

Is Peer Review a Game of Chance?

BRYAN D. NEFF AND JULIAN D. OLDEN

Peer review is the standard that journals and granting agencies use to ensure the scientific quality of their publications and funded projects. The peer-review process continues to be criticized, but its actual effectiveness at ensuring quality has yet to be fully investigated. Here we use probability theory to model the peer-review process, focusing on two key components: (1) editors' prescreening of submitted manuscripts and (2) the number of referees polled. The model shows that the review process can include a strong "lottery" component, independent of editor and referee integrity. Focusing on journal publications, we use a Bayesian approach and citation data from biological journals to show that top journals successfully publish suitable papers—that is, papers that a large proportion of the scientific community would deem acceptable—by using a prescreening process that involves an editorial board and three referees; even if that process is followed, about a quarter of published papers still may be unsuitable. The element of chance is greater if journals engage only two referees and do no prescreening (or if only one editor prescreens); about half of the papers published in those journals may be unsuitable. Furthermore, authors whose manuscripts were initially rejected can significantly boost their chances of being published by resubmitting their papers to other journals. We make three key recommendations to ensure the integrity of scientific publications in journals: (1) Use an editor or editorial board to prescreen and remove manuscripts of low suitability; (2) use a three-of-three or four-of-four decision rule when deciding on paper acceptance; and (3) use a stricter decision rule for resubmissions.

Keywords: Bayesian approach, citation, impact, probability, publication bias

Scientific journals operate within a highly developed peer-review system, which has been characterized as “the centerpiece of the modern scientific review process” (Glantz and Bero 1994). Nevertheless, it has long been known that the review process is not without problems (Huxley 1901, Lock 1986). Previous research has assessed reviewer integrity and the importance of journal impact factors on submission and acceptance decisions (Gordon 1977, Cole et al. 1981, Peters and Ceci 1982, Rothwell and Martyn 2000, Lawrence 2003). These studies have shown that referee reproducibility is poor—indeed, little better than chance. In a highly cited example, Cole and colleagues (1981) compared the rankings of 150 National Science Foundation grants that had been sent to two sets of referees and found surprisingly low correspondence (range of correlation across fields = 0.60–0.66). Given these findings, it is interesting that no study, to our knowledge, has addressed the importance of prescreening by editors or the number of referees needed to ensure consistency and reliability in the review process.

Nearly all biologists have considerable experience with acceptance and rejection of their manuscripts (Cassey and Blackburn 2004). In biology, statistics tell us that the more samples we collect, the more likely it is that our data set will reflect the true characteristics of the population. Peer review can be thought of in a similar way, where the “biological population” is the scientific community that is responsible for critically evaluating a given piece of research. Because not all referees are the same—for example, although some points of

concern in a review may overlap, many do not—one needs to collect enough “samples” to accurately represent the consensus of the scientific community. The minimum number of samples required to describe a population is two, but most biologists would agree that such a small sample rarely does a good job of capturing the true characteristics of the population. So why, one may reasonably ask, does the peer-review process typically poll only two referees?

In this article, we use probability theory to provide the first investigation of the effect of (a) prescreening (defined as an editor or editorial board assessing submission suitability and deciding whether or not a manuscript is sent out for review) and (b) the number of referees on the reliability of decision-making in the review process. We focus on the publication process in biology (though the results may be similarly applicable to grants), defining the “suitability” of a manuscript as the proportion of the scientific community that would deem the manuscript acceptable, possibly after revision, for publication in a given journal. “Suitability” is tractable because it captures both the scientific quality of a manuscript, as judged by the entire scientific community, and the standards of the journal. Thus, although a manuscript may be of high scien-

Bryan D. Neff (e-mail: bneff@uwo.ca) is a biologist at the Department of Biology, University of Western Ontario, London, Ontario N6A 5B7, Canada.

Julian D. Olden (e-mail: olden@wisc.edu) is a postdoctoral fellow at the Center for Limnology, University of Wisconsin–Madison, Madison, WI 53706.

© 2006 American Institute of Biological Sciences.

tific quality (e.g., appropriate experimental design, statistical analysis, interpretation of data), it may be unsuitable for a particular journal such as *Nature* or *Science* because, for example, the study is not leading edge (e.g., “sound but routine”). Alternatively, a manuscript may be of low scientific quality and unsuitable for all journals. Suitability ranges from 0 percent, where none of the scientific community deems the manuscript acceptable for publication in the targeted journal, to 100 percent, where all of the community deems the manuscript acceptable. All journals attempt to maintain high integrity of their papers, which we term “minimum standard suitability.” Thus, while impact may vary across journals, suitability should not; that is most of the community should agree that a published paper is appropriate for the particular journal. Our findings highlight the importance of prescreening and of using multiple referees for ensuring that only suitable manuscripts are published, and provide insight into how the peer-review process may be modified to ensure the scientific integrity of biological journals.

The Bayesian model

We modeled the publication process in three steps. First, manuscripts submitted to a journal are, or are not, prescreened for quality by an editor or editorial board before the decision is made to send, or not to send, the manuscript to a set number of referees for peer review. Second, if the manuscript is sent out for review, the referees recommend acceptance (possibly after revision) or rejection of the manuscript. Third, the editor accepts or rejects the manuscript on the basis of the referees’ recommendations and an acceptance decision rule. Thus, the publication fate of a manuscript depends on two key probabilities: (1) The manuscript must pass the suitability criteria of the prescreening process; and (2) the manuscript must pass the suitability criteria of the referees. The probability that a manuscript is accepted can therefore be calculated by two conditional probabilities:

$$\Pr(A | S) = \Pr(B | S) \times \Pr(A | S, B), \quad (1)$$

where $\Pr(A|S)$ is the probability that a manuscript is accepted (A) given its suitability (S), $\Pr(B|S)$ is the conditional probability that the manuscript is sent out for review (B), given its suitability (step 1 above); and $\Pr(A|S, B)$ is the conditional probability that a manuscript is accepted, given its suitability and given that it is sent out for review (steps 2 and 3 above). When no prescreening is done, all manuscripts are sent out for review and $\Pr(B|S) = 1$.

The acceptance decision rule used by editors can be defined broadly on the basis of the binomial theorem:

$$\Pr(A | S, B) = \sum_a^n \binom{n}{a} S^a \times (1 - S)^{n-a}, \quad (2)$$

where n is the number of referees and a is the minimum number of these referees who must recommend that the manuscript be accepted for it to be published. For example, when two referees are polled and the decision rule is that both

must recommend publication, then $n = 2$ and $a = 2$, and equation 2 becomes $\Pr(A|S, B) = S^2$. Editors can, of course, overrule a referee’s recommendation, but this is captured by the first conditional probability in equation 1 ($\Pr[B|S]$), albeit in this case it is performed after the refereeing.

Using the Bayesian approach, the mean suitability of published papers can then be calculated from

$$\bar{S}_A = k \cdot \int \Pr(A | S) \times \Pr(S) \times S \, dS, \quad (3)$$

where the integration is over all values of S (i.e., 0 to 100 percent), $\Pr(S)$ is the probability mass function associated with the suitability of submitted manuscripts, and k is the normalization constant. Here, for simplicity, we assume that $\Pr(S) = 1$, implying that similar numbers of manuscripts of each suitability are submitted to journals (results for a skewed distribution of submission suitability can be obtained from the authors). The mean suitability of published papers therefore depends on both the frequency with which manuscripts of varying suitability are sent out for review and the decision rule employed (equation 2).

We considered three probability functions to capture the scope of prescreening: (1) “no prescreening,” where all manuscripts are sent out for review and $\Pr(B|S) = 1$; (2) “editor prescreening,” where a preliminary prescreening process is used to remove manuscripts of particularly low suitability and $\Pr(B|S) = k \cdot S/(S + 1)$; and (3) “editorial board prescreening,” where an attempt is made to remove most unsuitable manuscripts before review, and $\Pr(B|S) = k/[1 + 100\exp(2.5 - 10S)]$. The k in each equation is a normalization constant (see figure 1 for plots of functions). We make the plausible assumption that prescreening involving multiple individuals of the editorial board will be more effective at weeding out unsuitable submissions than editor prescreening, although this will vary depending on the level of work effort and quality that the editor and editorial board invest in the project. We discuss below how journals can assess the effectiveness of their prescreening process and the specific shape of the corresponding function. We also consider a decision rule specifying that for a paper to be published, all polled referees must recommend acceptance, possibly after revisions (a typical rule used by journals).

Because only a limited number of referees are polled, the review process may have a “lottery” component. This lottery component will lead to some unsuitable manuscripts being published and, conversely, to some suitable manuscripts being rejected. The frequency of wrongful acceptance (f_{wa}) and wrongful rejection (f_{wr}) can be calculated using the following equations:

$$f_{wa} = k \cdot \int_0^{S^*} \Pr(A | S) \, dS \quad (4)$$

and

$$f_{wr} = k \cdot \int_{S^*}^1 [1 - \Pr(A | S)] \, dS, \quad (5)$$

where k represents normalization constants and S^* is the minimum standard suitability. For our analysis we set the minimum standard suitability to 80 percent, implying that if the scientific community were polled, then 8 out of every 10 referees would agree that the paper should have been published in the journal (similar results are obtained when other minimum standard suitability values, such as 70 percent or 90 percent, are considered). Based on the 80 percent value, and assuming that the probability mass function associated with the suitability of submitted manuscripts is $\text{Pr}(S) = 1$ (see above), the no-prescreening function implies that 80 percent of manuscripts submitted and sent out for review are unsuitable for publication (a value close to the actual rejection rate of manuscripts), the editor prescreening function implies that 69 percent of manuscripts submitted and sent out for review are unsuitable for publication, and the editorial board prescreening function implies that 42 percent of manuscripts submitted and sent out for review are unsuitable for publication.

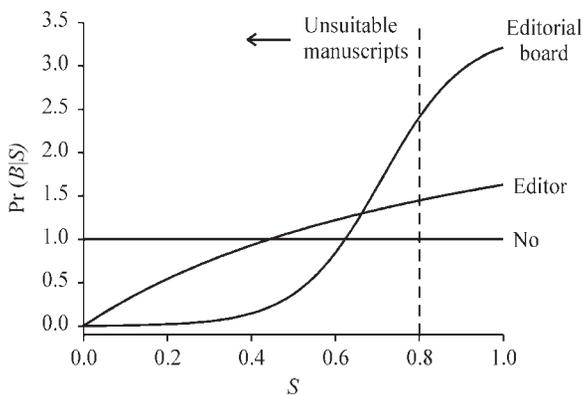


Figure 1. Functions used for the three types of prescreening, which we termed editorial board, editor, and no prescreening. The dashed vertical line represents the minimum standard suitability considered here. For journals with editorial board, editor, and no prescreening functions, 42 percent, 69 percent, and 80 percent, respectively, of manuscripts sent out for review are below this minimum level. $\text{Pr}(B|S)$ is the conditional probability that the manuscript is sent out for review, given its suitability S .

Empirical test of the model

We used two complementary approaches to examine the model predictions. First, we collected citation data for papers published in the year 2000 for 14 biological and ecological journals from the ISI Web of Science (<http://scientific.thomson.com/products/wos>). We recorded the number of times each paper was cited from the time it was published to June 2003 (ca. a 2.5-year period), and calculated the mean and coefficient of variation in the number of citations for each journal. We used the coefficient of variation in citations as one

measure of journal integrity because it controls for the profile of the journal (i.e., the actual number of citations) and instead focuses on variation in citation frequency. We expect journals that are successful at publishing primarily papers that meet their minimum suitability requirement to show consistency in the number of times each paper is cited. In contrast, journals that publish a substantial number of papers that are below their minimum suitability requirement should show less consistency in the number of times each paper is cited, because the papers with low suitability will be cited less frequently. Second, we contacted the editor in chief of each journal and requested the following information: (a) the percentage of submitted manuscripts that were typically sent out for review; (b) the percentage of submitted manuscripts accepted for publication; (c) the typical number of referees (excluding editors and associate editors); and (d) the use of a prescreening process, which was classified as “editorial board” when two or more individuals were involved, as “editor” when only one individual was involved (usually the editor or associate editor), or as “no” when all manuscripts were sent out for review (barring technical or formatting issues).

Results from the Bayesian model

Regardless of the prescreening process, increasing the number of referees increases the mean suitability of published papers by decreasing the chance that an unsuitable manuscript is wrongfully accepted (figure 2). Such errors occur when the referees selected to review a manuscript believe it should be published when in fact the community at large does not. For example, a manuscript of 50 percent suitability would be unsuitable on the basis of the 80 percent minimum standard of suitability, even though half of the community believes the paper is acceptable. Subsequently, there is a 25 percent chance that these favorable referees will be asked to review the manuscript when only two are polled.

Conversely, increasing the number of referees increases the probability that suitable manuscripts are wrongfully rejected (figure 2). Given a manuscript of 80 percent suitability, which meets our minimum standard, there is a 36 percent chance that at least one unfavorable referee will be polled when two are used. Thus, although increasing the number of referees helps to ensure that only suitable manuscripts are published, it also leads to a greater number being wrongfully rejected. For example, when there is no prescreening and two referees are used, 51 percent of all published papers fall below the minimum standard, whereas about 6 percent of manuscripts that fall above the minimum standard are wrongfully rejected (figure 2a, table 1). Furthermore, the mean suitability of published papers is actually below the desired minimum standard at 75 percent, and the 10th percentile is only 46 percent. The frequency of wrongful acceptance and wrongful rejection can be optimized at about eight referees, at which point the frequency of each is about 12 percent. In this case, the mean suitability of published papers would be

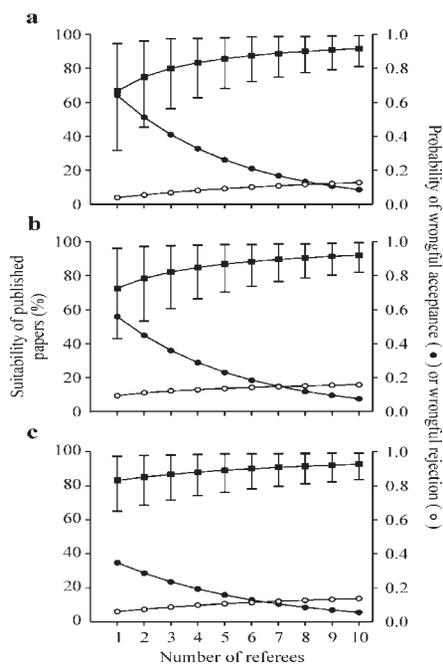


Figure 2. Probability analysis of the peer-review process. Panels show the probabilities of wrongful acceptance (filled circles), the probabilities of wrongful rejection (open circles), and the mean suitability and 80 percent confidence interval (filled squares) of published papers in journals with (a) no prescreening, (b) editor prescreening involving one individual, and (c) editorial board prescreening involving two or more individuals.

90 percent, and the suitability of the 10th percentile would be 79 percent.

When using an editor prescreening process that attempts to remove manuscripts of particularly low suitability before review, we found that the mean suitability of published pa-

pers increased, particularly when only a few referees were used (figure 2b). This is the result of the decreased frequency with which papers of low suitability are sent out for review and ultimately accepted (69 percent instead of 80 percent). For example, when two referees are used, 12 percent fewer manuscripts are wrongfully accepted. The frequency of wrongful acceptance could be reduced 18 percent by employing a three-of-three decision rule. Importantly, in the latter case the mean suitability of published papers rises above the minimum standard at 82 percent (table 1).

When using an editorial board prescreening process that attempts to remove most unsuitable manuscripts before review, we found that the mean suitability of published papers increased as compared with either editor prescreening or no prescreening (figure 2c). When two referees are used, the frequency of wrongfully accepted papers is 29 percent and the 10th percentile of the suitability of accepted papers is 69 percent. By employing three referees and a three-of-three decision rule, the frequency of wrongfully accepted papers can be decreased by 17 percent, to about a quarter (table 1).

Empirical test

We compared the results of our model with citation data for papers published in the year 2000 for 14 representative journals (table 2). Three lines of evidence support the model predictions. First, we found that the coefficient of variation in citation frequencies was lowest for journals using an editorial board prescreening process, intermediate for journals using an editor prescreening process, and highest for journals not using a prescreening process (table 2; ANOVA $F_{2,11} = 8.45, p = .006$). Second, journals employing a two-of-two decision rule accepted a significantly higher percentage of submitted manuscripts ($t_{11} = 2.45, p = .03$) and exhibited a marginally larger coefficient of variation in citations ($t_{12} = 1.77, p = .10$) compared with journals employing a three-of-three deci-

Table 1. Mean suitability of accepted papers and probability of wrongful acceptance and rejection for different levels of submission prescreening and for three decision rules.

Prior probability of suitability	Suitability value (80 percent confidence interval) or probability		
	Two-of-two decision rule	Three-of-three decision rule	Four-of-four decision rule
<i>Editorial board prescreening</i>			
Mean suitability of accepted papers	85 (69–98)	87 (72–98)	88 (74–98)
Probability of wrongful acceptance	0.29	0.24	0.19
Mean suitability of rejected papers	41 (8–76)	43 (8–78)	44 (9–80)
Probability of wrongful rejection	0.07	0.09	0.10
<i>Editor prescreening</i>			
Mean suitability of accepted papers	78 (53–97)	82 (61–98)	85 (66–98)
Probability of wrongful acceptance	0.45	0.36	0.29
Mean suitability of rejected papers	45 (9–85)	46 (9–86)	47 (9–86)
Probability of wrongful rejection	0.14	0.14	0.15
<i>No prescreening</i>			
Mean suitability of accepted papers	75 (46–96)	80 (56–97)	83 (63–98)
Probability of wrongful acceptance	0.51	0.41	0.33
Mean suitability of rejected papers	38 (7–73)	40 (8–76)	42 (8–78)
Probability of wrongful rejection	0.06	0.07	0.08

Note: Decision rules indicate the number of polled referees whose recommendation is needed to accept a manuscript. In a two-of-two decision rule, for example, both polled referees must recommend acceptance of a manuscript for it to be published.

Table 2. Summary of citation data collected from the ISI Web of Science in June 2003 for papers published in the year 2000.

Journal	2001 ISI impact factor	Mean number of citations	Coefficient of variation in citation data	Percentage of manuscripts sent for review	Percentage accepted	Typical number of referees
<i>Editorial board prescreening</i>	20.9	42.2	35.4	30.0	17.4	2.8
Cell	29.2	60.0	24.9	–	–	3
Nature	28.0	50.6	49.8	35.0	9.2	3
Proceedings: Biological Sciences	3.2	7.4	40.7	–	35.0	2
Science	23.3	50.6	26.4	25.0	8.0	3
<i>Editor prescreening</i>	7.2	14.8	47.2	82.1	31.5	2.1
American Naturalist	4.3	9.2	30.7	80.2	25.2	2
Biological Journal of the Linnean Society	2.3	3.8	57.9	95.0	50.0	2
Canadian Journal of Zoology	1.2	2.5	69.6	90.0	55.0	2
Ecology	3.7	8.3	37.7	85.0	30.0	2
Evolution	3.7	7.8	45.2	91.5	26.5	2
Molecular Ecology	2.5	6.0	57.2	83.0	35.0	2
New England Journal of Medicine	29.1	56.8	50.6	50.0	9.5	2
Proceedings of the National Academy of Sciences	10.9	24.5	28.9	–	20.5	3
<i>No prescreening</i>	1.0	2.0	99.8	85.0	60.0	2.0
Evolutionary Ecology Research	1.8	3.4	68.6	80.0	50.0	2
Journal of Freshwater Ecology	0.2	0.7	130.9	90.0	70.0	2

Note: Editorial board prescreening involves two or more individuals assessing the suitability of manuscripts before peer review, editor prescreening involves one individual, and no prescreening involves sending all manuscripts out for review (barring formatting issues). Values in bold represent averages for the three types of prescreening, and a dash indicates that data were not provided or were otherwise unavailable.

sion rule (figure 3a). These patterns are consistent with model predictions showing that variation in manuscript suitability and probability of wrongful acceptance are higher when using a two-of-two as opposed to a three-of-three decision rule (table 1). Comparisons including only those journals that use an editor prescreening process showed that *Proceedings of the National Academy of Sciences* is the only journal consistently employing three referees, and it exhibits the lowest coefficient of variation in citations (table 2). Third, we found a positive relationship between the percentage of manuscripts sent out for review (perhaps an indication of the effectiveness of the prescreening process at removing unsuitable manuscripts) and the coefficient of variation in citations, albeit this result was not significant ($F_{1,9} = 2.58, p = .14$, figure 3b).

Because of the apparent lottery component of the review process, repeat submission could be an effective strategy for publishing unsuitable manuscripts (figure 4). For example, when two referees are used, 44 percent of all manuscripts of only 50 percent suitability will be published in either the first or the second journal to which they are submitted (even assuming no modifications to increase the manuscript's suitability between submissions). When six journals are targeted sequentially, there is a better than 80 percent chance.

Implications for the peer-review process

Peer review is a valuable process that is central to ensuring scientific quality, yet it continues to be scrutinized in the natural sciences (Gura 2002) and has been a topic of recent interest in biology (Tregenza 2002, Cassey and Blackburn 2003, 2004, Grimm 2005, Leimu and Koricheva 2005). Research has shown that peer review can be sexist (Grant et al. 1997),

nepotistic (Forsdyke 1993), and biased with respect to the national language of the authors (Bakewell 1992) and the prestige of the authors' institutions (Garfunkel et al. 1994). This is true of ecology (Tregenza 2002) and the natural and medical sciences (Wennerås and Wong 1997, Jefferson et al. 2002). Here we provide one of the first studies to quantify the lottery component in the review process and show how manuscript prescreening, the number of referees, and the decision rule used by journals can influence the consistency in quality of published papers. It is not our intent to imply that editors do not also play a critical role in ensuring publication quality. Indeed, a premise of our model is that *all* editors are consistent in their intent to promote the scientific integrity of published papers, and we have shown that the ability of editors or boards to remove unsuitable manuscripts from the review process (or, equivalently, to overrule poor recommendations by referees) is a fundamental component of the review process.

Our study attempts to build on the idea that we need to improve the peer-review process by illustrating quantitatively that the degree of prescreening and the number of referees positively affect manuscript suitability, primarily by minimizing wrongful acceptance of unsuitable submissions. This finding is consistent with citation data, which show that journals sending a greater proportion of manuscripts out for review have greater variation in the quality of the papers published (as indicated by greater coefficient of variation in citation frequency). Moreover, we show the advantages of employing more stringent decision rules in the review process, a result also supported by the fact that journals employing a two-of-two decision rule accept a greater proportion of manuscripts

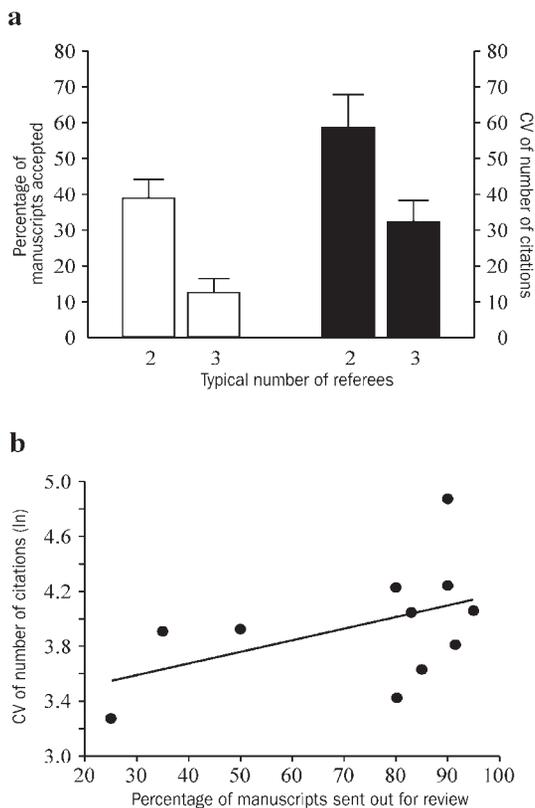


Figure 3. (a) Comparisons of the percentage of received manuscripts that were accepted (left side of figure) and coefficient of variation (CV) in the number of citations (right side of figure) for biological journals employing two-of-two or three-of-three decision rules. (b) Relationship between the percentage of received manuscripts that were sent out for review and the coefficient of variation in the number of citations (ln-transformed). Data are from table 2.

and have greater variation in publication quality than journals employing a three-of-three decision rule.

Our study and previous research raise an important question: Is there any way we can improve the peer-review process? Clearly the time is ripe for such a discussion, as indicated by the numerous journal editorials and international congresses focusing on peer review (e.g., the Fifth International Congress on Peer Review and Biomedical Publication; www.ama-assn.org/public/peer/peerhome.htm). The lottery problem illustrated in our study is not likely to be as complicated or sensitive as the plethora of other biases associated with the peer-review process, such as nepotism or sexism (Wennerås and Wong 1997), or even plagiarism (Marshall 1998). We argue that initiating a prescreening process, ideally involving several highly qualified individuals, and increasing the number of referees for ecological journals is a good starting point for improving the peer-review process. For example, our model shows that there are considerable benefits to employing three or four referees instead of just two, particularly for journals

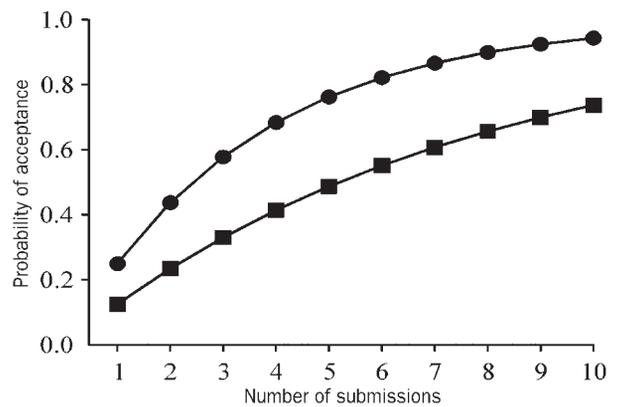


Figure 4. Relationship between the number of submissions and the probability of acceptance for a two-of-two decision rule (filled circles) and a three-of-three decision rule (filled squares), assuming an unsuitable manuscript of 50 percent suitability.

that do not use a prescreening process. Our model can be used by journals to assess the optimal approach, such as using more referees or improving the effectiveness of prescreening, to achieve a desired minimum suitability of their published papers.

The requirement of polling more referees may be a contentious issue because of the limited rewards for professionals who conduct reviews. Given that referees spend an average of 1.5 hours reviewing a manuscript (Lock and Smith 1990) and review an average of one manuscript per month (Yankauer 1990), moving from two to three referees would require an additional 45 minutes per month per referee. These estimates vary considerably, and thus the increase in referee time could be substantially more for some referees. However, we have also recommended that an editorial board (or the equivalent) be implemented to perform detailed prescreening. Editors typically are among the top researchers in their field of expertise, and as such may be able to assess the suitability of manuscripts much more efficiently than the average referee. Constructing editorial boards would most likely require some incentive for board members and may not be an option for all journals, but we argue that it is an important component of an effective peer-review process. If editorial boards are able to remove a majority of unsuitable manuscripts before review (for example, in our editorial board probability function, we assumed the board could remove 58 percent of unsuitable manuscripts), then the referee community will be less burdened with unsuitable manuscripts and more referees can be used. Generally, editors should be more efficient at identifying suitable manuscripts, in part because they are more familiar with the standards and scope of their journal. Furthermore, the manuscripts sent to referees would be of higher quality, and referees' time would be spent less on deciding the suitability of a manuscript (i.e., refereeing) and more on providing detailed comments on the study and thereby improving the manuscript.

There are several potential caveats to our modeling approach. First, defining a manuscript's or grant's suitability as a proportion of the scientific community that would deem it acceptable for publication or funding could bias against high-risk, high-payoff (i.e., "out of the box") research in favor of mainstream research. Such a bias could emerge if referees are prejudiced in favor of a particular scientific outcome or avenue of research (e.g., based on the paradigm they believe has the most merit; see Alatalo et al. 1997, Jennions and Møller 2002). For manuscript reviews, it is important that referees evaluate the scientific process and not the scientific outcome. To this end, journals should be active in removing individuals from their refereeing community who have demonstrated a prejudice. Second, the coefficient of variation in the number of citations may be a poor measure of journal integrity if it is confounded by the scope of the journal. For example, journals with broader scope may also have greater variation in citation numbers because of the different disciplines covered (although the data in table 2 appear inconsistent with this possibility). Third, our recommendation to increase the number of referees polled for each submission is likely to lead to an increase in the number of manuscripts wrongfully rejected. For example, when an editorial board is used, our model suggests that the probability of wrongful rejection could increase from 7 percent to 10 percent when four referees are used instead of only two (see table 1). This 3 percent increase needs to be evaluated in comparison to the predicted 19 percent reduction in wrongful acceptance (i.e., a reduction from 29 percent to 10 percent for four rather than two referees; see table 1). It is possible that wrongful rejection carries a greater cost to the advancement of science than does wrongful acceptance.

Testing the success of any change to peer-review practices requires that journals adopt a modified review process for a limited period (for example, a single year) and closely monitor the outcome. Ideally, journals would implement two methods simultaneously (e.g., a new and an old method) and randomly assign manuscripts to one or the other method. Societies that publish multiple journals could conduct different peer-review strategies for different journals. Tracking the outcome of the manuscripts, such as review time and subsequent citation record, would then provide a direct comparison of the utility of the two methods. The effectiveness of a particular prescreening practice, and the shape of the actual probability function used in our model, could be assessed by having the editor or editorial board rank manuscripts according to their suitability (i.e., identify what proportion of the referee community they think will recommend acceptance, possibly after revision, of the manuscript in the targeted journal) and then send all of the manuscripts out to 10 or more referees. A journal with a good prescreening practice should be able to identify most manuscripts that fall below some minimum level of suitability. The flexibility of journals in undertaking such experiments in peer review has been illustrated in the past (e.g., *Interfaces*; Armstrong 1982), yet, perhaps surprisingly, such experiments are rarely conducted.

A striking result from our study was that the lottery component of the review process creates the opportunity for unsuitable manuscripts to be published by repeated submission. For example, when six journals are sequentially targeted, there is a better than 80 percent chance that a manuscript of low suitability—one that only half the scientific community would recommend acceptance—is accepted. These results are supported by data showing that over 90 percent of the papers rejected by one journal are eventually published—some even unaltered—and not necessarily in a lower-tier journal (Wilson 1978). This undesirable outcome could be circumvented if a more strict decision rule were employed for repeat submissions, or if the authors of repeat submissions were required to submit the previous reviews and their responses to the reviewers (see the discussion in Rissgård 2003). Reporting integrity could be maintained if editors implemented a full-disclosure strategy in which they recorded and cross-referenced key information, such as title, authors, and abstract, of rejected manuscripts on a central repository. Our results on repeat submissions highlight a potential negative consequence of Carlstedt's (2002) recommendation that simultaneous submission of a manuscript to multiple journals be permitted to expedite the review process. Of course, any measure to prevent the publication of unsuitable manuscripts must be tempered by the fact that there can be a large number of manuscripts that are wrongfully rejected (our model suggests this number could be as high as 14 percent).

In conclusion, our paper makes three key recommendations to ensure the integrity of scientific publications in journals: (1) Use an editor or editorial board to prescreen and remove manuscripts of low suitability, (2) use a three-of-three or four-of-four decision rule when deciding on the acceptance of papers, and (3) use a stricter decision rule for repeat submissions. Implementation of these recommendations should help add to the integrity of the peer-review process and, ultimately, ensure publication quality in biology.

Acknowledgments

We thank Beth MacDougall-Shackleton, Steve Ormerod, Trevor Pitcher, and three anonymous reviewers for helpful comments, and Karrienne DeBaeremaeker for assistance in collecting the citation data. We are grateful to the editors in chief of the journals for providing submission and rejection data. This work was supported by the Natural Sciences and Engineering Research Council of Canada to B. D. N. and J. D. O., and a David H. Smith postdoctoral conservation fellowship to J. D. O.

References cited

- Alatalo RV, Mappes J, Elgar MA. 1997. Heritabilities and paradigm shifts. *Nature* 385: 402–403.
- Armstrong JS. 1982. Is review by peers as fair as it appears? *Interfaces* 12: 62–74.
- Bakewell D. 1992. French research—publish in English or perish. *Nature* 356: 648.
- Carlstedt RA. 2002. Cortex forum on peer-review multiple submissions. *Cortex* 38: 411.

Cassey P, Blackburn TM. 2003. Publication rejection among ecologists. *Trends in Ecology and Evolution* 18: 375–376.

———. 2004. Publication and rejection among successful ecologists. *BioScience* 54: 234–239.

Cole S, Cole JR, Simon GA. 1981. Chance and consensus in peer review. *Science* 214: 881–886.

Forsdyke DR. 1993. On giraffes and peer review. *Federation of American Societies for Experimental Biology* 7: 619–621.

Garfunkel JM, Ulshen MH, Hamrick HJ, Lawson EE. 1994. Effect of institutional prestige on reviewers' recommendations and editorial decisions. *Journal of the American Medical Association* 272: 137–138.

Glantz SA, Bero LA. 1994. Inappropriate and appropriate selection of 'peers' in grant review. *Journal of the American Medical Association* 272: 114–117.

Gordon M. 1977. Evaluating the evaluators. *New Scientist* 73: 342–343.

Grant J, Burden S, Breen G. 1997. No evidence of sexism in peer review. *Nature* 390: 438.

Grimm D. 2005. Suggesting or excluding reviewers can help get your paper published. *Science* 309: 1974.

Gura T. 2002. Peer review—unmasked. *Nature* 416: 258–260.

Huxley L. 1901. *Life and Letters of Thomas Henry Huxley*. New York: Appleton.

Jefferson T, Alderson P, Wager E, Davidoff F. 2002. Effects of editorial peer review—a systematic review. *Journal of the American Medical Association* 287: 2784–2786.

Jennions MD, Møller AP. 2002. Relationships fade with time: A meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings: Biological Sciences* 269: 43–48.

Lawrence PA. 2003. The politics of publication. *Nature* 422: 259–261.

Leimu R, Koricheva J. 2005. What determines the citation frequency of ecological papers? *Trends in Ecology and Evolution* 20: 28–32.

Lock S. 1986. *A Difficult Balance—Editorial Peer Review in Medicine*. Philadelphia: ISI Press.

Lock S, Smith J. 1990. What do peer-reviewers do? *Journal of the American Medical Association* 263: 1341–1343.

Marshall E. 1998. The Internet: A powerful tool for plagiarism sleuths. *Science* 279: 474.

Peters DP, Ceci SJ. 1982. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences* 5: 187–195.

Rissgård HU. 2003. Misuse of the peer-review system: Time for counter-measures? *Marine Ecology Progress Series* 258: 297–309.

Rothwell PM, Martyn CN. 2000. Reproducibility of peer review in clinical neuroscience: Is agreement between reviewers any greater than would be expected by chance alone? *Brain* 123: 1964–1969.

Tregenza T. 2002. Gender bias in the refereeing process? *Trends in Ecology and Evolution* 17: 349–350.

Wennerås C, Wong A. 1997. Nepotism and sexism in peer-review. *Nature* 387: 341–343.

Wilson JD. 1978. Peer review and publication. *Journal of Clinical Investigation* 61: 1697–1701.

Yankauer A. 1990. Who are the peer reviewers and how much do they review? *Journal of the American Medical Association* 263: 1338–1340.



TC5000

New from Meiji Techno

Designed with Meiji's ICOS Infinity Corrected Optical System - the new TC Series of Inverted Microscopes by Meiji Techno incorporates world class optics in a cost effective platform offering a higher standard in specimen observation.

With a host of new features and options, the TC Series makes cell checking faster, clearer and easier than ever before.

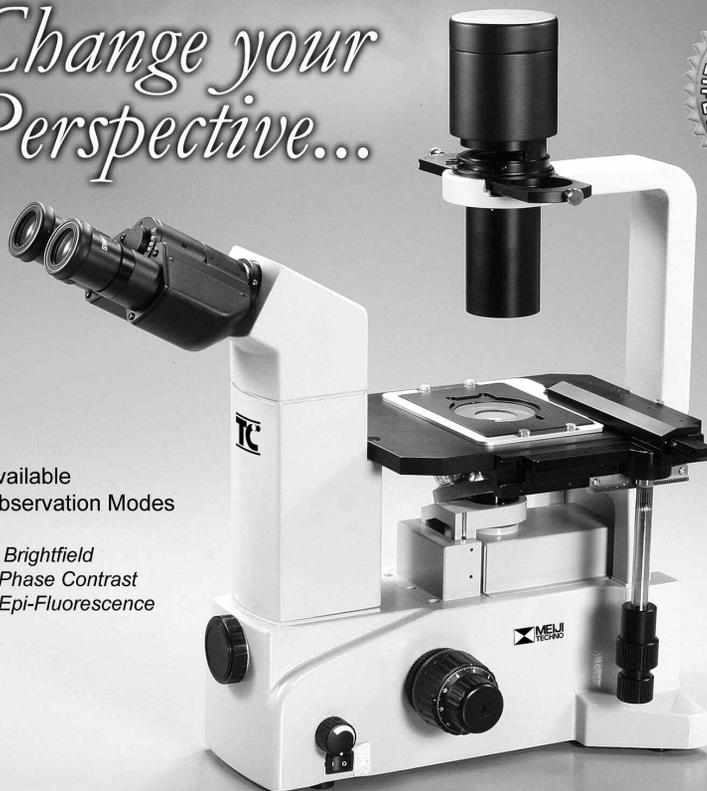
Ask your Meiji dealer for a demonstration today!

www.meijitechno.com

Change your Perspective...

Available Observation Modes

- Brightfield
- Phase Contrast
- Epi-Fluorescence



Meiji Techno America 3010 Olcott St., Santa Clara, CA 95054-3207
1-800-832-0060 toll free or visit us on the web at: www.meijitechno.com