

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

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.

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

it already is. With 400 papers a year, it is really no joke to spend another 2 hours or so on each paper to conduct this experiment in a responsible manner.

In the meantime, John Crowley's committee has thought a bit more about the experiment. This committee's point of view is that the editorial recommendation of the referee is a more interesting variable to study than the quality of the referee's report. Perhaps we should study both variables. After all, the main complaint seems to be that people get their papers rejected, and I shall argue below that only negative recommendations in referees' reports of high quality can contribute to this.

I realize that these remarks of mine are neither new nor particularly interesting. If I am to make any contribution to a better understanding of this issue, I should perhaps try to explain how a journal like *The Annals of Statistics* functions. It seems to me that many people have a somewhat unrealistic view of this. Having had the interesting experience of spending some 30 hours a week editing this journal for 3½ years, I am perhaps well qualified to do so, and it may shed some new light on the issue we are discussing.

Editorial decisions to publish or to reject papers are an imperfect business at best. What an Editor can do is to appoint 30 of the best people that can be found to the editorial board. Together their competence should range over the entire field, and ideally their total expertise should be distributed over the field in a way that will match the distribution of the submissions. Of course this never works out, so the Editor is continually faced with the problem that a paper is submitted that could be handled ideally by Associate Editor X, but unfortunately this was also true for four earlier papers during the past month, so X is already badly overloaded. Associate Editor Y is a reasonable alternative, but not quite as knowledgeable as X in the area of the paper. Moreover, Y will start moaning if the Editor does this to him too often. Of course the Editor will also make mistakes from time to time and send papers to the wrong Associate Editor without realizing that another member of the editorial board would be much more knowledgeable. Believe it or not, assigning submissions is one of the Editor's hardest problems, and a good deal of time is spent in trying to optimize this simple decision which is probably at the root of the majority of editorial errors.

The Associate Editor has a similar problem finding referees. Good referees are scarce and overworked, and Associate Editors are continually trying to expand the group of people they are working with. In fact there is such a shortage of good referees that I would estimate that about 40% of all referees' reports are fairly useless. This is not such a disaster as it may seem, because the vast majority of bad reports are easily

recognizable. They produce mostly generalities and little substance, and for the expert it is easy to see that the referee has not really read or tried to understand the paper. Such reports will generally be ignored by the Associate Editor and the Editor, and as a result they do not do much harm. I would not be surprised if an experiment would show that such referees would be overly impressed by well-known authors or unduly critical of less-established ones, but this will not make much difference for the final editorial decision.

A good referee's report will comment on the content of the paper and the way it is written, but will also evaluate the relevance of the paper. The validity of comments on the actual content can usually be verified rather easily and does not lead to wrong decisions very often. In the rare cases in which a paper is incorrectly rejected because of wrongly perceived mistakes, the author will be quick to point this out, and the paper will be reconsidered.

Critical remarks on writing style are much more controversial. Even though they do not normally lead to incorrect decisions to accept or reject a paper, they often infuriate authors. The fact of the matter is that there is a large amount of bad writing around, and it comes not only from inexperienced authors. There are just too many people who are unable or unwilling to write down an argument leading from A to B, without getting sidetracked to C, D and E on the way and messing up the original argument. Then there is a large group of people who do not bother to explain what they are doing; others explain too much in a 40-page paper, with about 5 pages of new content. Happily, the *Annals* now encourages informative introductions, but there is still an upper bound on the amount of useful information. It is not true that a paper gets better with increasing length. Economy of expression, a degree of mathematical elegance and the modesty not to tell the reader everything you know are often greater virtues. Because of the pressure to publish, too few people polish their results a little. Rather than looking for logical and clear mathematical arguments, they submit the first, and often laborious, proof that comes to mind. While this is understandable for new researchers, it is less so for senior people who are not as much subjected to career pressure.

When it concerns inexperienced authors, the *Annals* has a time-honored policy to try to help a little to improve the mathematics as well as the style of writing. The philosophy behind this is simple: it pays to put this work in early because it will improve the quality of future submissions by the same author. Of course most Ph.D. advisors at major universities know very well what makes a good paper, and it would be extremely helpful if they would pay a little more attention to these aspects when a student is preparing

a thesis or the papers originating from it. The mentorship program proposed by Deborah Nolan's New Researchers Committee may provide some relief in other cases. I would hope that it would not be restricted to women, because non-native English speakers and graduates from small schools need the same kind of help too.

It is the judgement of relevance, however, that creates most of the difficulties. Put inelegantly, it is based on bias, pure and simple. As long as this is a conscious bias, it is not necessarily a bad thing: it is also called the editorial policy of the journal! After all, the *Annals* is not supposed to publish everything that is formally new and correct. An additional criterion is that a paper should contribute something to the development of the field. It should contain at least one novel idea that could potentially advance the subject. Papers that make only a small technical advance of no great importance or novelty are not acceptable.

Each of us can point to certain areas in statistics that have seen vigorous development for many years, which is slowly coming to an end. Very many papers may still be written in such an area because grand traditions do not die easily, but there may be very little real progress. As the group of people working in this area shrinks, it also draws together more closely. After a while it is almost impossible to obtain an impartial report on a paper in such an area. The in-crowd will tend to be too sympathetic, whereas the profession at large may be overly negative. In such a situation the Editor's choice of an Associate Editor for handling manuscripts in the area more or less decides the fate of a large number of papers.

This is merely an example of a conscious bias which is presumably introduced after a good deal of thought and is part of the scientific policy of the journal. There is no reason to assume, however, that unconscious biases are not equally, or even more important. Most of us like the particular area we are working in, and Associate Editors are no exception to this rule. Some may be overly sympathetic to papers in their own area or overly critical of developments in this area that they do not like for good reasons or bad. Referees have only a very limited influence on such patterns. In the first place they are selected by the Associate Editor, and in the second place the Associate Editor is sufficiently knowledgeable to recognize bias on the part of the referee and discount it when making his recommendation.

How about the Editor? Does he or she have any power once an Associate Editor has been chosen to handle a paper? Of course the Editor can influence matters, but there are limitations. If the Editor feels an Associate Editor is making a mistake, the Associate Editor's recommendation will be ignored. However, the Editor must explain this decision very carefully to the

Associate Editor and discuss things with him or her to prevent a serious divergence of opinion. You cannot reverse an Associate Editor's decision three times in a row and expect this person to stay on board! Also, the Associate Editor is the real expert and the Editor will only be able to challenge his or her judgement in a limited number of areas. Basically the relation between Editor and Associate Editor is one of trust in each other's intellectual honesty.

The point I hope to have made is that this business is full of biases that are mainly introduced by the choice of an Editor, his or her choice of an editorial board and the Editor's decision which Associate Editor to assign a particular manuscript to. Referees are not a main source of bias, and blinding them is not going to change very much. Also, bias toward the subject matter of a paper is very much stronger and has much more influence on editorial decisions than bias for or against the author. Conscious subject matter bias is not necessarily bad: it constitutes the editorial policy of the journal. It is the unconscious variety that we have to watch out for.

While explaining these matters, I have probably reinforced some people's conviction that this is a dark and murky area that is occupied by an old-boys network, out to protect its own interests at the expense of defenseless authors. All I can say is that it is not. Admittedly, editorial boards do consist of moderately old boys (and girls) with an average age somewhere in the forties, but they are not much of a network. There is no attempt at continuity in the choice of Editors, and new Editors bring in large numbers of new people that they think highly of, and whose judgement they feel they can trust. My own experience is that all members of the editorial board were acutely conscious of the dangers of bias and doing their best to keep this in check. Controversial recommendations were argued in detail, and difficult decisions were discussed with individual Associate Editors and among the editorial board as a whole. If editing a journal remains an imperfect business, this is not because of, but in spite of the best efforts of a dedicated group of people.

Let me end with some points that are not really related to the discussion of double-blind refereeing but have come up in connection with the problems that young researchers are facing. First there is the sometimes excessive waiting time for a first report on a paper. This is the Editor's worst nightmare. Utterly reliable Associate Editors suddenly "burn out," and it is impossible to get anything out of their hands. Usually prompt referees repeatedly promise reports within a few weeks and never deliver. By the time this becomes clear, reassigning the paper means even more loss of time. Of course, these problems occur mostly with difficult papers and are unrelated to the name of the

author. However, young researchers are clearly affected much more than senior ones.

Everybody agrees that such things should not happen, but I know of no way of preventing them occurring from time to time. Draconic suggestions like firing Associate Editors or blacklisting referees the moment they do not deliver on time, sound fine in theory but are not realistic in practice. An Editor has to work with volunteers who are willing in principle to donate their time and effort, but sometimes just cannot get themselves in the right frame of mind to do the job. I believe we all know the feeling.

Another sore point is the often matter-of-fact and sometimes downright unpleasant tone of referees' reports that upsets inexperienced authors. Of course there should be no place for unpleasant remarks in the refereeing process since the idea is to judge papers, not people, as I believe Steve Stigler pointed out in his *JASA* days. In bad cases, the Editor will edit the referee's report before sending it to the author, but time simply does not permit doing this all the time. Unpleasant though this is, I am afraid we have to live with a bit of incivility, and experienced authors know how to turn this to their advantage by reacting in an overly polite and reasonable, if somewhat ironic, fashion. This will alert the Associate Editor to the problem and put the referee on the defensive.

Of course it is always a good idea to be reasonable and civil even if you feel that you have been treated badly and that your best efforts are not getting the applause they deserve. Like most other people, (Associate) Editors do not like to be yelled at. This elementary truth seems to have escaped quite a few authors. A growing number of them complain loudly about even the mildest criticism of their paper and the slightest delay in handling it. This is not helping to smooth tempers either. Since the set of referees is a subset of the set of authors, one could hope that the authors who complain loudest perform better and more responsibly as referees. I have often wondered about this, but to put it mildly, I have never found any evidence pointing in this direction.

ADDITIONAL REFERENCES

- ASTIN, H. S. (1973). Career profiles of women doctorates. In *Academic Women on the Move* (A. S. Rossi and A. Calderwood, eds.) 139-161. Russell Sage Foundation, New York.
- ASTIN, H. S. and BAYER, A. E. (1973). Sex discrimination in academe. In *Academic Women on the Move* (A. S. Rossi and A. Calderwood, eds.) 333-356. Sage Foundation, New York.
- BILLARD, L. (1991). The past, present and future of women in the mathematical sciences. *Notices Amer. Math. Soc.* 38 707-714.
- BILLARD, L. (1992). Twenty years later: Is there parity for women in academia? Technical Report, Dept. Statistics, Univ. Georgia, Athens, GA.
- BILLARD, L. (1993). A different path into print. *Academe* 79
- BILLARD, L. (1994). Women and academic careers. In *International Encyclopedia of Education*. Pergamon Press, Oxford. To appear.
- BLACKBURN, R. T., BEHYMER, C. E. and HALL, D. E. (1978). Research note: Correlates of faculty publications. *Sociology of Education* 51 132-141.
- CHAMBERLAIN, M. K. (1988). *Women in Academe: Progress and Prospects*. Sage Foundation, New York.
- COLE, J. R. (1979). *Fair Science: Women in the Scientific Community*. Free Press, New York.
- COLE, J. R. and ZUCKERMAN, H. (1984). The productivity puzzle: Persistence and change in the patterns and publications of men and women scientists. In *Advances in Motivation and Achievement* (P. Maehr and M. W. Steinkamp, eds.) 217-258. JAI Press, Greenwich, CT.
- DAVIS, D. and ASTIN, H. S. (1987). Reputational standing in academe. *Journal of Higher Education* 58 261-275.
- FISH, M. and GIBBONS, J. D. (1989). A comparison of the publications of female and male economists. *Journal of Economic Education* 20 93-105.
- FIDELL, L. S. (1970). Empirical verification of sex discrimination in hiring practices in psychology. *American Psychologist* 25 1094-1098.
- FREEMAN, B. C. (1977). Faculty women in the American university: up the down staircase. *Higher Education* 6 165-188.
- LEFKOWITZ, M. R. (1979). Education for women in a man's world. *Chronicle of Higher Education* 56 (Aug. 6).
- MICHELSON, R. A. (1989). Why does Jane read and write so well? The anomaly of women's achievement. *Sociology of Education* 62 47-63.
- OVER, R., OVER, J., MEUWISSEN, I. and LANCASTER, S. (1990). Publication by men and women with same sex and cross-sex Ph.D. supervision. *Higher Education* 20 381-391.
- PERSELL, C. H. (1983). Gender, rewards and research in education. *Psychology of Women Quarterly* 8 33-46.