Single- Versus Double-Blind Reviewing: An Analysis of the Literature

Richard Snodgrass
rts@cs.arizona.edu

1 Introduction

The peer review process is generally acknowledged as central to the advancement of scholarly knowledge. It is also vital to the advancement of individual careers.

With so much at stake, it is important to examine, and re-examine, issues pertaining to review quality on an ongoing basis. Thus it is appropriate that controversy has arisen in our field pertaining to the practice of double-blind reviewing. "As scientists, we should rather welcome all occasions to reflect on the act of writing, evaluating, editing and publishing research findings. The issue of double-blind refereeing, which recurs periodically in scientific circles, provides us with such an opportunity" [Genest 1993, page 324].

Most database journals employ single-blind reviewing, in which the reviewer is unknown to the author, but the identity of the author is known to the reviewer. Others employ double-blind reviewing, in which the identity of the author and the reviewer are not known to each other. The arguments for double-blind reviewing are that it is fairer and that it produces higher quality reviews. The arguments advanced against double-blind reviewing include that it has little effect, that it makes it more difficult for reviewers to comprehensively judge the paper, and that it is onerous to administrate [Ceci & Peters 1984].

To shed light on this controversy, we examine the now substantial scholarly literature regarding blind reviewing. This literature includes empirical studies from biomedicine, communication, computer science, economics, education, medicine, public health, physics, and psychology, retrospective analyses from computer science, ecology, economics, and medicine, and a quantitative meta-analysis from psychology. It is useful and instructive to learn what other disciplines, using diverse approaches, have discovered about blind reviewing.

In the following, we first define the various terms used in the literature. We then examine in some depth the general issues of fairness, review quality, and efficacy of blinding. As will be seen, in most cases the results are mixed. We end with a list of recommendations from scholarly societies and a brief summary of this complex sociological question.

2 Terminology

ACM defines a *refereed journal* or *refereed conference* as one that "is subjected to a detailed peer review, following a defined, formal process according to a uniform set of criteria and standards." [1]. This is distinguished from *formally reviewed material* ("subjected to a structured evaluation and critique procedure following a defined process uniformly applied as with refereeing, only without requiring that the tests of scholarly originality, novelty and importance be applied"), *reviewed* ("subjected to a more informal and not necessarily uniform process of volunteer review, with standards dependent upon the publication and the type of material"), *highly edited* ("professionally edited, usually by paid staff, with primary emphasis on exposition, graphic presentation, and editorial style rather than on content and substance"), and *unreviewed* ("published as submitted, with or without copyediting"). "Reviewing" in the present document refers to peer review for a refereed journal or conference.

Peer review is the use of predetermined reviewers, in the case of program committees, or ad hoc reviewers, in the case of reviewers for most journals, who individually read the submitted manuscript and prepare a written review. Sometimes, as in the case of some conference program committees, reviewers will subsequently either physically or electronically meet to discuss the papers to arrive at an editorial decision. For most journals, the Associate Editor handling the paper or the Editor-in-Chief will make the final editorial decision.

In the vast majority of refereed database conferences and journals, the identity of the reviewer(s) is not revealed to the author(s), ostensibly to ensure more objective reviewing. This is termed *single-blind reviewing* or, less frequently, "one-eyed review" [Rosenblatt & Kirk 1980]. (Incidentally, the terms "reviewer" and "referee" are used interchangeably in the literature.)

There are other sources of confidentiality in the review process. For most journals, the identity of the reviewer is not revealed to other reviewers; such is not the case for program committees. Some conferences, such as IEEE ICDE, utilize area program chairs. Generally but not always it is known which area program chair mediated the editorial decision for a submission. The associate editor for a journal submission is usually revealed to the author, except when one of the authors is himself/herself an

[1] ACM Policy on Pre-Publication Evaluation, at http://www.acm.org/pubs/prepub_eval.html

associate editor (cf. the *TODS* policy [Snodgrass 2003]). Of course, the Editor-in-Chief and the Program Chair are known to everyone. The important point here is that the term "single-blind reviewing" applies only to hiding the identity of the reviewer from the author.

In an effort to achieve more objective reviewing, a venue can also request that the identity of the author be removed from the submitted manuscript, a process termed *blinding the manuscript*. When the identity of the authors and their institutions is kept from the reviewers, this is termed double-blind reviewing. Note that the Editor-in-Chief and Program Chair, and generally the Associate Editor, are made aware of this information via a separate cover sheet not shared with the reviewers.

The psychological sciences utilize a different terminology that conveys a subtle philosophical shift. When the identities of the authors and reviewers are not revealed to each other, it is termed in these sciences a *masked* reviewing process. Note the symmetry of this terminology. The American Psychological Association *Guide to Preparing Manuscripts for Journal Publication* [Calfee & Valencia 2006] states, "Peer review is the backbone of the review process. Most APA journals, like the majority of other professional publications, practice anonymous, or masked, reviews. Authors and reviewers are unaware of each other's identities in most instances, an arrangement designed to make the process more impartial." The implication is that revealing either the reviewer's identity or the author's identity breaks the mask. Presumably single-blind reviewing would then be termed "non-masked," but the APA doesn't use the term. (The term "unmasking" denotes revealing the identity of a reviewer to a co-reviewer [van Rooyen 1999]; we don't consider that practice here.)

The present paper will use the terms single-blind and double-blind reviewing, as well as their respective three-letter acronyms, *SBR* and *DBR*.

Venues differ in who does the blinding/masking of a submission. We will use the term "author masking" when the author removes identification from the paper before submitting it and "editorial masking" when such identification (generally, author name and affiliation) is removed in the editorial process before sending the manuscript to the reviewers. Procedures differ in how aggressive is the required author masking and the actual editorial masking. Self-citations and other first-person references in the body of papers are generally retained in editorial masking. While author masking can be more thorough, because authors would know what kind of information is revealing, authors through various devious means can circumvent both kinds of masking.

3 Literature Reviews

Because of the centrality of peer review to the propagation of scientific knowledge, one would expect that peer review has been thoroughly studied, with its benefits and potential pitfalls exhaustively documented. Such is not the case. Prior to 1975 research on peer review was relatively scarce, with discussions based more on personal observations rather than systematic data gathering. Campanario has written the most comprehensive (61 page!) summary of the research that has been done on peer review, generally over the two decades of 1975–1995. Part 1 of this summary covered the participants in the system: the credentials of referees, editorial board members, and editors; how editors and editorial board members are appointed; how referees are chosen; reviewer incentives and tasks; and systemic problems of reliability, accuracy, and bias: reliability of review; accuracy of review; is the system biased towards positive results; and is the system biased against replication [Campanario 1998a]. Part 2 covered current research findings about fraud, favoritism, and self-interest in peer review [Campanario 1998b]; this part included a three-page section on DBR.

The Institute of Mathematical Statistics (IMS) formed the Ad Hoc Committee on Double-Blind Refereeing in February 1991, at least partly in response to a report of the New Researchers Committee [Altman et al. 1991]. This committee issued a report that contained a three-page literature review [Cox et al. 1993].

Other reviews include one of the voluminous literature (over 600 items) on the more general topic of journal reviewing, including several paragraphs related to DBR [Dalton 1995], another on 68 papers on empirical evidence concerning journal peer review [Armstrong 1997], including one page on DBR, and a summary of the evidence for the effectiveness of peer review in general, including about a page and a half on DBR [Fletcher & Fletcher 1997]. Finally, two international congresses on editorial peer review have been held, with papers revised and re-reviewed and appearing in the *Journal of the American Medical Association (JAMA)* [Rennie 1990, Rennie & Flanagin 1994]. Relevant papers from these congresses are discussed in the following sections.

These literature reviews emphasize three primary aspects relevant to blind reviewing, fairness to authors (to unknown authors or to authors affiliated with unknown institutions, to less-published or to proficient authors, to both genders), review quality, and blinding efficacy. The following sections will address each aspect in turn.

4 Fairness

The fundamental argument for double-blind reviewing is that it is fairer to authors (and thus, indirectly, to readers). The argument proceeds as follows: The judgment of whether a paper should be accepted for publication should be made on the basis of the paper alone: is what the submission states correct, insightful, and an advancement of the state-of-the-art? The editorial judgment should not be made on extenuating circumstances such as who wrote the paper or the professional affiliations of the authors. By blinding the submission, the reviewers cannot take these peripheral aspects, which are not relevant, into account in their review.

The analysis of fairness in the extant literature concerns (a) fairness to unknown authors or institutions, (b) fairness to prolific or to less-published authors, and (c) gender equity. There is also the related issue of the perception of fairness. The following sections will elaborate on each of these concerns.

4.1 Fairness to Unknown Authors or Institutions

Some evidence from retrospective and experimental studies suggest that when the authors' names and affiliations are known, reviewers may be biased against papers from unknown authors or institutions, termed "status bias" [Cox et al. 1993]. An anecdote illustrates this possibility. The psychologist Robert Rosenthal wrote of his experience in the prestigious journal *Behavioral and Brain Sciences* with "the 15 to 20 articles I had written while at UND [University of North Dakota] that I was not able to publish in mainstream psychological journals. 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" [Rosenthal 1982, page 235].

We now examine the studies that attempt to detect status bias, in chronological order.

A retrospective study of manuscripts that had been submitted to *The Physical Review* between 1948 and 1956 found that "some 91 per cent. of the papers by physicists in the foremost departments were accepted as against 72 per cent. from other universities" [Zuckerman & Merton 1971, page 85]. Two possible explanations were offered: status bias and "differences in the scientific quality of the manuscripts coming from different sources" [ibid].

An early experiment found that "the effect of institutional prestige failed to attain significance in any one of the measures" [Mahoney et al. 1978, page 70]. "Experimental manuscripts were sent to 68 volunteer reviewers from two behavioristic journals. ... Institutional affiliation was also manipulated on the experimental manuscripts, with half allegedly emanating from a prestigious university or a relatively unknown college" [ibid].

Another retrospective study, this of the records of reviews of a society which publishes research journals in two areas of the physical sciences, found large differences in how papers from minor and major universities are reviewed: "minor university authors are more frequently evaluated favourably (ie less critically) by minor university referees, while major university authors are more often evaluated favourably by major university referees than they are by those affiliated to minor universities. It would therefore appear that when referees and authors in these areas of the physical sciences share membership of national or institutional groups, the chances that the referees will be less critical are increased. ... Personal ties and extra-scientific preferences and prejudices might, of course, be playing a part as well. But it appears that, even in the absence of these personal factors, the scientific predispositions of referees still bias them towards less critical evaluation of colleagues who come from similar institutional or national groups, and so share to a greater extent sets of beliefs on what constitutes good research" [Gordon 1980, pp. 274–5].

Peters and Ceci performed a famous experiment [Peters & Ceci 1982] that gave some credence to the existence of such bias.[2] In this study, twelve papers published by investigators from prestigious and highly productive American psychology departments in high-quality journals were altered with fictitious names and institutions substituted for the original ones and then formally resubmitted to the journals that had originally refereed and published them 18 to 32 months earlier. Only three were detected as resubmissions; of the remaining nine, eight were rejected, in many cases based on "serious methodological flaws."

Peters and Ceci 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

[2] This experiment and the associated paper have generated much controversy. A special issue of *Behavioral and Brain Science* was dedicated to the paper and 55 (!) commentaries, along with an authors' response that was almost as long as the original paper. "In the course of the Commentary, just about every aspect of the peer-review problem is brought up and subjected to critical scrutiny" [Harnad 1982, page 186].

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" [Peters & Ceci 1982, page 192]. Their proposed solution: "If institutional affiliation or professional status can in fact bias peer review - and this bias proves to have no validity, or negative validity - then one possible solution to this problem (as several critics have recommended) would be to establish blind reviews as standard journal policy" [ibid, page 194].

A seminal experiment [Blank 1991] demonstrated status bias in reviewing more directly. In this experiment, every other paper that arrived at the *American Economic Review* was designated as double-blind. For these papers, an editorial assistant removed the name and affiliation of the author from the title page and typically scanned the first page for additional titles or notes that would identify the author (i.e., editorial masking). This experiment lasted for two years.

The relevant issue was "whether the ratio of acceptance rates between institutional ranks in the blind sample differs from the corresponding ratio in the nonblind sample." [Blank 1991, page 1053–1054]. It was found that this ratio did not differ for those at top-ranked departments and those at colleges and low-ranked universities. 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 nonblind sample; in some cases, the acceptance rate dropped by more than 7 percentage points. She found similar differences with referee ratings between SBR and DBR.

A retrospective study of single-blind reviews for the *Journal of Pediatrics* and published in *JAMA* found only partial evidence for status bias, that "for the 147 brief reports, lower institutional rank was associated with lower rates of recommendation for acceptance by reviewers ($P < .001$). ... For the 258 major papers, however, there was no significant relationship between institutional rank and either the reviewer's recommendations (P=.409) or the acceptance rate (P=.508)" [Garfunkel et al. 1994, page 138].

Another retrospective analysis of single-blind reviews also published in *JAMA* found evidence of status bias at a coarse geographical level [Link 1998]. In this analysis of original research articles submitted to *Gastroenterology* during 1995 and 1996, 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, page 246].

The experimental evidence is mixed concerning status bias present for top-ranked authors and institutions. The evidence is quite compelling that status bias is possible, perhaps prevalent, in SBR for most other authors and institutions, presumably for those papers most needing the critical evaluation of reviewers.

4.2 Fairness to Prolific Authors

There have been several studies that have looked at the impact of blinding on prolific authors, with conflicting results.

The Mahoney experiment discussed in the previous section also suggested that "self-citation may be a determinant of a reviewer's evaluation of a manuscript" [Mahoney et al. 1978, page 70]. In half of the papers sent to volunteer reviewers, "the author defended his contentions by referencing three of his own "in press" publications. For the other half, these same prepublication references were also cited, but were attributed to someone else. ... Reviewers rated the article as more innovative and publishable if the fictitious author included self-references in the manuscript than if no self-references were included" [ibid].

An experiment published in *JAMA* on 57 consecutive manuscripts submitted to the *Journal of Development and Behavioral Pediatrics* that were randomly assigned to either blinded or unblinded review (that is, using editorial masking) found that "contrary to the original hypothesis of this study, senior authors with more previous articles received significantly better scores from the blinded reviewers (r=−.45), but not from the nonblinded reviewers (r=−.14)" [Fisher et al. 1994, page 145]. The authors "interpret 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" [ibid, page 146].

A retrospective study of two database conferences [Madden & DeWitt 2006] found no impact on prolific authors. This study considered papers authored by those designated variously as a "famous person" or "prolific researcher" or "more senior researcher," defined as "those individuals who have published 20 or more papers in SIGMOD and VLDB conferences" [ibid, page 29]. The analysis found that DBR reviewing (with author masking) in the ACM SIGMOD conference "has had essentially no impact on the publication rates of more senior researchers in the database field" [ibid, page 30]. This result mirrors those of Blank's study, which found that acceptance rates for top-ranked institutions (where presumably most of these prolific researchers resided) were not affected by DBR.

An independent analysis of this data, using medians rather than means, reached the opposite conclusion: "that double blind review in SIGMOD do have its impact on the performance of 'famous person' " [Tung 2006]. The reason for the differing conclusions over the same data may be that this data, consisting of yearly counts of papers by prolific authors and by others, is too coarse to make a final determination, as it doesn't taken into consideration the varying submission rates of individuals and the varying participation by prolific authors.

These contradictory results render it impossible to say anything definitive about the impact of blinding on prolific authors. However, there does seem to be evidence of some kinds of bias with SBR.

4.3 Gender Equity

When reviewers know the identity of the author(s) of the submitted manuscript, gender bias is also a possibility. Several disciplines have launched in-depth studies based on concerns of gender equity.

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 [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 them rated essays bearing the names of male authors (e.g., John T. McKay), whereas the other half rated the same essays with the names of female authors (e.g., Joan T. McKay). The results indicated that those essays where the author was male was rated higher.

A quantitative meta-analysis of this and similar studies (over one hundred) over the intervening two decades found that "the average difference between ratings of men and women is negligible" [Swim et al. 1989, page 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. Interestingly, "there was some indication, however, that women will be rated less favorably than men when less information is presented" [ibid, page 421] and "there was also some indication of greater bias when the stimulus material was a resumé or application" [ibid, page 422]. More relevant to the issue of peer review is the observation from the APA task force report that when "Joan and John's work was high in quality, the effect size was close to zero (-.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 (-.24). ... these 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, page 45]. This is again consistent with previously-discussed studies that considered status bias.

Blank's experiment, described earlier, was in fact initiated due to concerns of gender bias. The *American Economic Review* journal had employed SBR for most of its recent history, except during a period of 1973–1979 when the then-current editor adopted DBR [Borts 1974]. In the mid-1980's, 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, page 1045]. As a result, Blank was asked by the current editor of the *AER* and the Board of Editors to design and run a randomized experiment looking into this potential effect.

Due to the careful randomization design of this experiment, one can compare acceptance rates between the blind and nonblind samples, and indeed, there were striking differences. "For women, there is no significant difference in acceptance rates between the two samples. For men, acceptance rates are significantly higher in the nonblind sample." [ibid, page 1053]. When reviewers knew that that paper was authored by a male, they accepted a higher percentage (15%, versus 11%) than if the paper was blinded. "One can compare acceptance rates between the blind and nonblind samples without other control variables because the randomization process guarantees that papers by women (and men) in each sample have identical distributions of characteristics" [ibid].

Blank emphasized the core issue: "whether the *ratio* of male to female acceptance rates in the nonblind sample is different from that in the blind sample. 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" [ibid]. Here the results were statistically insignificant, perhaps because there were too few observations of papers authored by women.

Would DBR result in a large increase in acceptances of papers by women? "While there is some indication in these data that women do slightly better under a double-blind system, both in terms of acceptance rates and referee ratings, these effects are relatively small and statistically insignificant. Thus, this paper provides little evidence that moving to a double-blind reviewing system will substantially increase the acceptance rate for papers by female economists" [ibid, page 1063]. Interestingly, the *American Economic Review* now employs DBR.

The Modern Language Association's (MLA) experience was striking: going to DBR resulted in a large increase in acceptances by female authors. "Contributed

papers at MLA meetings had first to survive a review stage before acceptance to be read. Prior to 1974, these papers were refereed with the author's name intact. In 1974, double-blind refereeing was tried with the effect that the number of women and of new investigators having papers accepted doubled from previous years. This number doubled again when repeated in 1975, until, by 1978, the proportion of acceptances among women and new researchers was comparable to that for men. The MLA Board subsequently decided in 1979 to use double-blind refereeing for all their publications" [Billard 1993, page 321]. The impetus for this change was the perception of gender bias. "A number of women complained to the Modern Language Association in the United States that there were surprisingly few articles by women in the association's journal, compared to what would be expected from the number of women members. It was suggested that the review processes were biased. The association vigorously denied this but under pressure instituted a blind reviewing procedure under which the names of the authors and their institutional affiliations were omitted from the material sent to the reviewer. The result was unequivocal: There was a dramatic rise in the acceptance of papers by female authors" [Horrobin 1982, page 217].

It is possible that the small observed effect in Blank's study (in contrast to the MLA experience) was due to the low number of submissions by women to *AER*.

These studies show that revealing author identity, specifically the gender of the author, can sometimes have an effect on acceptance rates.

4.4 The Perception of Fairness

A *perception* of possible bias may be just as damaging as actual bias.

The Institute of Mathematical Statistics (IMS) New Researchers' Committee (NRC) report stated, "The NRC feels that the current system [SBR] has the potential for bias or perceived bias against NRs [new researchers], women and identifiable minorities, (a disproportionate number of the latter two categories are NRs)" [Altman et al. 1991, page 165]. In a response to discussants of that report, the NRC reasserted a year later, that "much of the value of double-blind refereeing lies in the community perception of fairness" [Altman et al. 1992, page 266].

The experience with this controversy at the IMS indicated a split between new researchers, which "strongly endorses double-blind refereeing. ... It seems likely that [this] represents the majority opinion among new researchers, although support for double-blind refereeing is not unanimous among new researchers" and senior members: " 'negative but sympathetic' ... seems to be a majority view among those senior enough to have been involved in the editing process" [Cox et al. 1993, page 311]. However, a survey to IMS members "indicates strong support for double-blind refereeing in the IMS journals" [ibid].

A responder to the IMS report [Cox et al. 1993] stated, "Refereeing 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. Whether 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, page 324] (emphasis in original).

5 Quality of Reviews

Does blinding impact the quality of reviews? Two counter-balancing effects have been claimed. One possibility is that SBR, by revealing the authors' identity, increases the quality of reviews by supplying to the peer reviewer relevant information about the prior accomplishments and about publication and citation rates. On the other hand, such identity information might be used by reviewers as short cuts, thus reducing review quality. Perhaps DBR, in not revealing the author's identity, permits the reviewer to focus more on the paper itself, which can increase the quality of the review. We examine the scientific evidence of each effect in turn. Overall, the evidence is mixed on both effects.

One study [Abrams 1991] investigated a related question: whether peer evaluations of grant proposals are the best available predictor of future output of influential science. While reviewing grant proposals is different than reviewing papers submitted to refereed conferences and journals, some of these results do have relevance here.

This study examined the commonly-held perception that "individual scientists seldom fluctuate between periods of producing large amounts of good work and periods when they produce only a small amount of poor-quality work. This is the basis for hiring and promotion decisions at universities and research institutions" [ibid, pp. 112–3]. The study found a high degree of correlation. Data concerning the members of the US National Science Foundation Ecology Panel in 1988 was examined. Among the eleven senior members of this panel, high correlations between citation rates over time as well as high temporal correlations in publication rates were observed. For a group of forty-five scientists drawn from a 1973 ecology textbook who had published papers in major ecology journals, high temporal consistency among the top-ranked individuals was again observed. The conclusion is 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" [ibid, page 115].

There is evidence that even with this potentially useful information, journals using SBR "publish a larger fraction of papers that should not have been published than do journals" using DBR [Laband 1994, page 147]. This study used nonlinear regression and ordered probit techniques to estimate the impact of DBR on citations of a sample of 1051 articles published in 28 economics journals during 1984. The analysis found that "articles reviewed single-blind are less likely than those reviewed double-blind to be identified correctly as the highest-impact articles (those with nine or more citations in the ensuing 5 years). By the same token, articles reviewed single-blind are more likely than those reviewed double-blind to be misidentified as the lowest-impact papers (those with no citations in the ensuing 5 years). ... We conclude that the single-blind review process apparently suffers from a type I error bias to a greater extent than the double-blind review process." [ibid, page 149].

Several limitations of the Laband study have been pointed out: "There are difficulties with this analysis, the main one being that the papers considered were only reviewed in one way, either blinded or not blinded. Also, controlling for the status of the journal in which each papers appeared is inevitably a difficult process. Papers selected for the 'market leader' journals by whatever process must be more likely to be cited than those selected for more specialised or less well respected competitors" [Poutney 1996, page 1059].

The evidence is thus very mixed about whether information about prior accomplishments, coupled with the observed correlation with future accomplishments, results in better judgment about a specific submission before a reviewer. "Some argue 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, page 295]. It has also been observed that "Some referees believe that they can judge better if they know the author because the manuscript can be evaluated in the context of the author's entire corpus of work, but this claim is rare. More frequent is advocacy of anonymity for authors" [Dalton 1995, page 236]. Another asserted, "it should be the work itself, and not the reputation of the author, which influences ... As statisticians, one of our maxims is that the data should speak for themselves, so likewise should we let the work speak for itself without undue influence from outside pressures" [Billard 1993]. We now examine the scientific evidence that DBR can increase review quality.

A study of sixty articles drawn from the *Journal or Abnormal Psychology* found that "the articles by scholars affiliated with high-status institutions were cited considerably more often than the articles by scholars at low-status institutions" [Perlman 1982]. "Therefore, it appears that an institution's prestige is a valid predictor, and editors may be justified in using this as a factor in their decision making. Advocates of blind review, however, may still object to using either institutional affiliation or an individual's reputation as criteria in selecting articles. They could claim that the excellence of the manuscript should not only be apparent over time, it should also be immediately apparent without the aid of status cues. Thus, even with a blind review process, assessors should identify a higher proportion of items submitted by scholars at prestigious institutions as worthy of publication" [ibid].

Another double-blind study of DBR, 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, page 1375]; see also [Evans et al. 1990]. Specifically, "editors graded the quality of blinded reviews better on three of the four quality dimensions from the editor's perspective: importance of the question, targeting key issues, and methods (all $P < .02$). The greatest difference was noted in the grades for methods. Editors graded blinded reviews from the author's point of view statistically significantly better on only one of the five quality measures: the blinded reviewer was graded as more knowledgeable (P=.05). The grades on the other dimensions of quality favored the blinded reviewer, except for courteousness. The editors' summary grades, taking into account both editor's and author's points of view, favored the blinded reviewers. The mean summary grade was 3.5 for blinded reviewers and 3.1 for unblinded reviewers. The mean difference between the blinded and unblinded reviewers was 0.41 (P=.007). The difference between the median grade for blinded and unblinded reviewers was 4.0 vs 3.0, respectively, an entire grade" [McNutt et al. 1990, pp. 1373–4]. Two limitations have been noted [Fletcher & Fletcher 1997]. "One difficulty with 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, page 316]. "The study recognized that the results may have been influenced by the nature of the journal – not a market leader, but with a very wide editorial remit – and that quite different results might be found in similar analyses of large journals, sub-specialty journals and basic science journals" [Poutney 1996, page 1059].

However, four other studies did not find that masking peer reviewers to author identity improves the quality of peer review.

In an experiment also published in *JAMA*, two randomizations were performed for submissions

over a six-month period to five biomedical journals [Justice et al. 1998]. The first assigned a quarter of the submissions to the journal's usual practice, which for four of the journals was SBR. For the rest of the submissions, one of the two reviewers was randomly selected to receive a manuscript that had been masked, by "removing author and institutional identity from the title page, running headers or footers, and acknowledgments of the manuscripts. Self-references in the text were not removed" (that is, editorial masking) [ibid, page 241]. Questionnaires were provided to editors, authors, and reviewers. Analysis of these questionnaires revealed that "authors and editors perceived no significant difference in quality between masked and unmasked reviews. We also found no difference in the degree to which the review influenced the editorial decision. ... When analysis was restricted to manuscripts that were successfully masked, review quality as assessed by editors and authors still did not differ" [ibid, page 240].

Three other studies [Tobias & Zibrin 1978, van Rooyen 1999, Smith et al. 2002], in the fields of specialized medicine and education, found similar, negative results

It is important to note that these studies utilized editorial masking. We will show shortly that masking success depends highly on how that masking is done. The authors of the *JAMA* study speculate that "poor overall masking success, in combination with the observation that an author's renown is strongly associated with masking failure," may contribute to this lack of a difference between unblinded and blinded reviews [Justice et al. 1998, page 242].

One must conclude that the jury is still out. It has not been shown convincingly that either SBR or DBR can, by revealing or by hiding the identity of the author and institution, increase the quality of the reviews of a submitted manuscript.

6 Efficacy of Blinding

Any benefits ascribed to double-blind reviewing assume that the blinding of the submitted manuscript has been successful, that reviewers cannot in fact identify the author(s) nor their institutions. "How 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, page 236]. "The notion that an experimented 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, page 295].

We earlier differentiated editorial and author blinding. Each can be done in various ways. "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, page 781].

Several studies have examined how well blinding works. These studies, across a wide range of disciplines, observed that blinding achieved success rates of 53% to 79%. We now review the studies chronologically.

A study of reviewers of papers 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 & Kirk 1980, page 389]. This works out to successful blinding 65% of the time, for editorial blinding in which the names of authors and their institutional affiliations are removed from the title page.

One retrospective study over six journals employing DBR with author blinding and representing a broad range of areas in psychology 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 multiauthored papers. ... There were no significant differences in the proportion of correct detections among the six journals, ranging from 26% to 42% ... nor was there any relationship between detection accuracy and the number of years of reviewing experience" [Ceci & Peters 1984, page 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" [ibid].

In the McNutt study described in the previous section, editorial blinding was successful to institution name for 73% of the reviewers and to author(s)' names for 76% of thr reviewers [McNutt et al. 1990]. The blinding process was of moderate thoroughness: "An editorial assistant copied each manuscript, retyped the title page, and removed authors' and institutions' identifiers using an opaque tape. To do this, the entire manuscript was scanned—headers and footers, body of text, tables, and figures. She made no attempt to remove references to the author's own work. ... Minimal changes were made, on average, to the body of the manuscript" [ibid, page 1372].

Submissions to the *American Journal of Public Health (AJPH)* are partially author blinded and partially editori-

ally blinded: "Contributors to *AJPH* are instructed to submit a second face sheet which includes only the title of the paper. These instructions are usually followed. We remove acknowledgments, but make no further effort to remove identifying page headers (when present contrary to instructions) or to change the text or references. We have been aware that a substantial portion of our manuscripts are not truly blinded because of text allusions and self-referencing" [Yankauer 1991, page 843–4]. A questionnaire revealed that "blinding could be considered successful 53% or 61% of the time, depending on whether successful blinding ignores identification or includes only correct identification. ... Self-referencing (61.8%) and personal knowledge (38.2%) were the two clues given for identification of author and/or institution. In both cases 16% of the indentifications were incorrect" [ibid, page 844].

Blank's 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. (Multiple-author papers are considered to be correctly identified in any of the author's names are known.)" [Blank 1991, page 1051] Note that in this experiment the papers were editorially blinded only by changing the first or second pages.

The study by Fisher et al. discussed in the previous section also considered (editorial) masking success, "in which the cover page and any identifying data on the top of bottom of each page had been removed; so as not to alter the quality of the manuscript, no effort was made to delete information in the text when the authors might have identified themselves" [Fisher et al. 1994, page 144]. 54% of reviewers were thereby successfully blinded.

The study by Justice et al. also discussed in the previous section found that with (editorial) masking, "manuscripts by authors with whom the unmasked reviewer was familiar ... were less likely to be successfully masked (53%) (that is, the masked reviewer was more likely to correctly guess author identity) than those of authors who were not known to the unmasked reviewer (79%)" [Justice et al. 1998, page 241].

Another experiment in *JAMA* found that "a long-standing policy of masking did not increase masking success" [Cho et al. 1998, page 243]. This study included four medical journals that did not mask author identity and three medical journals with a policy of DBR. Papers were then editorially blinded: "Each journal masked eligible manuscripts 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" [ibid, page 244]. 60% of reviewers were masked. "There was no significant difference in masking success between journals with a policy of masking and those without (P=.92)" [ibid, page 245].

Another randomized study from a few years ago in a medical journal showed that "with successful blinding defined as either author not identified or author identified incorrectly, 170 reviewers (58%) were successfully blinded" [van Rooyen 1999, page 235]. Blinding was editorial: "Blinding consisted of removing authors' details from the title page and acknowledgments. No attempt was made to remove authors' details from within the text of the manuscript, the illustrations, or the references" [ibid, page 234].

Most of these studies utilized editorial masking, achieving success rates of 53% to 79%, with a gross average across studies of 62%. The success rate was lower for known authors. Self-referencing was a major clue to reviewers. And there were incorrect guesses. So even minimal editorial blinding can be somewhat effective. "Clearly, the feasibility and success of blinding depends both on the amount of effort put into the blinding process and on factors related to the type and circulation of the involved journal" [Fisher et al. 1994, page 145].

A data mining experiment took a different tack, seeing whether a computer program could identify authors using only the citations included in the paper. Two automatic methods for author identification were considered: "(1) a (dynamic) vector-space model that represents both papers and author histories, and (2) tallying (discriminative) self-citations." [Hill & Provost 2003, page 179]. A very large archive of physics papers gathered as part of the KDD Cup 2003 competition was used in the study. "The self-citation based methods generally worked better. However, the vector-space models are able to match (with much lower accuracy) even when self-citations are removed. With the best method, based on discriminative self-citations, authors can be identified 45% of the time. Additionally, the top-10% most prolific authors can be identified 60% of the time. ... authors with 100 or more prior publications can be identified 85% of the time" [ibid].

Blank's conclusions apply to the many studies generally. "On the one hand, a substantial fraction—almost half—of the blind papers in this experiment could be identified by the referee. This indicates the extent to which no reviewing system can ever be fully anonymous. On the other hand, more than half of the papers in the blind sample *were* completely anonymous. A substantial fraction of submitted papers are not readily identified by reviewers in the field. ... Those blind papers that are correctly identified by the referees ... are skewed in favor of authors who are better known or who belong to networks that distribute their working papers more widely" [Blank 1991, pp. 1051–2]. Ceci and Peters conclude that "Although there are occasional lapses in the preparation of manuscripts by au-

thors and failures to screen manuscripts by editorial staff, we are impressed by the overall efficiency of blind review" [Ceci & Peters 1984, page 1494].

7 Recommendations of Scholarly Societies

Several scholarly societies have weighed in on this question.

An examination published in the *Journal of Business Ethics* of past abuses of the editorial process warranted the proposal to "use the double blind process in which referees do not know authors nor authors, referees" so as to "protect referees from influence and participants from damage in a negative decision" [Carland et al. 1992, page 103].

The Institute of Mathematical Statistics (IMS) New Researchers' Committee (NRC) report stated, "after extensive discussion, the consensus of the NRC is that the advantages of the double-blind system outweigh the costs, and we recommend that IMS journals evaluate the benefits of adopting such a system" [Altman et al. 1991, page 166]. In a response to discussants of this report, the NRC reasserted a year later, that "the NRC strongly supports double-blind refereeing for its potential to remove separate consideration, perceived or otherwise" [Altman et al. 1992, page 266].

The view of the IMS Ad Hoc Committee on Double-Blind Refereeing in 1993 "may be summarized as cautiously receptive to double-blind refereeing. We are not convinced that the benefits outweigh the disadvantages; but we are open to the possibility. We recommend that if a change in journal policy is contemplated that an experiment be conducted to assess the merits of double-blind refereeing before any permanent change is made" [Cox et al. 1993, page 311].

The APA Task Force on Women in Academe states that "Because of the potential for bias, APA has mandated that editors of APA journals offer masked review as an option; however, mandatory masked review of articles should be instituted as policy" [Fouad et al. 2000, page 34]. The report includes in its recommendations, to "Implement a policy of mandatory masked review for all APA peer-reviewed publications" [ibid, page 45]. The APA Guide states, "Most APA journals, like the majority of other professional publications, practice anonymous, or masked reviews" [Calfee & Valencia 2006]. Of the 47 APA journals that accept paper submissions, almost half require masking and most of the rest allow it on request.

There is a pattern here. "It is therefore only to be expected that senior established researchers will tend to seek the status quo, being less inclined to want to move to double-blind refereeing, while new (and also women and researchers in lower status named institutions) researchers will tend to prefer that double-blind refereeing be introduced" [Billard 1993, page 322].

8 Prevalence

Disciplines vary widely in their use of single- and double-blind review, but the historical trend is clear.

A non-randomized survey in 1989 showed that "in chemistry, physics, math, and psychology, the responding journals indicate that they use a single-blind reviewing system. ... Biology appears to have both single-blind and double-blind journals, as does history and anthropology. Political-science and sociology journals report uniformly using double-blind reviewing methods" [Blank 1991, pp. 1043–45]. Overall, considering 37 journals in nine disciplines, 79% were single-blind and 21% were double-blind.

"Four surveys of the frequency of binding have been published, but none is based on a random sampling of journals, and their size and response rates often leave something to be desired. Their results suggest that the majority of scientific journals do not practice blind review and that blinding may be more common in the social sciences than in the physical and medical sciences" [Yankauer 1991, page 843].

A more recent survey of 553 journals selected from eighteen disciplines revealed that DBR is increasing in prevalence, in comparison with previous studies: "Across 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 & Sawallis 2003, page 54]. Within computer science (29 journals, 15 responding), 57% were single-blind and 43% were double-blind.

Within ACM, all journals are single-blind and about 80% of ACM conferences are single-blind (as of 2000 [Snodgrass 2000]).

Similarly, the database field has traditionally relied on single-blind reviewing. Until recently, all of its conferences and journals have been single-blind. The ACM SIGMOD conference adopted double-blind reviewing in 2001 [ibid].

9 Summary

"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" [Genest 1993, page 324].

As the noted statistician Lynne Billard, who has written extensively on this topic, has remarked, "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, page 320].

We have attempted here to summarize the many studies of the varied aspects of blind reviewing within a large number of disciplines.

Concerning the central issue of fairness, Blank's summary in 1991 of the literature still holds true fifteen years later. "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" [page 1045]. Many studies provide evidence that DBR is fairer to authors from less-prestigious institutions and to women authors. Such differences are likely to matter even more for highly-selective conferences and journals.

Concerning quality of reviews, it is not known definitely whether either SBR or DBR results in a higher quality of reviews.

Most of the studies discussed here utilize editorial blinding, which has been shown to be successful about 60% of the time, across many disciplines. Removing text allusions and self-citations would increase success rates to perhaps 75%.

The prevalence of DBR has increased dramatically over the last fifteen years, to the point where most scientific journals now employ double-blind reviewing.

There is an administrative cost to DBR, to the journal as well as to the author. These costs vary depending on how the blinding is done, with the efficacy directly related to the effort expended.

Journals strongly desire to fairly evaluate submitted manuscripts, while simultaneously keeping costs in control. The policy question before each journal and each scholarly publisher is thus the following. Is the documented benefit of equity worth the administrative cost? At what price fairness?

Acknowledgments

The author thanks Tamer Özsu for comments on a previous version and Merrie Brucks for her help throughout.

References

[Abrams 1991] P. A. Abrams, "The Predictive Ability of Peer Review of Grant Proposals: The Case of Ecology and the US National Science Foundation," *Social Studies of Science* 21(1):111–132, February 1991.

[Altman et al. 1991] N. Altman, D. Banks, P. Chen, D. Duffy, J. Hardwick, C. Léger, A. Owen, and T. Stukel, "Meeting the Needs of New Statistical Researchers," *Statistical Science* 6(2):163–174, May 1991.

[Altman et al. 1992] N. Altman, J. F. Angers, D. Banks, D. Duffy, J. Hardwick, C. Léger, M. Martin, D. Nolan (Chair), A. Owen, D. Politis, K. Roeder, T. N. Stukle and Z. Ying, "Rejoinder," *Statistical Science* 7(2):265–266, May 1992.

[Armstrong 1997] J. S. Armstrong, "Peer Review for Journals: Evidence on Quality Control, Fairness, and Innovation," *Science and Engineering Ethics* 3:63–84, 1997.

[Bachand & Sawallis 2003] R. G. Bachand and P. P. Sawallis, "Accuracy in the Identification of Scholarly and Peer-Reviewed Journals and the Peer-Review Process Across Disciplines," *The Serials Librarian* 45(2):39–59, 2003.

[Billard 1993] L. Billard, "Comment," *Statistical Science* 8(3):320–322, August 1993.

[Calfee & Valencia 2006] R. C. Calfee and R. R. Valencia, "APA Guide to Preparing Manuscripts for Journal Publication," `http://www.apa.org/journals/authors/guide.html`, viewed June 21, 2006.

[Blank 1991] R. M. Blank, "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The American Economic Review," *American Economic Review* 81(5):1041–1067, December 1991.

[Borts 1974] G. Borts, "Report of the Managing Editor," *American Economic Review* 64:476–82, May 1974.

[Campanario 1998a] J. M. Campanario, "Peer Review for Journals at It Stands Today—Part 1," *Science Communication* 19(3):181–211, March 1998.

[Campanario 1998b] J. M. Campanario, "Peer Review for Journals at It Stands Today—Part 2," *Science Communication* 19(4):277–306, June 1998.

[Carland et al. 1992] J. A. Carland, J. W. Carland, and C. D. Aby, Jr., "Proposed Codification of Ethicacy in the Publication Process," *Journal of Business Ethics* 11:95–104, 1992.

[Ceci & Peters 1984] S. J. Ceci and D. P. Peters, "How blind is blind review?" *American Psychologist* 39:1491–1494, 1984.

[Cho et al. 1998] M. K. Cho, A. C. Justice, M. A. Winker, J. A. Berlin, J. F. Waecklerle, M. L. Callaham, and D. Rennie, "Masking author identity in peer review—What factors influence masking success?" *Journal of the American Medical Association* 280(3):243–245, July 15, 1998.

[Cox et al. 1993] D. Cox, L. Gleser, M. Perlman, N. Reid, and K. Roeder, "Report of the Ad Hoc Committee of Double-Blind Refereeing," *Statistical Science* 8(3):310–317, August 1993.

[Dalton 1995] M. Dalton, "Refereeing of Scholarly Works for Primary Publishing," in **Annual Review of Information Science and Technology (ARIST)**, Volume 30, M. E. Williams (ed.), American Society of Information Science, 1995.

[Evans et al. 1990] A. T. Evans, R. A. McNutt, R. H. Fletcher, and S. W. Fletcher, "The Effects of Blinding on the Quality of Peer Review: A Randomized Trial," *Clinical Research* 38(2), 1990.

[Fisher et al. 1994] M. Fisher, S. B. Friedman, and B. Strauss, "The effects of blinding on acceptance of research papers by peer review," *Journal of the American Medical Association* 272(2):143–146, July 13, 1994.

[Fletcher & Fletcher 1997] R. H. Fletcher and S. W. Fletcher, "Evidence for the Effectiveness of Peer Review," *Science and Engineering Ethics* 3(1):35–50, 1997.

[Fouad et al. 2000] N. Fouad, S. Brehm, C. I. Hall, M. E. Kite, J. S. Hyde, and N. F. Russo, **Women in Academe: Two Steps Forward, One Step Back**, Report of the Task Force on Women in Academe, American Psychological Association, 2000. `http://www.apa.org/pi/wpo/academe/repthome.html`, viewed June 21, 2006.

[Franzini 1987] L. R. Franzini, "Editors Are Not Blind," *American Psychologist*, page 104, January 1987.

[Garfunkel et al. 1994] J. M. Garfunkel, M. H. Ulshen, H. J. Hamrick, and E. E. Lawon, "Effect of Institutional Prestige on Reviewers' Recommendations and Editorial Decisions," *Journal of the American Medical Association* 272(2):137–138, July 13, 1994.

[Genest 1993] C. Genest, "Comment," *Statistical Science* 8(3):323–327, August, 1993.

[Goldberg 1968] P. Goldberg, "Are some women prejudiced against women?" *Transaction* 5:28–30, April, 1968.

[Gordon 1980] M. D. Gordon, "The role of referees in scientific communication," in **The Psychology of Written Communication: Selected Readings**, J. Hartley (ed.), London, England: Kogan Page, pp. 263–275, 1980.

[Harnad 1982] S. Harnad, "Peer commentary on peer review," *Behavioral and Brain Sciences* 5(2):185–186, 1982.

[Hill & Provost 2003] S. Hill and F. Provost, "The Myth of the Double-Blind Review," *SIGKDD Explorations* 2(5):179–184, December 2003.

[Horrobin 1982] D. F. Horrobin, "Peer review: A philosophically faulty concept which is proving disastrous to science," *Behavioral and Brain Sciences* 5(2):217–218, 1982.

[Justice et al. 1998] A. C. Justice, M. K. Cho, M. A. Winker, J. A. Berlin, and D. Rennie, "Does Masking Author Identity Improve Peer Review Quality?" *Journal of the American Medical Association* 280(3):240–242, July 15, 1998.

[Laband 1994] D. N. Laband, "A Citation Analysis of the Impact of Blinded Peer Review," *Journal of the American Medical Association* 272(2):147–149, July 13, 1994.

[Link 1998] A. M. Link, "US and Non-US Submission: An Analysis of Review Bias," *Journal of the American Medical Association* 280(3):246–247, July 15, 1998.

[Madden & DeWitt 2006] S. Madden and D. DeWitt, "Impact of Double-Blind Reviewing on SIGMOD Publication Rates," *ACM SIGMOD Record* 35(2):29–32, June 2006.

[Mahoney et al. 1978] M. J. Mahoney, A. E. Kazdin, and M. Kenigsberg, "Getting Published," *Cognitive Therapy and Research* 2(1):69–70, 1978.

[McGiffert 1988] M. McGiffert, "Is Justice Blind? An Inquiry into Peer Review," *Scholarly Publishing* 20(1):43–48, October 1988.

[McNutt et al. 1990] R. A. MuNutt, A. T. Evans, R. H. Fletcher, and S. W. Fletcher, "The Effects of Blinding on the Quality of Peer Review," *Journal of the American Medical Association* 263(10):1371–1376, March 9, 1990.

[Perlman 1982] D. Perlman, "Reviewer "bias": Do Peters and Ceci protest too much?" *Behavioral and Brain Sciences* 5(2):231–232, 1982.

[Peters & Ceci 1982] D. P. Peters and S. J. Ceci, "Peer-Review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again," *Behavioral and Brain Sciences* 5(2):187–195, 1982.

[Pitkin 1995] R. M. Pitkin, "Blinded Manuscript Review: An Idea Whose Time Has Come?" Editorial, *Obstetrics and Gynecology* 85(5 Part 1):781–782, May 1995.

[Poutney 1996] M. Poutney, "Blinded Reviewing," Editorial, *Developmental Medicine and Child Neurology* 38(12):1059–1060, December 1996.

[Rennie 1990] D. Rennie, "Editorial Peer Review in Biomedical Publication: The First International Conference," *Journal of the American Medical Association* 263(10):1317, March 9, 1990.

[Rennie & Flanagin 1994] D. Rennie and A. Flanagin, "The Second International Congress on Peer Review in Biomedical Publication," *Journal of the American Medical Association* 272(2):91, July 13, 1994.

[Rosenblatt & Kirk 1980] A. Rosenblatt and S. A. Kirk, "Recognition of Authors in Blind Review of Manuscripts," *Journal of Social Service Research* 3(4):383–394, Summer 1980.

[Rosenthal 1982] R. Rosenthal, "Reliability and Bias in Peer-Review Practices," *Behavioral and Brain Sciences* 5(2):235–236, 1991.

[Smith et al. 2002] J. A. Smith, R. Nixon, A. J. Bueschen, D. D. Venable, and H. H. Henry, "The Impact of Blinded versus Unblinded Abstract Review on Scientific Program Content," *Journal of Urology* 168(5):2123–2125, November 2002.

[Snodgrass 2000] R. T. Snodgrass, "Chair's Message," *ACM SIGMOD Record* 29(1):3, March 2000.

[Snodgrass 2003] R. T. Snodgrass, "Developments at *TODS*," *ACM SIGMOD Record* 32(4):14–15, December 2003.

[Swim et al. 1989] J. K. Swim, E. Borgida, G. Maruyama, and D. G. Myers, "Joan McKay versus John McKay: Do gender stereotypes bias evaluations?" *Psychological Bulletin* 105(3):409–429, 1989.

[Tobias & Zibrin 1978] S. Tobias and M. Zibrin, "Does Blind Reviewing Make a Difference?" *Educational Researcher* 7(1):14–16, January 1978.

[Tung 2006] A. K. H. Tung, "Impact of Double Blind Reviewing on SIGMOD Publication: A More Detail Analysis," 2 pages, July 2006.

[van Rooyen 1999] S. van Rooyen, F. Godlee, S. Evans, R. Smith, and N. Black, "Effect of Blinding and Unmasking on the Quality of Peer Review: A Randomized Trial," *Journal of General Internal Medicine* 14(10):622–624, October 1999.

[Yankauer 1991] A. Yankauer, "How Blind is Blind Review?" *American Journal of Public Health* 81(7):843–845, July 1991.

[Zuckerman & Merton 1971] H. Zuckerman and R. K. Merton, "Patterns of Evaluation in Science: Institutionalization, Structure and Functions of the Referee System," *Minerva* 9(1):66–100, January 1971.