantee that potential members of preferential treatment groups are recognized as such and served in accordance. Under the present system, I do not think that editorial boards could seriously pretend to be able to determine with any degree of certainty the status or background of all submitting authors without explicitly soliciting that information.

My third point is that there is absolutely nothing in double-blind refereeing to prevent a journal from giving preferential treatment to target groups of researchers, provided that such policy be explicit and public. Manuscripts submitted from members of such groups could be flagged to referees without revealing the author's identity or institutional affiliation. As explained above, I personally doubt the value of such "affirmative action" policies. If known publicly and applied in conjunction with double-blind refereeing, however, they would at least have the merit of being fair to all within the targeted groups while limiting other forms of bias based on gender, nationality and the like.

As concerns editorial assistance with respect to presentation, I would simply like to comment that the need for counsel is felt by *all* authors, young and old, particularly when they venture into new areas of research. Different topics call for different journals and for different exposition styles. One may well master the art of describing new theoretical results to a mathematically inclined audience and experience difficulties motivating statistical concepts and procedures in an applied context. Also, I do not think it can be taken for granted that all people become more efficient communicators as their career progresses. The act of scientific publishing nearly always involves an iterative process between authors, editors and reviewers, resulting in a certain amount of paper honing. In that regard, we all stand to learn and would all welcome guidance.

Potential Disadvantage 2: *It may be advantageous, particularly for new researchers, to publicize their pre-publication work through the refereeing process.*

In my opinion, it is neither efficient nor particularly

desirable to disseminate research work through the refereeing network. What are we talking about here? Half a dozen people in all? An Editor, an Associate Editor and two or three referees? Let us not mix apples with oranges: refereeing is a mechanism for ensuring quality in scientific publications; editorial boards are not advertising agencies. Researchers, young and old, who wish to make their contributions known had better give seminars and present communications in scientific meetings. Unless investigators were exceptionally aggressive on these fronts, making presentations in all parts of the world, their anonymity should be preserved with high probability, inasmuch as refereeing of their articles is concerned. If they wished, certain writers could also choose to send advance copies of their work to selected individuals, after taking the precaution of publishing them in their institutional Technical Report Series to establish priority. If well planned and sufficiently extensive, such mailings might have the effect of revealing a contributor's name to most of the natural referees, but that would then be the author's doing. It would be interesting to see how many people would go to that extent, presumably in the hope of influencing the peer review process in their favor; very few, I bet. Should double-blind refereeing become the norm, this way of proceeding might even come to be regarded as unethical, just as attempts to influence an Associate Editor through direct communication are today.

Potential Disadvantage 3: *The name of the author is a relevant piece of information that makes the refereeing process more efficient, in that judgments about the likely impact of the work are legitimately influenced by the author's name, background and the importance of his or her previous work.*

This is, in capsule, exactly what advocates of double-blind refereeing argue against. To me, an author's name, status and institutional affiliation are irrelevant information that should in no way affect judgment about a paper's scientific contents and suitability for publication. (This would not be true of a research proposal submitted to a granting agency, however. A grant application requests funding for investigations to come. It is a promise, and as far as promises go, what's past is prologue. Also, note that we review articles, but that we grant people.) I would thus have serious reservations against using a referee who insisted that he/she needed to have access to these personal and social attributes of scientists in judging the merits of their contributions. To make an analogy, what would you think if, in the context of a clinical trial, a medical doctor refused to tell whether or not a patient has recovered from a certain illness prior to release of information concerning the treatment that was administered? Is that not precisely the sort of bias-prone behavior double-blind studies are designed to avoid? My profound conviction is that the assess-

ment of an investigation's merits should not depend on anything but the contents of the submitted contribution, and that as a result, the refereeing process should not be more lenient for some, and less so for others. Implementation of anonymous refereeing in the IMS journals would foster recognition of that fact throughout the profession. The strong resistance to double-blind refereeing observed currently in certain academic circles rather perpetuates the notion, among those who are suspicious of the system, that factors such as name, institutional affiliation and social status in the scientific community somehow come into play in the editorial process. This open door to the intrusion of particularistic standards is sometimes conjectured to profit a mythical "old boys club," whose influence has even been argued by some to be beneficial to science and to the community of scientists. While it is clear that double-blind refereeing would not be completely effective in controlling the influence of such networks, assuming they exist, editorial boards would at least have to take the responsibility for their effects and be accountable for them. It is also worth noting that the objectivity conferred to referee reports by double-blind procedures would increase their value and might then help keep Associate Editors and Editors honest.

Potential Disadvantage 4: *Some distortion in the refereeing process may be introduced by referees trying to guess the author, sometimes of course incorrectly.*

Even if a manuscript does not tip it off, it is undeniable that an author's identity can sometimes be guessed correctly, especially if the line of work is highly specialized. Nowadays, however, most journals receive contributions worldwide from an ever expanding pool of researchers. Thus, there is *always* a chance that an author might be unknown to the referees: foreigner, fresh Ph.D. working in their mentor's area of expertise, established investigator exploring new horizons. My point is that so long as a doubt remains, biases against unknowns cannot surface, and that is really what matters. Now I am not clear on the sorts of distortions alluded to in the Reid Committee Report, that trying to guess authors' identity could introduce in the peer review process. Again, what seems to be implied here is that refereeing somehow might be done differently if an author's identity were known; this in itself is an admission of bias in the system. I think most statisticians will agree that people who spend more time trying to identify authors than looking at their work have grave misconceptions concerning their responsibilities as referees. If I, as an Associate Editor, came to be informed of this behavior, I would not hold the opinion of such referees in any great regard and would probably stop consulting them altogether. In fact, it is rather likely that their reports should be superficial and of little editorial assistance anyway. Frankly, researchers

who like guessing games should have plenty to do already trying to unmask the referees of their own papers!

Potential Disadvantage 5: *There is some increased editorial burden in changing the cover sheet of the submitted paper to remove the authors' names.*

As indicated in the Reid Committee Report, this is only a slight burden that can be transferred to the contributors of articles, when anonymous refereeing becomes a journal's policy. I would just want to emphasize that the editorial board of *The Canadian Journal of Statistics* experienced no difficulty whatsoever in completing this transfer, and that many psychology and social sciences journals have successfully operated under double-blind refereeing policies for much longer periods of time, not to mention *Psychometrika*.

In summary, my reading of the situation is that none of the above constitutes a real argument against double-blind refereeing. While I appreciate the cautiousness of the IMS Council, and its desire to collect its own data and proceed to a trial run before full-scale implementation of this policy in its journals, I would contend that abundant literature and the experience of many scientific publications, including statistics journals, provide ample evidence already that anonymous refereeing bears no strong disadvantages and many potential benefits. One should thus be careful not to invest too much energy on experimentation. While the preceding reports both make good suggestions about the design of such an experiment, I am afraid that too much time and effort may be required to reach a definite conclusion. It is important to realize

that for all the extra work such a study will impose on editorial boards of IMS journals, it is not likely to prove that the current system is *better* than double-blind reviewing. It could only fail to detect bias or show that it is not statistically significant. Meanwhile, the *potential* for bias will always remain.

As I have tried to argue, double-blind refereeing is a simple, low-cost procedure that neither increases editorial workload nor reduces referee collaboration in any significant way. Although it may not be totally effective in eliminating all possibilities of bias (no procedure could be!), it would *at the very least* alleviate perception of unfairness within the statistical community. This in itself would be a source of professional satisfaction and stimulation. But beyond public relations, the introduction of anonymous reviewing is likely to put editorial responsibility where it belongs and to send referees back to their prime duty: the conscientious, objective assessment of the scientific merit of research manuscripts. If double-blind refereeing could accomplish this, would we not all be grateful for it?

DISCLAIMER

Although I was a member of the Board of Directors of the Statistical Society of Canada and an Associate Editor for *The Canadian Journal of Statistics* at the time when its double-blind refereeing policy was adopted, the opinions expressed herein are mine alone. They should not be construed to represent in whole or in part either the official views of the Society or those of its journal's present or past editorial board.

Comment

Willem R. van Zwet

Apparently many people hold strong views on the issue of double-blind refereeing, and to me the main virtues of the report of Nancy Reid's committee are its measured tone and balanced views. Being one of the group of people who "have had experience of the editorial process", I am not convinced that double-blind refereeing will make much difference one way or the

other. An often heard argument that is also noted by the committee is that even if there would be no bias, it is important to eliminate the perception that there is. Of course there is something to be said for this, but on the other hand, our society today is rich in such perceptions, and one cannot eliminate all of them in a lifetime. As an onlooker from another continent, I cannot help noticing a certain similarity to the political correctness ideology that appears to be so powerful in the U.S. right now.

The Reid committee has not taken this path and wisely proposes to carry out an experiment first. This may indeed produce some interesting facts, but it will also make the Editor's life even more miserable than

Willem R. van Zwet is Professor of Mathematics, University of Leiden and University of North Carolina. Present address: Department of Mathematics, University of Leiden, P.O. Box 9512, 2300 RA Leiden, The Netherlands.