Bulletin of The British Psychological Society (1979), **32**, 390—391 *Printed in Great Britain*

Editorial role in author—referee disagreements

Andrew M. Colman

The referee is the lynchpin about which the whole business of science is pivoted (Ziman, 1968).

I wonder how many readers have encountered problems similar to the following. A manuscript was recently returned to me and my co-authors by the editor of a well-known journal together with two anonymous referees' reports. The editor invited us to modify the paper along the lines suggested by the referees. Unfortunately, there were two peculiarities about the referees' reports. In the first place, they were diametrically at odds with each other: one criticized the statistical treatment of the data and recommended a complete re-analysis, while the other went out of his way to praise the statistics as 'both sophisticated and appropriate'. Secondly, the latter referee wanted us to acknowledge a possible source of artifact in the results which, as far as we could judge, was ruled out by the randomization procedure we had used. The mutual contradictoriness of the referees' reports placed us in a sort of Laingian double-bind, while the second referee's criticism seemed — from a purely logical point of view — to be erroneous.

We decided to appeal to the editor to cut the Laingian knot by adjudicating between the referees on nthe statistical question, and to decide for himself about the logic of the other criticism. In the event, the editor declined to make any comments and left the final decision in the hands of the referees. The second referee, needless to say, refused to back down and insisted, rather more shrilly than before, that the paper should not be published unless the source of artifact was acknowledged. Eventually, after several resubmissions, we somehow managed to perform the miraculous feat of satisfying both referees without unduly compromising the paper, and the story had a happy ending.

I have often found referees' reports helpful and to the point (even when they have not recommended unqualified acceptance!) but occasionally they have been less than impressive. Like authors, referees are of course human and therefore fallible. Refereeing is a difficult and often tedious job which brings neither fame nor fortune, and it is hardly surprising that blunders sometimes occur. Incompetent refereeing leads to two types of error: acceptance of bad papers (false positives) and rejection of good papers (false negatives). Examples of false positives are not difficult to find in leading journals; some of the transparently incompetent papers of the late Sir Cyril Burt spring readily to mind. False negatives, on the other hand, seldom come to light because it is difficult for an author to advertise the fact that one of his papers has been rejected without appearing not only professionally incompetent but also small-minded and resentful. Occasionally, however, the secret leaks out. It is now well known that the famous paper by Garcia & Koelling (1966) which revolutionized modern learning theory was rejected by a number of respectable journals, with the consequence that flavour aversion learning was not widely accepted until 15 years after the first clear and rigorous evidence for it had been obtained (Revusky, 1977, p. 63). One further example from a different context is worth mentioning. Karl Popper's constroversial but enormously influential *Poverty of Historicism* papers (Popper, 1944*a*, *b*, 1945) failed to reach the appropriate readers for many years because the manuscripts were all originally turned down by *Mind* (Popper, 1976, p. 119). In spite of the usual conspiracy of silence surrounding false negatives, there are good reasons for believing that they occur quite frequently (Ziman, 1970; Gordon, 1977; Revusky, 1977).

When disagreements arise between authors and referees, it is unreasonable to assume that the referees are always right. In general, the authors of a manuscript have spent many months considering and reconsidering the adequacy of their arguments, their experimental design, their interpretation of the results and so on. They have frequently discussed these matters informally with their colleagues and presented their work for scrutiny at departmental seminars or at learned conferences before submitting it in the form of a journal article. They are therefore often at least as familiar with its strengths and weaknesses as is a referee who spends perhaps an hour or two examining it. The referee, in any event, often lacks an intimate knowledge of the specific area of research in which the authors are immersed. The submitted manuscript may indeed be seriously flawed — it often is — but if the authors claim to be able to rebut the referees' criticisms of it, their arguments should be taken seriously by the editor.

The editor of a top-level journal has an unenviable task to perform. He is typically embarrassed by the number of manuscripts which land on his desk. Most of the best psychological journals have rejection rates in excess of 80 per cent, and some reject as many as 99 per cent (details are given by Markle & Rinn, 1977). The editor's main task is to see that the best papers are published, but how is this end to be achieved?

In an ideal world, some omniscient decision-maker would rank-order all the available manuscripts each time an issue was due, and the ones at the top of the pile would be published. In practice, however, no one is competent to judge submissions in diverse areas, so different referees have to evaluate different manu-

0007-1692/79/1010-0390$02.00/0

scripts. But this is a highly unsatisfactory state of affairs. An analogy may be helpful at this point. Suppose that 100 candidates have applied for places on a course, but only 20 places are available. Few psychologists, I imagine, would have much faith in a selection procedure whereby each applicant was evaluated by a different interviewer, and each interviewer was kept more or less in the dark about the quality of the competing candidates. If, for some reason or other, such a decentralized evaluation procedure were unavoidable, the validity of the selection would probably be improved if some individual were designated to make the final selection, using summary notes from each interviewer as a guide. A similar argument is applicable, I believe, in the case of submissions to a learned journal: the editor ought to make the final decisions about acceptance or rejection, using referees' reports merely as a guide.

The editor ought, furthermore, to be alert to all the possible motives for a referee's recommendation regarding a specific manuscript. The following spring to mind. The referee may recommend rejection because the paper is genuinely weak, because he is imposing inappropriately stringent standards, because he is not sufficiently familiar with the field to appreciate fully the significance of the work, because he has not studied the paper closely, because the paper contradicts one of his pet theories, or because of negative personal attitudes towards the author(s). He may recommend acceptance because the paper is genuinely good, because he is a relatively lenient referee, because his own work is favourably cited in the paper, because he has not devoted enough attention to the paper to discover its fatal weaknesses, because the paper supports one of his pet theories, because he is a fan of the author(s), because he believes the author(s) may guess who he is, or because the author(s) recently refereed one of his own papers or may do so in the future. At all times, but particularly when the referees disagree among themselves about a particular manuscript, the editor ought, in my view, to use his judgement in deciding on the trustworthiness of their evaluations.

There is very little empirical evidence concerning the reliability and validity of the refereeing system, but the few studies which have been carried out are far from encouraging. Scott (1974) reported an inter-referee reliability of 0.26 on various characteristics of 328 manuscripts submitted to the *Journal of Personality and Social Psychology*. Mahoney (1976) sent 75 manuscripts to guest reviewers from a 'well-known psychological journal' and found the referees to be much more critical of manuscripts whose results contradicted their presumed theoretical standpoints than they were of manuscripts which differed from these only in reporting 'congenial' results. There is no evidence for any positive correlation between referees' evaluations and the scholarly impact of published papers as measured by citation counts; in fact a survey conducted by the Primary Communications Research Centre at Leicester University has revealed a slight negative correlation (Gordon, 1977).

I have two positive suggestions to make. Firstly, referees should be asked to comment as fully as possible on the manuscripts which they are sent, but it should be made clear that the decision whether or not to publish is not up to them. Only the editor, when he has all the evidence in front of him, is in a position to make comparative evaluations among all the competing manuscripts. Secondly, whenever there is any room for argument about a referee's criticism of a manuscript, the author(s) should as a matter of course be invited to comment upon them, and the editor should defer his decision until he has examined the arguments on both sides. He should not hesitate to refer back to the referees and authors for further clarification until he feels able to make a fair decision. This would be a great improvement over the current system whereby authors are tried and condemned by anonymous judges without even being invited to defend themselves.

These recommendations are hardly radical, nor can they be expected magically to transform a process of dubious validity into a completely fair one, but they may eliminate some of the more blatant errors to which the system is prone, and improve the quality of the journals.

References

Garcia, J. & Koelling, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science,* **4,** 123—124.

Gordon, M. (1977). Evaluating the evaluators. *New Scientist,* **73,** 342—343.

Mahoney, M. J. (1976). *Scientist as Subject: The Psychological Imperative.* Cambridge, Mass.: Ballinger.

Markle, A. & Rinn, R. C. (1977). *Author's Guide to Journals in Psychology, Psychiatry and Social Work.* New York: Hawarth.

Popper, K. (1944*a*). The poverty of historicism. I. *Economica,* **11,** 86—103.

Popper, K. (1944*b*). The poverty of historicism. II. *Economica,* **11,** 119—137.

Popper, K. (1945). The poverty of historicism. III. *Economica,* **12,** 69—89.

Popper, K. (1976). *Unended Quest: An Intellectual Autobiography.* London: Fontana.

Revusky, S. (1977). Interference with progress by the scientific establishment: Examples from flavor aversion learning. In N. W. Milgram, L. Krames & T. M. Alloway (eds), *Food Aversion Learning*, pp. 53—71. London: Plenum.

Scott, W. A. (1974). Interreferee agreement on some characteristics of manuscripts submitted to the *Journal of Personality and Social Psychology. American Psychologist*, **29,** 698—702.

Ziman, J. (1968). *Public Knowledge: The Social Dimension of Science*. Cambridge: Cambridge University Press.

Ziman, J. (1970). Some pathologies of the scientific life. *Nature, London,* **227,** 996—997.

Requests for reprints should be addressed to **A. M. Colman,** Department of Psychology, University of Leicester, Leicester LE1 7RH.