

Andrew M. Colman

Department of Psychology, University of Leicester, Leicester LE1 7RH, England

One of the most influential discoveries in modern science, the first law of thermodynamics (sometimes called the law of the

conservation of energy) was first reported by the German physician J. R. Mayer in 1842. But Mayer's revolutionary paper was rejected by the leading physics journal *Annalen der Physik* and was eventually published in a relatively obscure and much less appropriate chemical journal. It was therefore almost entirely ignored by physicists, and, possibly as a result of this, Mayer suffered a mental breakdown from which he never recovered (Ziman 1976, pp. 103-4). This is just one dramatic example of the fallible judgments to which journal editors and referees are sometimes prone; I have cited some other equally shocking examples elsewhere (Colman 1979).

Peters & Ceci (P & C) have reported some interesting empirical evidence from a controlled investigation of the peer-review system. Their data demonstrate that the system is vulnerable either to random error or to systematic bias, or possibly to both. The authors acknowledge that in order to test the bias hypothesis properly it would be necessary to compare the fate of resubmitted articles purporting to come from high-status institutions with the fate of others purportedly from low-status institutions. Since this manipulation was not performed, P & C's interpretation of their results as supporting the bias hypothesis lacks force, and I believe that their statistical arguments against the random error hypothesis are unsound.

Let us assume that the ultimate fate of a submitted manuscript or an undetected resubmission is a purely random event, unrelated to its quality or to the authors' reputations or their institutional affiliation. Suppose that there is a fixed probability p that the manuscript will be accepted and a probability of $q = 1 - p$ that it will be rejected. (In reality, of course, there are other possible outcomes apart from outright acceptance and outright rejection, but I shall ignore this complication as P & C have done.) If the number of submitted - or resubmitted and undetected - manuscripts is N , then the exact probability $P(x)$ that x of them will be accepted is given by the binomial probability function:

$$P(x) = \frac{N!}{x!(N-x)!} p^x q^{N-x}, x = 0, \dots, N.$$

The logical derivation of this formula is explained from first principles in Colman (1981, chapter 4). P & C correctly state that if $N = 9$, $p = .43$, and $q = .57$, then the probability of less than two acceptances - that is, one or zero acceptances - is $P(1) + P(0) = .046$.

This does not, however, answer the question, How improbable is the observed outcome of less than two acceptances out of nine on the basis of chance alone given the actual acceptance rate (20 percent) of the journals studied? The required probability is obtained by setting $N = 9$, $p = .20$, and $q = .80$. Then, according to my electronic abacus, the probability of less than two acceptances is $P(1) + P(0) = .44$. This means that, on the assumption of purely random selection, the probability of an outcome as extreme as that observed by Peters and Ceci is .44, which is certainly not low by any standards. Furthermore, the expected number of acceptances, given the above parameters, is $Np = 1.80$, which is fairly close to the observed outcome of one acceptance (the standard deviation is 1.20). In other words, if the experiment were repeated many times, then between one and two manuscripts, on average, would be accepted per experiment. It seems imprudent, therefore, to reject the hypothesis that the fate of the manuscripts resubmitted by P & C was determined purely randomly. This does not, of course, prove that bias was absent, but Occam's razor bids us to reject the bias hypothesis in favor of the null hypothesis of random selection.

P & C attempted to show that the relative frequency of favorable reviews by referees and editors was significantly less for the resubmissions than for the original submissions. Using a conservative estimate of the latter, they rejected the null hypothesis of no significant difference on the basis of a chi

square test. Unfortunately, this conclusion is invalid because editors' reviews are clearly influenced by those of their referees; hence the crucial assumption of stochastic independence of observations underlying the chi square statistical model was violated in P & C's calculation. It cannot, therefore, be inferred that the resubmissions received significantly fewer favorable reviews than the original submissions, or that the observed outcome was "quite improbable," as P & C claim.

Whether referees and editors are systematically biased or operate in a quasi-random fashion, the peer-review system evidently lacks validity. When referees claim to have found serious flaws in a manuscript, therefore, there is no a priori reason to assume that they are right and the authors are wrong. If editors lack sufficient specialized knowledge to evaluate the criticisms, how ought they to respond to unfavorable referees' reports? The following procedure, which has been successfully used by the new journals *Current Psychological Research* and *Current Psychological Reviews*, seems to me to be most fair. The authors should be sent the referees' criticisms and be invited to rebut them if they consider them invalid. The original manuscript, together with the referees' criticisms and the authors' rebuttals, should then be submitted to an independent arbiter for a final verdict. This procedure would perhaps eliminate some of the more blatant injustices of the peer-review system and act as a corrective to referee error.

A. M. Colman is executive editor of *Current Psychological Research* and *Current Psychological Reviews*. Ed.

Criterion problems in journal review practices

John D. Cone

Department of Psychology, West Virginia University, Morgantown, W. Va. 26506

The provocative paper by Peters & Ceci (P & C) further documents persistent problems of unreliability and possible bias in the peer-review practices common to professional journals in the behavioral and physical sciences. As an associate editor of a psychology journal (*Behavioral Assessment*) I was surprised at P & C's finding that the same article was not recognized by the editors who had handled it just 18 to 32 months earlier. This is especially surprising in view of the manuscript's acceptance the first time around, since accepted papers are usually handled several times as they wind their way through revision, copy-editing, and final processing for publication. The forgetting of rejected manuscripts would be less surprising.

Nonetheless, the P & C results are compelling, and the editors apparently did forget. It is not the editors' faulty memory that is the primary focus of this study, nor of these comments, however. Editors of APA journals typically handle hundreds of manuscripts each year, and it is not expected, or even desirable, that they remember each one. Perhaps associate editors should be expected to do better, but even for them the implications of the P & C findings are that editorial recollection is but one element in a complex, often hastily enacted process that requires serious study and our commitment to overhauling.

Such study would begin with an analysis of the variables controlling the reviewing process. Disagreement among referees is not surprising when it is realized that reviewers, typically prominent and overextended researchers themselves, work independently, under tight time constraints, with minimal criteria to guide them, with no opportunity to question the author for clarification, with minimal feedback as to the adequacy of their reviews, and with few rewards for the long hours devoted to the process. Judgments by independent experts concerning simpler sets of stimuli than those repre-