CHAPTER

5

Blind Peer Review by Academic Journals

Emily A. Largent[1], *Richard T. Snodgrass*[2]

[1]Program in Health Policy, Harvard University, Cambridge, MA, USA;

[2]Department of Computer Science, The University of Arizona, Tucson, AZ, USA

OUTLINE

Introduction 75

Overview of Peer Review 77
- *Process of Peer Review* 77
- *Types of Peer Review* 78
- *A Note on the Evidence* 78

Double-Blinding as a Means of Enhancing Fairness 79
- *Fairness to Authors from Less-Prestigious Institutions* 79
- *Fairness to Non-US Authors* 81
- *Fairness to Prolific Authors* 81
- *Fairness to Female Authors* 82
- *The Perception of Fairness* 83

Blinding as a Means of Improving the Quality of Reviews 84
- *Effects of (Not) Revealing Author Identity* 84
- *Effects of (Not) Revealing Reviewer Identity* 86

Breaking the (Double) Blind 87
- *How Double-Blinding Is Accomplished* 87
- *Blinding Success Rates* 87
- *Impact of Technology on Blinding* 89

Preferences for Open, Single-, or Double-Blind Review 90

Conclusion 92

References 93

INTRODUCTION

Peer review, "the evaluation of research findings for competence, significance, and originality by qualified experts" (Benos et al., 2007, p. 145), is the gold standard for assessment of manuscripts submitted to journals.[1] "Peer review was originally conceived to provide

[1]This chapter is based, in part, on prior work by Richard T. Snodgrass (2006, 2007).

http://dx.doi.org/10.1016/B978-0-12-802460-7.00005-X

advice for the editor, the equivalent of asking the knowledgeable colleague down the hall for an opinion. By the 1960s and 1970s, however, it had come to be the measure of quality for journals—high-quality journals use strong peer-review systems" (Bordage and Caelleigh, 2001, p. 904). Correspondingly, publications "are not held in high esteem if they do not first pass through this process [of peer review]" (Benos et al., 2007, p. 146).

The stakes of the peer review process are quite high for scholars, scholarship, and society. Fletcher and Fletcher (1997) explain:

> Progress in science depends on both the dissemination of ideas and evidence as well as on vigorous critique by peers, both of which take place in journals. In basic science, inefficiency results from conclusions that exceed or fall short of study results, or are difficult to understand, as colleagues' time and money is invested in building on or refuting incorrect results. In applied fields such as medicine and engineering, people can be harmed if the information in journals is wrong, and clinicians are misled about diagnosis and treatment, or engineers are misinformed regarding the risks of buildings, bridges, and nuclear energy plants. Also, scientists' careers, including grant awards, promotion and reputation, depend on their success as authors. (pp. 36–37)

While Fletcher and Fletcher frame their discussion in terms of the sciences, their observations are applicable to other academic disciplines as well. Peer review outcomes—that is, which articles are ultimately published—affect individual authors, journal readers, and the course of future research (Hojat et al., 2003).

At present, although peer review is generally accepted as an essential element in the pursuit of knowledge, as one that imparts legitimacy and credibility, considerable debate persists on how best to implement it. Richard Smith, former editor of the *British Medical Journal*, has enumerated the faults of peer review: "In addition to being poor at detecting gross defects and almost useless for detecting fraud, it is slow, expensive, profligate of academic time, highly subjective, something of a lottery, prone to bias, and easily abused" (Smith, 2006, p. 179). Nevertheless, Smith, like others, argues that the correct approach is not to abandon peer review but to identify ways to improve it, ideally through evidence-based interventions. It is, he says, "a system full of problems but the least worse we have" (Smith, 2006, p. 178).

Given the centrality of peer review to the production and propagation of knowledge, one might be surprised to learn that research on the efficacy of peer review was relatively scarce prior to 1975 and that clear-cut evidence is still lacking today. What discussion existed before 1975 was based more on anecdote than on systematically gathered data. This changed when, in 1974, the then-editor of the *New England Journal of Medicine* called for research into peer review, arguing that "data were needed to justify the American process that 'faithfully and meticulously depended on a system of peer review'" as compared to the process at the prestigious British medical journal *Lancet*, which did not (Weller, 2001, pp. 10–11). Although research into peer review was subsequently conducted (Armstrong, 1997; Campanario, 1998a,b), a Cochrane review published in 2007 concluded that "little empirical evidence is available to support the use of editorial peer review" (Jefferson et al., 2007, p. 2) and urgently called for the launch of a "large, well-funded programme of research" into its effects (Jefferson et al., 2007, p. 2). Further inquiry, descriptive and normative, is doubtlessly still needed.

In this chapter, we will examine the extant scholarly literature related to one aspect of the peer review practices of academic journals: blinding, specifically, attempting to mask the identity of authors to reviewers and/or of reviewers to authors. A central critique of peer review is that it is subjective, overly dependent on the opinions of individual reviewers.

This subjectivity leads to inconsistency in publication recommendations and introduces the possibility of bias. Types of bias relevant to peer review include gender bias, racial bias, status bias (the tendency to more favorably evaluate manuscripts submitted by famous authors from prestigious institutions), ideological bias (the tendency to favor a particular ideological or theoretical orientation to research and analysis), bias against negative results (the tendency to reject those studies with negative findings), and confirmatory bias (the "tendency of some reviewers to accept outcomes that agree with commonly accepted theories and to discredit those that do not") (Hojat et al., 2003, p. 78). Blinding is commonly advocated as an approach to alleviate the effects of biases. Blind peer review may help to mitigate some of these effects, but clearly, it will be extremely difficult (if not impossible) to address all of the just-enumerated biases through blinding.

We begin this chapter by providing a brief background on peer review. We then examine empirical evidence related to (1) blinding as a means of reducing bias; (2) blinding as a means of yielding higher quality reviews; (3) the feasibility and efficacy of blinding; and (4) stakeholder (i.e., author, reviewer, editor) preferences for blinding. Although the evidence for the first three is generally inconclusive, one finds widespread support for blinded peer review. Ultimately, we argue that, as the evidence currently stands, a determination that blinding is "worth it" requires weighing the acknowledged weaknesses of blinding against gains in the perceived fairness of peer review.

OVERVIEW OF PEER REVIEW

Peer review is intended to improve published manuscripts in at least two ways. First, it helps editors winnow out the best papers—in terms of both overall quality and appropriateness for a particular journal—from all of those submitted. Second, by offering comments both minor and substantive, it helps authors improve their papers, with the relevant baseline being what would have been published had the manuscript simply been published as submitted. To the extent that it succeeds in these two aims, peer review should improve the overall quality of the scholarly literature.

While journals in all branches of academe share "the core ethos and values of peer review, it has evolved in diverse ways to best fit the environments and circumstances of the various sciences and disciplines" (Bordage and Caelleigh, 2001, p. 904).

Process of Peer Review

Although peer review processes vary from journal to journal, there are common features. Editors can summarily reject a manuscript upon submission. For those manuscripts that survive this first cut, the editor typically selects two or more reviewers to whom the manuscript is sent. "Reviewers who are experts in some aspect of the submitted work (its contents, methods, or importance), are external to the staff of the journal itself, and are free of financial conflict of interest, provide not only their opinion about whether the work should be published but also suggestions for both editors and authors about how the written presentation could be improved" (Fletcher and Fletcher, 1997, p. 36). After receiving the completed peer reviews and advice from the journal's editorial staff, an editor will decide whether to reject, request

a revision, or publish a particular manuscript. Given that the rejection rate for articles submitted to peer-reviewed journals is quite high, as is the rate of acceptance made conditional upon revisions in accordance with reviewers' comments, it is unsurprising that peer reviewers have been described as "gatekeepers" (Bradley, 1981, p. 31; Hojat et al., 2003) to publication.

Types of Peer Review

There are two general types of peer review: *open* or *blind*.

In open peer reviewing, the identities of both the author(s) and reviewers (sometimes called "referees") are revealed. Reviewers will be aware of who authored a given manuscript, and authors will be able to attribute the feedback they receive to particular individuals. Open peer review has a number of possible variations. For instance, the names of the reviewers—and even their reviews—might be published alongside the final paper.

Within the category of blind peer reviewing, a further distinction is drawn between *single-blind* and *double-blind* peer review. In single-blind review, the reviewer knows the identity of the author, while the reviewer is anonymous to the author. In double-blind review, both the reviewer and the author are anonymous to each other. We will merely note that some have argued for adoption of so-called *triple-blind* peer review in which neither the reviewer nor the editor would be privy to identifying information about the author (Jump, 2015).

Regarding terminology, we wish to acknowledge that some people are highly offended by the use of "blindness" as a rhetorical device to signify lack of knowledge and find it demeaning to people with disabilities. They have, therefore, lobbied journals to eschew the phrase "blind review" and suggest use of neutral but equivalent phrases such as "anonymous review" or "masked review" instead, a practice some journals and professional associations have adopted (Tremain, 2011). While sympathetic to this position, we have retained the traditional terms, as they remain widespread and will be most familiar to our readers.

Disciplines vary widely in their use of single- and double-blind review. A survey of 553 journals selected from 18 disciplines revealed that "[a]cross the disciplines, the majority of surveyed journals used double-blind reviews (58%), 37% employed single-blind, and only 5% made use of open review" (Bachand and Sawallis, 2003, p. 54).

A Note on the Evidence

Because "[p]eer review is not a discipline-specific field, literature on the subject could and does exist in almost every scholarly field with a journal publication outlet" (Weller, 2001, p. 8). That fact has two related, though distinct, implications for our present undertaking. First, it is useful and instructive to learn what a range of disciplines, using diverse approaches, have discovered about blind review. Second, rather than presenting an exhaustive summary of the literature on blinding in peer review, we have chosen to group studies thematically to facilitate exploration of the broad issues that cut across disciplines: whether blinding a reviewer to an author's identity enhances fairness; whether blinding authors and/or reviewers enhances the quality of reviews; and finally, whether blinding a reviewer to the identity of an author is even feasible. This thematic approach is also pragmatic: because the relevant studies ask slightly different questions, use different designs, and have different outcome measures, there is a barrier to conducting formal *meta*-analysis (Jefferson et al., 2007).

We caution that many of the studies presented below have limitations that threaten their internal and external validity. In studies that used nonrandomized designs, for instance, findings may be confounded by other factors such as the quality of the submitted manuscripts. A majority of the studies have small sample sizes, and in some instances, confidence intervals are not sufficiently narrow to draw definite conclusions. A common methodological issue for studies examining the effects of blinding on fairness and quality is the difficulty of ensuring robust blinding. Additionally, the character of a particular academic discipline or of the journal(s) at which a study was conducted may limit generalizability. We flag these concerns in a general way here so that readers can keep them in mind; when appropriate, further limitations of particular studies will be noted below.

DOUBLE-BLINDING AS A MEANS OF ENHANCING FAIRNESS

Ideally, the determination that a manuscript will or will not be accepted for publication should be made on the basis of the manuscript alone: is what the submission states correct, insightful, and an advancement of the state-of-the-art? "Bias occurs if a review is influenced by other criteria, such as geographic location or institutional prestige" (Ross et al., 2006, p. 1675). The fundamental argument in favor of double-blind review is that when a manuscript is so blinded, reviewers cannot take irrelevant criteria (i.e., their biases) into account, and so judge the manuscript on its own merits. Advocates of blinding suggest that this will both improve the accuracy of the peer review process by allowing the best work to shine and promote fairness to authors.

Any number of an author's personal and professional attributes could be subject to bias. We will consider: affiliation with a less-prestigious institution; non-U.S. origins; prolificness; and gender. Finally, we will consider the effect of blinding on perceived fairness.

Fairness to Authors from Less-Prestigious Institutions

Some evidence from retrospective and experimental studies suggests that when reviewers know authors' names and affiliations, they may exhibit bias against papers from unknown authors or less-prestigious institutions. The "inclination to give disproportionate credit to those who are already famous" (Hojat et al., 2003, p. 79) is variously called the "Matthew effect" or "status bias" (Cox et al., 1993).

An anecdote illustrates the possibility. The psychologist Robert Rosenthal has related how he was unable to publish between 15 and 20 manuscripts that he wrote while at the University of North Dakota (UND) (Rosenthal, 1982). Rosenthal continues:

> After I had been at Harvard a few years, most of those same articles *were* published in mainstream journals. My anecdote does not demonstrate that journal articles were biased against papers from UND and biased toward papers from Harvard. There are plausible rival hypotheses that cannot be ruled out. My belief, however, is that location status bias may well have played some role in the change in publishability of my stack of papers. (p. 235)

With this anecdote in mind, we now examine the studies that attempt to detect status bias.

Several retrospective studies have found evidence of status bias. One found that "some 91 per cent of the papers by physicists in the foremost departments were accepted [for publication at *The Physical Review*] as against 72 per cent from other universities" (Zuckerman and Merton, 1971, p. 85). A retrospective study of reviews for journals in the physical sciences found large differences in how papers from more- and less-prestigious universities were reviewed (Gordon, 1980). The investigator concluded that "the scientific predispositions of referees…bias them toward less critical evaluation of colleagues who come from similar institutional…groups" (Gordon, 1980, pp. 274–275). A third retrospective study, this one of single-blind reviews for the *Journal of Pediatrics*, found partial evidence for status bias (Garfunkel et al., 1994). Lower institutional rank was associated with lower rates of recommendation for acceptance for some types of papers but not others (Garfunkel et al., 1994).

By contrast, Mahoney et al. (1978) performed an early experiment in which institutional affiliation was manipulated so that half of the manuscripts sent to 68 volunteer reviewers allegedly came from a prestigious university, while the other half allegedly came from a relatively unknown college. They found that "the effect of institutional prestige failed to attain significance" (p. 70) on how reviewers rated various aspects of the manuscripts or on their summary recommendation regarding publication.

Two additional experiments bear discussing at greater length. First, Peters and Ceci (1982) performed a famous experiment in which 12 papers published by investigators from prestigious American psychology departments in high-quality journals were altered to include fictitious names and less-prestigious institutional affiliations. The altered papers were then formally resubmitted to the same journals that had published them just 18–32 months prior. Remarkably, only three of the manuscripts were detected as resubmissions. Of the remaining nine manuscripts, eight were rejected, in many cases based on the presence of "serious methodological flaws." In their analysis, Peters and Ceci (1982) put forth the possibility of status bias: "The predominantly negative evaluations of the resubmissions may reflect some form of response bias in favor of the original authors as a function of their association with prestigious institutions. These individuals may have received a less critical, more benign evaluation than did our unknown authors from 'no-name' institutions… The near perfect reviewer agreement regarding the unacceptability of the resubmitted manuscripts, coupled with the presumably near perfect agreement among the original reviewers in favor of publishing, provide additional convergent support for the response bias hypothesis" (p. 192).

Second, a seminal experiment by Blank (1991) demonstrated status bias in reviewing more directly. In her experiment, which lasted for 2 years, every other paper submitted to the *American Economic Review* was designated for double-blind review. Her question was "whether the ratio of acceptance rates between institutional ranks in the blind sample differs from the corresponding ratio in the nonblind sample" (pp. 1053–1054). It was found that this ratio did not differ for those at top-ranked departments nor for those at colleges and low-ranked universities. However, all other groups, in that important gray area where editorial judgment is most needed, had substantially lower acceptance rates in the blind sample than in the non-blind sample; in some cases, the acceptance rate dropped by more than seven percentage points.

While several retrospective studies have suggested the possibility of status bias, evidence from controlled experiments is inconclusive. Double-blind review may moderately reduce bias experienced by authors from less-prestigious institutions—and may be especially important for authors in the relatively undifferentiated middle.

Fairness to Non-US Authors

Evidence suggests that reviewers may exhibit bias when they know a manuscript's authors come from outside the United States.

A retrospective analysis of single-blind reviews found evidence of status bias at a coarse geographical level (Link, 1998). In this analysis of original research articles submitted to *Gastroenterology* over a 2-year period, it was found that "reviewers from the United States and outside the United States evaluate non-US papers similarly and evaluate papers submitted by US authors more favorably, with US reviewers having a significant preference for US papers" (Link, 1998, p. 246). While not from journals, a relevant retrospective analysis of all abstracts submitted to the American Heart Association's annual Scientific Sessions research meeting before and after implementation of double-blinded peer review found that blinding "significantly attenuated the association between country and likelihood of abstract acceptance" (Ross et al., 2006, pp. 1677–1678). Finally, a randomized study of 40 consecutive manuscripts submitted to *Dermatologic Surgery* found a nominal (but not statistically significant) difference in likelihood of acceptance associated with the geographic origin of the manuscript (Alam et al., 2011).

Although limited, the evidence suggests that reviewers have some preference for US authors. If so, double-blind review may modestly reduce bias experienced by authors from outside the United States.

Fairness to Prolific Authors

The experimental evidence is mixed concerning bias experienced by prolific authors.

The Mahoney et al. (1978) experiment, discussed above, suggests that reviewers may be biased in favor of prolific authors. The volunteer reviewers, who were not blinded, rated papers "as more innovative and publishable if the fictitious author included self-references in the manuscript than if no self-references were included" (Mahoney et al., 1978, p. 78).

By contrast, a randomized controlled trial involving manuscripts submitted to the *Journal of Developmental and Behavioral Pediatrics* found that "senior authors with more previous articles received significantly better scores from the blinded reviewers…, but not from the nonblinded reviewers" (Fisher et al., 1994, p. 145). They "interpret[ed] this finding to indicate that the blinded reviewers, especially those who were really blinded and could not guess author identity, may have recognized improved quality in the work of those authors with more previous publications. In contrast, reviewers who were aware of author identity did not give better scores to the more experienced authors, likely indicating that various types of bias may have entered into their thinking" (Fisher et al., 1994, p. 146).

A retrospective study found no impact of double-blind review on the rate at which prolific authors have their papers published in the Special Interest Group on Management of Data (SIGMOD) conference (Madden and DeWitt, 2006). However, an independent analysis of this data reached the opposite conclusion (Tung, 2006).

These contradictory results render it impossible to say anything definitive about the impact of blinding on prolific authors, though the strongest of the studies suggests that blinding could be beneficial.

Fairness to Female Authors

Sexism is a concern in peer review broadly (Wennerås and Wold, 1997). It is, therefore, unsurprising that several disciplines have launched in-depth studies in response to concerns related to gender equity in publication. These studies suggest the existence of bias against female authors.

A classic and much-referenced study showed that even when the work of a woman was identical to that of a man, the former was judged to be inferior to the latter (Goldberg, 1968). In this study, scholarly essays in a number of academic fields were presented to female college students. All of the students rated the same essays, but half of the essays carried the names of male authors (e.g., John T. McKay), whereas the other half carried the names of female authors (e.g., Joan T. McKay). Those essays attributed to male authors were rated higher. A quantitative *meta*-analysis of this and similar studies conducted over the two subsequent decades found that "the average difference between ratings of men and women is negligible" (Swim et al., 1989, p. 409). Consistent with this analysis, 73% of studies found no significant effect for the Joan–John manipulation, 20% found that John's work was rated higher, and the remaining 7% found that Joan's work was rated higher (Swim et al., 1989). More relevant to the issue of peer review is the observation from the American Psychological Association (APA) Task Force on Women in Academe that when "Joan and John's work was high in quality, the effect size was close to zero (–0.02 [the negative sign indicating a lower evaluation of female-authored work]); the effect was larger when Joan and John's work was medium in quality (–0.24)… [T]hese results seem to indicate that evaluation of absolutely outstanding articles will not be biased, but articles of ambiguous merit may be judged based on the author's gender" (Fouad et al., 2000, p. 45).

The seminal experiment by Blank (1991), described earlier, was initiated due to concerns of gender bias. At the time it was conducted, the *American Economic Review* had employed single-blind reviewing for most of its recent history. In the mid-1980s, the American Economic Association's Committee on the Status of Women in the Economics Profession formally expressed its concern about "the potential negative effect on women's acceptance rates of a single-blind system" (Blank, 1991, p. 1045). The Board of Editors at the *American Economic Review* subsequently asked Blank to design and run a randomized experiment looking into this. Her careful randomization process guaranteed that papers by women (and men) in each sample had identical distributions of characteristics. This made it possible to compare acceptance rates between the blind and non-blind samples without other controls, and indeed, there were striking differences. Blank found no significant difference in acceptance rates between the two samples for women. By contrast, acceptance rates for men were significantly higher in the non-blind sample. When reviewers knew that a male authored a paper, they accepted a higher percentage (15% vs 11%) than if the paper was blinded.

Blank (1991) emphasized that the core issue was not whether women's acceptance rates differed from men's but "whether the *ratio* of male to female acceptance rates in the non-blind sample is different from that in the blind sample" (p. 1053). She went on: "In both samples, women's acceptance rates are lower than men's, but the differential in the blind sample is smaller. While women in the blind sample have an acceptance rate only 1 percentage point below that of men, their rate is 3.8 percentage points lower in the non-blind sample" (p. 1053). The results were statistically insignificant, perhaps because there were too few observations

of papers authored by women. Blank concluded that "this paper provides little evidence that moving to a double-blind reviewing system will substantially increase the acceptance rate for papers by female economists" (p. 1063).

The Modern Language Association's (MLA) experience was strikingly different. The switch to double-blind reviewing resulted in a large increase in acceptances for female-authored publications, such that the acceptance rate for women eventually became comparable to that for men (Billard, 1993). The impetus for the MLA's change to double-blind review was the perception of gender bias by female MLA members. Although the MLA vigorously denied bias in its review processes, it instituted a blind reviewing procedure. "The result was unequivocal: There was a dramatic rise in the acceptance of papers by female authors" (Horrobin, 1982, p. 217).

More recently, a retrospective study of reviews for *Behavioral Ecology* found that in the 4 years following the introduction of double-blind review, the absolute number of manuscripts accepted for publication increased, which benefited both male first-authors and female first-authors (Budden et al., 2008). The magnitude of this difference was, however, significantly larger for females than for males: there was a 7.9% increase in the proportion of papers with a female first-author and a corresponding decrease in papers with a male first-author. Budden et al. found that the increase in female authorship at *Behavioral Ecology* was greater than that seen in most of the comparable single-blind journals over the same time period. Yet, an independent analysis of this data, using a general linear mixed model rather than testing each journal separately, reached the opposite conclusion, finding that "the increase in female authorship over time in *BE* [*Behavioral Ecology*] is not exceptionally different from the changes in the other journals in the field" (Webb et al., 2008, p. 352).

The evidence suggests that double-blinded peer review could have a favorable effect on acceptance rates for female authors. The consequences of such a shift could be far-reaching and extend well beyond publications: "If females are less successful in publishing research on account of their gender, then given current practices associated with appointment and tenure, and the need for women dramatically to out-compete their male counterparts to be perceived as equal, any such publication bias impedes the progress of women to more advanced professional stages" (Budden et al., 2008, p. 6).

The Perception of Fairness

Although the evidence we have just reviewed about the existence of bias in peer review is mixed, the fact is that "[r]efereeing 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" (Genest, 1993, p. 324). For some, "[w]hether 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" (Genest, 1993, p. 324). It seems inescapable to conclude that "much of the value of double-blind refereeing lies in the community perception of fairness" (Altman et al., 1992, p. 266).

Resnik et al. (2008) conducted a survey of government researchers concerning their perceptions about ethical problems with journal peer review. Bias was the second most commonly reported problem, with 50.5% of respondents saying they had encountered at least one biased

reviewer during the peer review process. A limitation of this opinion survey is that it does not provide objective data on whether bias actually existed. The authors note that answering yes to the question "a reviewer was biased" may, for example, "result from the respondent's own displeasure with the reviewer's comments or recommendations" (Resnik et al., 2008, p. 308). Nevertheless, documenting perceived bias "can be an important finding in its own right, because a scientist may change his/her behavior in response to what he/she perceives to be a problem… A researcher who is concerned that a reviewer is incompetent or biased may choose to ignore the reviewer's comments rather than address the concerns (which may in fact be valid)" (Resnik et al., 2008, pp. 308–309). If Resnik et al. are correct, even if bias does not actually exist, the mere perception of bias could minimize the effectiveness of peer review (i.e., improving the quality of published articles by incorporating feedback from reviewers) by eroding trust.

Concern about effects of perceived bias or unfairness is evident in the recent decision by the journal *Nature* to begin offering optional double-blind reviewing. The executive editor of Nature Publishing Group "admits that it will be difficult to demonstrate conclusively that double-blind review eliminates bias. 'But we feel that is not a reason not to offer it. If scientists feel they would be better served this way, why would we not offer it?'" (Jump, 2015, n.p.).

Trust in the peer review process is both intrinsically and instrumentally valuable. Trust may be promoted by fairness in the procedures (i.e., how people are treated) and/or fairness in the outcome (i.e., publication decisions). If double-blind reviewing promotes trust in the process, regardless of how it affects outcomes, this may be the strongest argument yet in favor of double over single-blind peer review.

BLINDING AS A MEANS OF IMPROVING THE QUALITY OF REVIEWS

As described above, peer review is intended to improve published manuscripts by aiding editors in identification of high-impact articles for publication and by offering constructive feedback for authors to incorporate in the form of revisions, whether minor or substantive. High-quality reviews should further both of these ends.

The quality of reviews may be evaluated along two dimensions: content, which "refers to the substance and scientific scrutiny evident in the review," and format, which "addresses practicalities of the peer review and editorial processes" such as timeliness and observation of good reviewing etiquette (Feurer et al., 1994, p. 98). According to a survey conducted by Kearney and Freda (2005), "[T]he most universally desired characteristic [of reviews] was specific, detailed guidance to the author on how to improve the manuscript. Editors preferred reviews that demonstrated that the paper had 'obviously been read carefully…' …A balanced, objective, and fair review that included comments on strengths and weaknesses was also valued, as was a review that conveyed suggestions in a helpful, positive, or constructive tone" (p. 449).

Effects of (Not) Revealing Author Identity

Contradictory theories have been advanced regarding both the mechanism and direction of the effect of double-blinding on review quality.

On the one hand, advocates of open or single-blind peer review offer a variety of arguments in favor of their position. One is that revealing an author's identity increases the quality of reviews by supplying peer reviewers with relevant information about the author's prior accomplishments, publications, and citation rates (Dalton, 1995). Another belief is "that information about the authors' institutional affiliation helps referees evaluate manuscripts because they constitute presumptive 'proof' that the research described was actually done" (Campanario, 1998b, p. 295). For instance, it might "reduce the number of questions referees need to ask about whether a lab has the expertise to perform a 'tricky' technique successfully" (Jump, 2015, n.p.). On the other hand, advocates of double-blind review argue that reviewers might use identifying information as an inappropriate shortcut or faulty heuristic in their decision-making, whereas not revealing an author's identity permits the work to speak for itself (Billard, 1993). Advocates of blinding reviewers claim that the excellence of the manuscript should be readily apparent, even in the absence of identifying information (Perlman, 1982).

The evidence is mixed. Before proceeding, we note that most of the studies used their own rating instruments, and most scales appeared not to be validated. We will begin by looking at the evidence suggesting that double-blinding does, in fact, enhance review quality.

A double-blind study of double-blind reviewing (!), carried out at the *Journal of General Internal Medicine*, found that "blinding reviewers improves the quality of review from the editor's perspective" (McNutt et al., 1990, p. 1375). A potential limitation to the study is that "the referees receiving the blinded copy of the manuscript would have been aware that they were part of an experiment and may in consequence have been more careful with their reports" (Cox et al., 1993, p. 316).

Jadad et al. (1996) sought to determine whether blinding would improve the quality of recommendations to include clinical trials in meta-analyses and systematic reviews. Reviewers were randomly assigned to open or blind assessment of the clinical trials. "Blind assessments of the reports produced significantly lower and more consistent scores" (Jadad et al., 1996, p. 9). Jadad et al. characterized this as an advantage of blinding, stating, "This may be very important for editors of journals to reduce bias in manuscript selection which could be introduced in open peer review" (p. 10).

Despite evidence that "scientists who have done large amounts of good quality work in the recent past are likely to continue doing so in the near future" (Abrams, 1991, p. 115), Laband and Piette (1994) looked at citation rates and found that "articles reviewed single-blind are less likely than those reviewed double-blind to be identified correctly as the highest-impact articles" (p. 148).

Now, we will turn our attention to studies that failed to find that blinding reviewers to author identity improves the quality of peer review. One of these was a retrospective study (Isenberg et al., 2009); others, however, were randomized experiments.

In one randomized experiment, reviewers for five biomedical journals either conducted their manuscript review according to the journal's usual practice, which for four of the journals was single-blind, or one of the two reviewers was randomly selected to conduct double-blind peer review (Justice et al., 1998). Justice et al. found that "authors and editors perceived no significant difference in quality between masked and unmasked reviews [and] no difference in the degree to which the review influenced the editorial decision" (p. 240).

In another randomized experiment, a paper that had already been accepted for publication by *BMJ* was altered to introduce eight weaknesses in design, analysis, or interpretation (Godlee et al., 1998). Study participants were then asked to comment on the paper as part of a study into ways of improving peer review. Blinding "had no effect on the quality of reviewers' reports as judged by the number of weaknesses that they identified in the manuscript" (Godlee et al., 1998, p. 239).

In a randomized trial of 527 consecutive manuscripts submitted to *BMJ*, reviewers were randomized to receive either a blinded or unblinded version of a manuscript (van Rooyen et al., 1998). The study concluded that blinding has little effect on the quality of reviews. There was no significant difference in the recommendations regarding publication between the blinded and unblinded groups.

Finally, the Alam et al. (2011) study, described above, found no significant difference between the scores given to manuscripts by blinded and unblinded reviewers. There was also no difference in word count, which was used as a proxy for review quality, between blinded and unblinded reviews. These "results would suggest that blinded reviewers are no more or less likely to provide quality reviews than unblinded reviewers" (Alam et al., 2011, p. 566).

Although some studies suggest that author anonymity enhances the quality of reviews, other studies, generally with superior design, do not document such an effect. This suggests a negligible effect of blinding on review quality.

Effects of (Not) Revealing Reviewer Identity

Several studies have examined the effects of disclosing *reviewer* identity—to authors, to other reviewers, and even to a journal's readership.

In the randomized trial conducted by van Rooyen et al. (1998), described above, manuscripts were additionally randomized as to whether reviewers' identifies were revealed to their co-reviewers or not, which was (somewhat confusingly) called "unmasking." van Rooyen et al. explained that "[a]lthough 2-factor analysis of variance revealed a statistically significant difference [in review quality] between the masked and unmasked groups ($P = 0.04$), absolute differences were of no editorial significance" (p. 236). Similarly, Das Sinha et al. (1999) found that informing reviewers that their reviews would be exchanged with co-reviewers did not affect quality.

In another randomized trial, 125 manuscripts submitted to *BMJ* in early 1998 were each sent to two reviewers who were randomized to remain anonymous or to have their identity revealed to the manuscript's author(s) (van Rooyen et al., 1999). No significant difference in quality between anonymous reviewers and identified reviewers was found. In addition, no significant difference was found in either the recommendation regarding publication or time taken to review the manuscript. van Rooyen et al. (1999) did find, however, that asking reviewers to be identified significantly increased the likelihood of reviewers declining to review. Interestingly, *BMJ* announced that it would let authors know the identity of reviewers in 1999 (Smith, 1999).

Walsh et al. (2000) found notably different results. They conducted a trial in which reviewers for the *British Journal of Psychiatry* were randomized to either a signed or an unsigned group. Upon comparing the quality of reviews, Walsh et al. found that "the total mean score was significantly higher in the signed group than the unsigned group" (p. 48). They noted that

"signed reviews were significantly more courteous and less abusive than unsigned reviews" (p. 48) and offered more constructive comments. Signed reviewers were significantly more likely than unsigned reviewers to recommend rejection of manuscripts.

In an interesting twist, a study conducted at *BMJ* randomized manuscripts to either have the reviewer's signed report made available only to the author or have it made available to the public on the *BMJ* Web site alongside the published paper (van Rooyen et al., 2010). The intervention was revealing to reviewers, after they agreed to review but before they undertook their review, that their signed report might appear on the Web site. "Forewarning reviewers that their signed reviews might be published had no discernable effect on the overall quality of their reviews" (van Rooyen et al., 2010, n.p.).

At this point, the jury is still out. There is little to suggest that reviewer anonymity enhances the quality of peer review, though there may be practical reasons for favoring anonymity, such as retaining reviewers.

BREAKING THE (DOUBLE) BLIND

Many benefits ascribed to double-blinding assume that blinding is successful in shielding the identity of the author from the reviewer. Of course, "[h]ow truly anonymous any party can be in a world in which referees are selected for their in-depth knowledge of a small slice of the universe of knowledge is open to question" (Dalton, 1995, p. 236). "The notion that a…referee can identify the author of a given paper in a specialty journal has been used by many to derogate the claim of an advantage to double-blind review" (Campanario, 1998b, p. 295).

How Double-Blinding Is Accomplished

Double-blinding may be accomplished in different ways. As a first cut, it is useful to differentiate author blinding from editorial blinding. *Author blinding* occurs when the author removes identifying information from a manuscript before submitting it, whereas *editorial blinding* occurs when identifying information is removed in the editorial process before the manuscript is sent to reviewers. Authors may be in a better position than editors to blind because they are more likely to know what kind of information could be identifying, though they may have incentives to evade the blind.

Another important variable is how much effort is expended on the task. "To simply block out the name and affiliation from the title page requires minimal effort, to block out self-references adds a little more, and scrutinizing the manuscript for any internal cues necessitates laborious line-by-line study. Therefore, the efficacy of the blinding process will vary directly with the effort expended on it" (Pitkin, 1995, p. 781). The feasibility of blinding is also "related to the type and circulation of the involved journal" (Fisher et al., 1994, p. 145).

Blinding Success Rates

Several studies have examined how well blinding works in practice, that is, how well a double-blinded manuscript hides the identity of the author(s). These studies, across a wide range of disciplines, have found that blinding achieved success rates of 53–79%.

A study of individuals who reviewed manuscripts submitted to the *Journal of Social Service Research* during 1978 showed that "in 56% of the reviews, the referees did not venture a guess as to the identity of the author. In another 4%, the referees guessed wrong. In an additional 5%, the referees made correct guesses about some bit of identifying information, but they did not guess the name of the author" (Rosenblatt and Kirk, 1981, p. 389). This works out to successful blinding 65% of the time for editorial blinding.

One retrospective study over six psychology journals employing author blinding showed that "35.6% of the 146 reviewers were correct in their identifications of the author or of at least one of the authors in the case of multi-authored papers… There were no significant differences in the proportion of correct detections among the six journals,… nor was there any relationship between detection accuracy and the number of years of reviewing experience" (Ceci and Peters, 1984, p. 1493). When "editorial staff oversights in not removing title pages of manuscripts before sending them to reviewers or authors' oversights in preparing their manuscripts, such as explicit flagging of former work ('In our earlier work…') or inappropriate inclusion of personal acknowledgments in the body of the text,…are excluded from the analysis, overall only 25.7% of reviewers are able to detect authors' identities, with very little variation among the six journals" (Ceci and Peters, 1984, p. 1494).

In another study, reviewers for the *American Journal of Public Health*, which relies on a combination of author and editorial blinding, were asked to identify author and institution for the manuscript they reviewed (Yankauer, 1991). Reviewers claimed to be able to identify the author and/or institution in 47% of cases. Of those cases, correct identification of the author or institution occurred 84% of the time. Blinding could, therefore, be considered successful 53% or 61% of the time, depending on whether successful blinding is defined to encompass self-reported instances of identification or is defined more narrowly to include only correct instances of identification.

Another experiment found that "a long-standing policy of masking did not increase masking success" (Cho et al., 1998, p. 243). This study included three medical journals with a long-standing policy of double-blind review and four medical journals that did not mask author identity. During the experiment, manuscripts that met the study's inclusion criteria were editorially blinded "by removing author and institutional identity from the title page, running headers or footers, and acknowledgments of manuscripts. Self-references in the text were not removed" (Cho et al., 1998, p. 244). Reviewers were asked if they thought they could identify any of the principal authors of the manuscript, and if yes, to name the author(s). If any reviewer correctly identified one or more authors, the reviewer was considered unmasked. Overall, 60% of reviewers were successfully blinded.

A prospective review of manuscripts submitted to two radiology journals with double-blind peer review policies found that, despite explicit instructions to authors, 34% of the manuscripts "contained information that potentially or definitely unblinded the identities of the authors or their institutions" (Katz et al., 2002, p. 1417). The most frequent violations of blinded peer review policies were "the statement of the authors' initials in the body of the manuscript, referencing the authors' own work in press, identifying references as the authors' previous work, revealing the identity of the institution in the figures, and stating the identity of the authors' institutions in the manuscript" (Katz et al., 2002, p. 1416). Editors at the two journals correctly identified the authors, their institutions, or both in 74% of the manuscripts

in which the authors had been unsuccessful in their blinding. Unblinding occurred in 25% of all of the manuscripts reviewed.

A survey of reviewers for nursing journals found that blinding was generally successful at masking authors' identities (Baggs et al., 2008). Respondents were asked what percentage of the time they could identify authors even if they did not receive the authors' names. Sixty-two percent reported that they could never identify authors of papers, and another 17.5% reported that they could identify authors less than 10% of the time.

Several of the studies discussed previously also looked at the efficacy of blinding. In the McNutt et al. (1990) study, editorial blinding successfully masked the institution name for 73% of the reviewers and the authors' names for 76% of the reviewers. Blank's (1991) experiment found that "among all referee surveys received for blind papers, slightly over half (50.9 percent) claim to know the author. Ten percent of these referees are incorrect, however, so that only 45.6 percent of the authors in the blind sample are correctly identified" (p. 1051). The study by Fisher et al. (1994) evaluated the success of an editorial blinding process and found 54% of reviewers were thereby successfully blinded. Justice et al. (1998) concluded that "[s]uccess in masking reviewers to author identity was generally low (68%)" (p. 241). Manuscripts by authors with whom the reviewer was familiar were less likely to be successfully masked than manuscripts by authors who were not known to the reviewer (Justice et al., 1998). In the study by Godlee et al. (1998), 26% of the 90 respondents who had been blinded to the authors' identities nevertheless correctly named the authors. van Rooyen et al. (1998) found that 58% of reviewers were successfully blinded. Finally, Isenberg et al. (2009) found that 77.7% of the reviewers reported that they had no idea about the author's identity.

Overall, these studies suggest that even minimal editorial blinding can be somewhat effective. How one characterizes success rates may depend on whether one's philosophical viewpoint is more "glass-half-full" or "glass-half-empty."

Impact of Technology on Blinding

Going forward, technology may alter the feasibility and success of double-blinding. A handful of studies have explored this possibility. A data-mining experiment by Hill and Provost (2003) found that a computer program that tallied self-citations could identify the authors of physics papers using only the citations included in the paper 45% of the time. Authors with 100 or more prior publications could be identified 85% of the time.

In some disciplines, authors have strong incentives to disseminate their research results, for example, as a working paper, before eventual publication in an academic journal. These incentives may make blinding of manuscripts more difficult. Holm (2011) took a sample of 32 manuscripts from two economics journals that use double-blind review and asked participants in his study to use information about each paper (e.g., title and abstract) and the Internet search engine of their choice to identify the author. It took participants between 24 and 33 s to identify working papers and 34–62 s to identify conference papers using Google or Google Scholar. Holm concluded that "information about nearly all accepted papers was available on the Internet before their respective review processes were expected to be finished" (p. 26). He claims "author anonymity in a double-blind review process is no longer credible" (p. 26).

Neither of these studies spoke to reviewers' incentives to deliberately break a blind. They strongly suggest, however, that should an author wish to do so, the cost and difficulty of breaking a blind are increasingly low.

PREFERENCES FOR OPEN, SINGLE-, OR DOUBLE-BLIND REVIEW

Given the equivocal nature of much of the evidence on blinding, it is reasonable to ask how stakeholders in the peer review process feel about blinding. Studies conducted across a variety of disciplines suggest that there is a preference (though not unanimous) for double-blind review by reviewers, authors, and editors. Reasons for their preferences are relatively consistent across studies and reflect the debates in the literature. We now review these studies chronologically.

In the survey conducted by Yankauer (1991), described above, respondents were asked whether they favored double-blind review. In all, 75% of respondents favored blinding, and 8% stated that they had no preference. "The principal reason given for their opinion by those favoring blind review was the elimination of bias. The principal reason given for their opinion by those opposing blind review was failure of the attempt to blind" (Yankauer, 1991, p. 844). Yankauer noted that additional reasons given for opposing blind review were contradictory: respondents stated both that identification will not influence judgment and that identification assists judgment. Interestingly, on follow-up, only eight of the 30 reviewers who initially opposed blind review stated that they would sign all their reviews; 19 indicated they would not do so, and three would do so only in some instances.

A survey of Institute of Mathematical Statistics (IMS) members indicated support for double-blind review in the IMS journals (Cox et al., 1993). Nearly half of respondents (46%) agreed or strongly agreed with the statement that the IMS should institute double-blind review, while 24% were neutral, and 29% responded negatively. There was, apparently, a split between new researchers, a majority of whom endorsed double-blind refereeing (Altman et al., 1991, 1992), and more senior members, whose views were characterized as "negative but sympathetic" (Cox et al., 1993, p. 311). "An informal poll of some past editors of the *Annals* indicated little support for double-blind refereeing" (Cox et al., 1993, p. 311).

In the survey conducted by Kearney and Freda (2005), discussed above, 80% of nursing journal editors said that blinding was important in peer review, and 53% said they could see no benefit to unblinding either the author or reviewer. The editors expressed belief that double-blinding promoted objectivity "by protecting authors and reviewers from personal and emotional responses that might ensue from being identifiable or identifying the other party" (Kearney and Freda, 2005, p. 448).

When reviewers for *Medical Education* were surveyed, "74% of respondents said they would be willing to sign their reviews in the spirit of creating a more transparent review process; 20% were not sure and 6% would be unwilling" (Snell and Spencer, 2005, p. 95). First-time reviewers were more likely than more experienced reviewers to say 'No' or 'Not Sure.' Proponents of signing reviews "offered comments such as: 'peer review should be more open transparent and honest'; 'it is more fair and constructive'; 'reviewers will take more time with reviews'; and 'the quality of reviews would be better if the reviewer had to reveal their identity: if your opinions are based on sound judgment you should be prepared to stand by them'" (Snell and Spencer, 2005, p. 95). Meanwhile, comments from opponents included:

"'it's important to be able to make frank disclosures to the editors'; 'I might be tempted to be too soft'; 'this might alter my responses'; and 'there's the inevitable embarrassment of saying no to friends'" (Snell and Spencer, 2005, p. 95).

Subsequent to the survey conducted by Snell and Spencer (2005), *Medical Education* conducted a survey of both authors and reviewers who had submitted or reviewed a manuscript for the journal in 2003 and 2004 (Regehr and Bordage, 2006). Overall, 68% of respondents indicated a preference for some process of blinding that concealed author names. "Almost three-quarters of respondents (72%) indicated a desire for a procedure in which reviewer names were concealed or only revealed at the discretion of the reviewer (double-blind, single-blind, or optional single-blind)" (Regehr and Bordage, 2006, p. 837). Regehr and Bordage (2006) found that the primary reasons cited in favor of concealing author identity were: "to facilitate fairness (to obtain less biased and more objective reviews); to allow the manuscript to stand on its own merit; and to meet the need to avoid personal conflicts or rivalries with reviewers, especially given the small size of the medical education research community and the need to further protect junior members" (p. 835). Similar reasons were given for concealing reviewer identity: "to facilitate more honest reviews; to avoid 'bad blood' and personal tension among colleagues and friends, including retaliations (e.g., interfering with career development or grant applications); to allow the review to stand on its own merit; to encourage trust in the editor's choice of reviews, and to allow authors not to have to deal with knowing the identity of a reviewer bearing bad news or returning a harsh critique of one's work" (Regehr and Bordage, 2006, p. 835). The primary reasons given for revealing author identity were: "to allow for the preparation of better reviews by facilitating more understanding of the context of the study; to meet the need for greater transparency, and to acknowledge that masking does not always work because of specific clues in the manuscript. The main reasons for revealing reviewer identity were: to allow the author to appreciate the reviewer's qualifications and credibility; to foster more constructive and concrete reviews; to meet the need for accountability (by allowing the reviewer to stand by his or her critique); to facilitate fairness, and to allow for the possibility of author contact with the reviewer" (Regehr and Bordage, 2006, pp. 835–836).

The majority (73%) of respondents to a survey of ecologists and evolutionary biologists expressed a preference for double-blinding; 21% for open peer view; and only 6% for single-blinding (the peer review method typically employed by ecological journals) (Smit, 2006). The "clear preference for the double-blinded system existed across all groups, regardless of age, gender, academic position, and number of publications" (Smit, 2006, p. 713).

In the Baggs et al. (2008) survey, discussed above, nursing journal reviewers expressed a preference to maintain double-blinding, the most common form of blinding encountered by members of the sample. Eighty-one percent of respondents disagreed or strongly disagreed with the statement: "Reviewers should be informed of the identities of authors of a manuscript" (Baggs et al., 2008, p. 134). "Respondents thought that blinding decreased the potential for interpersonal conflict, encouraged honest and fair assessments on the part of reviewers, and protected junior reviewers from senior authors" (Baggs et al., 2008, p. 137).

It is interesting that some of the same arguments—such as facilitating fairness—are made both in favor of and against unblinding, which may be attributable to the inconclusive nature of the evidence on blinding. Nevertheless, a preference for blinding apparently predominates. These preferences may vary by discipline; as mentioned before, disciplines vary widely in their use of open, single-, and double-blind review.

CONCLUSION

At the outset, we remarked that although peer review is the gold standard for assessment of manuscripts submitted to journals, discussions regarding how best to accomplish it are ongoing. Such discussions are academic only in the sense that they relate to scholarship; they have great practical relevance. Much is at stake in peer review. Genest (1993) observed:

> 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. (p. 324)

Picking up on the central issue of fairness, we began this chapter with an exploration of data related to double-blinding as a means of imparting fairness to the peer review process. In 1991, Blank wrote:

> In summary, the literature on single-blind versus double-blind reviewing spans a wide variety of disciplines and provides rather mixed results. Few of the empirical tabulations provide convincing evidence on the effects or non-effects of refereeing practices, largely because of their inability to control for other factors in the data. If not fully convincing, however, there is at least a disturbing amount of evidence in these studies that is consistent with the hypothesis of referee bias in single-blind reviewing. (p. 1045)

Twenty-five years later, her conclusion holds. Despite a lack of conclusive evidence that bias exists, we have argued that even the perception of bias in peer review is troubling because "[t]hreats to the impartiality of review appear to threaten peer review's psychological and epistemic legitimacy" (Lee et al., 2013, p. 3). That, in turn, is a threat to all of us who are producers and/or users of knowledge.

Is double-blind reviewing a good or even plausible answer to this threat? The evidence is inconclusive, and reasonable people disagree. Many stakeholders in the peer review process favor double-blinding precisely because it is perceived to be fairer than the alternatives. Others think that blinding runs counter to the transparency that should inhere in the pursuit of knowledge. It is far from clear that blinding actually imparts fairness or enhances the quality of reviews; moreover, blinding is not always successful. Complicating matters, "the issue of double-blind refereeing today is one fraught with emotional overtones both rational and irrational, often subconsciously culturally based, and so is difficult for many of us to resolve equitably no matter how well intentioned" (Billard, 1993, p. 320).

The policy conundrum before each journal then, is this: how do we value fairness, and how does that value weigh against the costs and limits of blinding? We do not think that there is one "right" answer. Rather, each publication—and the community that it serves (i.e., producers and users of knowledge)—must carefully deliberate, in light of the existing literature and its own values and interests, and determine what is right for it.

References

Abrams, P.A., 1991. The predictive ability of peer review of grant proposals: the case of ecology and the U.S. National Science Foundation. Social Studies of Science 21 (1), 111–132.

Alam, M., Kim, N.A., Havey, J., Rademaker, A., Ratner, D., Tregre, B., West, D.P., Coleman, W.P., 2011. Blinded vs. unblinded peer review of manuscripts submitted to a dermatology journal: a randomized multi-rater study. British Journal of Dermatology 165 (3), 563–567.

Altman, N., Banks, D., Chen, P., Duffy, D., Hardwick, J., Leger, C., Owen, A., Stukel, T., 1991. Meeting the needs of new statistical researchers. Statistical Science 6 (2), 163–174.

Altman, N., Angers, J.F., Banks, D., Duffy, D., Hardwick, J., Leger, C., Martin, M., Nolan, D., Owen, A., Politis, D., Roeder, K., Stukle, T.N., Ying, Z., 1992. Rejoinder Statistical Science 7 (2), 265–266.

Armstrong, J.S., 1997. Peer review for journals: evidence on quality control, fairness, and innovation. Science and Engineering Ethics 3 (1), 63–84.

Bachand, R.G., Sawallis, P.P., 2003. Accuracy in the identification of scholarly and peer-reviewed journals and the peer-review process across disciplines. The Serials Librarian 45 (2), 39–59.

Baggs, J.G., Broome, M.E., Dougherty, M.C., Freda, M.C., Kearney, M.H., 2008. Blinding in peer review: the preferences of reviewers for nursing journals. Journal of Advanced Nursing 64 (2), 131–138.

Benos, D.J., Bashari, E., Chaves, J.M., Gaggar, A., Kapoor, N., LaFrance, M., Mans, R., Mayhew, D., McGowan, S., Polter, A., Qadri, Y., Sarfare, S., Schultz, K., Splittgerber, R., Stephenson, J., Tower, C., Walton, R.G., Zotov, A., 2007. The ups and downs of peer review. Advances in Physiology Education 31 (2), 145–152.

Billard, L., 1993. Comment Statistical Science 8 (3), 320–322.

Blank, R.M., 1991. The effects of double-blind versus single-blind reviewing: experimental evidence from *the American Economic Review*. The American Economic Review 81 (5), 1041–1067.

Bordage, G., Caelleigh, A.S., 2001. A tool for reviewers: "Review criteria for research manuscripts". Academic Medicine 76 (9), 904–908.

Bradley, J.V., 1981. Pernicious publication practices. Bulletin of the Psychonomic Society 18 (1), 31–34.

Budden, A.E., Tregenza, T., Aarssen, L.W., Koricheva, J., Leimu, R., Lortie, C.J., 2008. Double-blind review favours increased representation of female authors. Trends in Ecology & Evolution 23 (1), 4–6.

Campanario, J.M., 1998a. Peer review for journals as it stands today—part 1. Science Communication 19 (3), 181–211.

Campanario, J.M., 1998b. Peer review for journals as it stands today—part 2. Science Communication 19 (4), 277–306.

Ceci, S.J., Peters, D.P., 1984. How blind is blind review? American Psychologist 39 (12), 1491–1494.

Cho, M.K., Justice, A.C., Winker, M.A., Berlin, J.A., Waeckerle, J.F., Callaham, M.L., Rennie, D., 1998. Masking author identity in peer review: what factors influence masking success? PEER Investigators. Journal of the American Medical Association 280 (3), 243–245.

Cox, D., Gleser, L., Perlman, M., Reid, N., Roeder, K., 1993. Report of the ad hoc committee on double-blind refereeing. Statistical Science 8 (3), 310–317.

Dalton, M.S., 1995. Refereeing of scholarly works for primary publishing. Annual Review of Information Science and Technology (ARIST) 30, 213–250.

Feurer, I.D., Becker, G.J., Picus, D., Ramirez, E., Darcy, M.D., Hicks, M.E., 1994. Evaluating peer reviews. Pilot testing of a grading instrument. Journal of the American Medical Association 272 (2), 98–100.

Fisher, M., Friedman, S.B., Strauss, B., 1994. The effects of blinding on acceptance of research papers by peer review. Journal of the American Medical Association 272 (2), 143–146.

Fletcher, R.H., Fletcher, S.W., 1997. Evidence for the effectiveness of peer review. Science and Engineering Ethics 3 (1), 35–50.

Fouad, N., Brehm, S., Hall, C.I., Kite, M.E., Hyde, J.S., Russo, N.F., 2000. Women in Academe: Two Steps Forward, One Step Back. Retrieved from: http://www.apa.org/pi/women/programs/academe/taskforce-report.pdf.

Garfunkel, J.M., Ulshen, M.H., Hamrick, H.J., Lawson, E.E., 1994. Effect of institutional prestige on reviewers' recommendations and editorial decisions. Journal of the American Medical Association 272 (2), 137–138.

Genest, C., 1993. Comment Statistical Science 8 (3), 323–327.

Godlee, F., Gale, C.R., Martyn, C.N., 1998. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. Journal of the American Medical Association 280 (3), 237–240.

Goldberg, P., 1968. Are women prejudiced against women? Society 5 (5), 28–30.

Gordon, M.D., 1980. The role of referees in scientific communication. In: Hartley, J. (Ed.), The Psychology of Written Communication: Selected Readings. Kogan Page, London, pp. 263–275.
Hill, S., Provost, F., 2003. The myth of the double-blind review? Author identification using only citations. ACM SIGKDD Explorations Newsletter 5 (2), 179–184.
Hojat, M., Gonnella, J.S., Caelleigh, A.S., 2003. Impartial judgment by the "gatekeepers" of science: fallibility and accountability in the peer review process. Advances in Health Sciences Education 8 (1), 75–96.
Holm, H.J., 2011. Double-blind in light of the internet: a note on author anonymity. Information Economics and Policy 23 (1), 24–26.
Horrobin, D.F., 1982. Peer review: a philosophically faulty concept which is proving disastrous for science. Behavioral and Brain Sciences 5 (2), 217–218.
Isenberg, S.J., Sanchez, E., Zafran, K.C., 2009. The effect of masking manuscripts for the peer-review process of an ophthalmic journal. British Journal of Ophthalmology 93 (7), 881–884.
Jadad, A.R., Moore, A., Carroll, D., Jenkinson, C., Reynolds, D.J.M., Gavaghan, D.J., McQuay, H.J., 1996. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Controlled Clinical Trials 17, 1–12.
Jefferson, T., Rudin, M., Brodney Folse, S., Davidoff, F., 2007. Editorial peer review for improving the quality of reports of biomedical studies. The Cochrane Database of Systematic Reviews 18 (2).
Jump, P., 2015. Will Double-Blind Reviewing Bring Quality into Focus? Times Higher Education. Retrieved from: https://www.timeshighereducation.co.uk/news/will-double-blind-reviewing-bring-quality-into-focus/2018963.article.
Justice, A.C., Cho, M.K., Winker, M.A., Berlin, J.A., Rennie, D., 1998. Does masking author identity improve peer review quality? A randomized controlled trial. PEER Investigators. Journal of the American Medical Association 280 (3), 240–242.
Katz, D.S., Proto, A.V., Olmsted, W.W., 2002. Incidence and nature of unblinding by authors: our experience at two radiology journals with double-blinded peer review policies. American Journal of Radiology 179 (6), 1415–1417.
Kearney, M.H., Freda, M.C., 2005. Nurse editors' views on the peer review process. Research in Nursing & Health 28 (6), 444–452.
Laband, D.N., Piette, M.J., 1994. A citation analysis of the impact of blinded peer review. Journal of the American Medical Association 272 (2), 147–149.
Lee, C.J., Sugimoto, C.R., Zhang, G., Cronin, B., 2013. Bias in peer review. Journal of the American Society for Information Science and Technology 64 (1), 2–17.
Link, A.M., 1998. U.S. and non-U.S. submissions: an analysis of reviewer bias. Journal of the American Medical Association 280 (3), 246–247.
Madden, S., DeWitt, D., 2006. Impact of double-blind reviewing on SIGMOD publication rates. ACM SIGMOD Record 35 (2), 29–32.
Mahoney, M.J., Kazdin, A.E., Kenigsberg, M., 1978. Getting published. Cognitive Therapy and Research 2 (1), 69–70.
McNutt, R.A., Evans, A.T., Fletcher, R.H., Fletcher, S.W., 1990. The effects of blinding on the quality of peer review: a randomized trial. Journal of the American Medical Association 263 (10), 1371–1376.
Perlman, D., 1982. Reviewer "bias": do Peters and Ceci protest too much? Behavioral and Brain Sciences 5 (02), 231–232.
Peters, D.P., Ceci, S.J., 1982. Peer-review practices of psychological journals: the fate of published articles, submitted again. Behavioral and Brain Sciences 5 (2), 187–195.
Pitkin, R.M., 1995. Blinded manuscript review: an idea whose time has come? Obstetrics & Gynecology 85 (5), 781–782.
Regehr, G., Bordage, G., 2006. To blind or not to blind? what authors and reviewers prefer. Medical Education 40 (9), 832–839.
Resnik, D.B., Gutierrez-Ford, C., Peddada, S., 2008. Perceptions of ethical problems with scientific journal peer review: an exploratory study. Science and Engineering Ethics 14 (3), 305–310.
van Rooyen, S., Godlee, F., Evans, S., Smith, R., Black, N., 1998. Effect of blinding and unmasking on the quality of peer review. Journal of the American Medical Association 280 (3), 234–237.
van Rooyen, S., Godlee, F., Evans, S., Black, N., Smith, R., 1999. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. BMJ 318, 23–27.
van Rooyen, S., Delamothe, T., Evans, S.J., 2010. Effect on peer review of telling reviewers that their signed reviews might be posted on the web: randomised controlled trial. BMJ 341, c5729.

Rosenblatt, A., Kirk, S.A., 1981. Recognition of authors in blind review of manuscripts. Journal of Social Service Research 3 (4), 383–394.
Rosenthal, R., 1982. Reliability and bias in peer-review practices. Behavioral and Brain Sciences 5 (2), 235–236.
Ross, J.S., Gross, C.P., Desai, M.M., Hong, Y., Grant, A.O., Daniels, S.R., Hachinski, V.C., Gibbons, R.J., Gardner, T.J., Krumholz, H.M., 2006. Effect of blinded peer review on abstract acceptance. Journal of the American Medical Association 295 (14), 1675–1680.
Sinha, S.D., Sahni, P., Nundy, S., 1999. Does exchanging comments of Indian and non-Indian reviewers improve the quality of manuscript reviews? National Medical Journal of India 12, 210–213.
Smit, C., 2006. Peer review: time for a change? BioScience 56 (9), 712–713.
Smith, R., 1999. Opening up BMJ peer review. BMJ 318 (7175), 4–5.
Smith, R., 2006. Peer review: a flawed process at the heart of science and journals. Journal of the Royal Society of Medicine 99 (4), 178–182.
Snell, L., Spencer, J., 2005. Reviewers' perceptions of the peer review process for a medical education journal. Medical Education 39 (1), 90–97.
Snodgrass, R.T., 2006. Single-versus double-blind reviewing: an analysis of the literature. ACM SIGMOD Record 35 (3), 8–21.
Snodgrass, R.T., 2007. Editorial: single- versus double-blind reviewing. ACM Transactions on Database Systems (TODS) 32 (1), 1–31.
Swim, J., Borgida, E., Maruyama, G., Myers, D.G., 1989. Joan McKay versus John McKay: do gender stereotypes bias evaluations? Psychological Bulletin 105 (3), 409–429.
Tremain, S., 2011. Abelist Language and Philosophical Associations. Retrieved from: http://www.newappsblog.com/2011/07/ableist-language-and-philosophical-associations.html.
Tung, A.K., 2006. Impact of double blind reviewing on SIGMOD publication: a more detail analysis. ACM SIGMOD Record 35 (3), 6–7.
Walsh, E., Rooney, M., Appleby, L., Wilkinson, G., 2000. Open peer review: a randomised controlled trial. The British Journal of Psychiatry 176 (1), 47–51.
Webb, T.J., O'Hara, B., Freckleton, R.P., 2008. Does double-blind review benefit female authors? Trends in Ecology & Evolution 23 (7), 351–353.
Weller, A.C., 2001. Editorial Peer Review: Its Strengths and Weaknesses. Information Today, Inc, Medford, NJ.
Wennerås, C., Wold, A., 1997. Nepotism and sexism in peer-review. Nature 387 (6631), 341–343.
Yankauer, A., 1991. How blind is blind review? American Journal of Public Health 81 (7), 843–845.
Zuckerman, H., Merton, R.K., 1971. Patterns of evaluation in science: institutionalisation, structure and functions of the referee system. Minerva 9 (1), 66–100.