

Bias in Peer Review

Carole J. Lee

Department of Philosophy, University of Washington, 361 Savery Hall, Seattle, WA 98195. E-mail: c3@uw.edu

Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin

School of Library and Information Science, Indiana University, 1320 East Tenth Street, Bloomington, IN 47405. E-mail: {sugimoto, guozhang, bcronin}@indiana.edu

Research on bias in peer review examines scholarly communication and funding processes to assess the epistemic and social legitimacy of the mechanisms by which knowledge communities vet and self-regulate their work. Despite vocal concerns, a closer look at the empirical and methodological limitations of research on bias raises questions about the existence and extent of many hypothesized forms of bias. In addition, the notion of bias is predicated on an implicit ideal that, once articulated, raises questions about the normative implications of research on bias in peer review. This review provides a brief description of the function, history, and scope of peer review; articulates and critiques the conception of bias unifying research on bias in peer review; characterizes and examines the empirical, methodological, and normative claims of bias in peer review research; and assesses possible alternatives to the status quo. We close by identifying ways to expand conceptions and studies of bias to contend with the complexity of social interactions among actors involved directly and indirectly in peer review.

Nature and Purpose of Peer Review

Peer review is an established component of professional practice, the academic reward system, and the scholarly publication process. The fundamental principle is straightforward: experts in a given domain appraise the professional performance, creativity, or quality of scientific work produced by others in their field or area of competence. In most cases, reviewer identity is hidden (single-blind review) to encourage frank commentary by protecting against possible reprisals by authors; and, in some cases, author identities

will be masked from reviewers (double-blind review) to protect against forms of social bias. The structure of peer review is designed to encourage peer impartiality: typically, peer review involves the use of a “third party” (Smith, 2006, p. 178), someone who is neither affiliated directly with the reviewing entity (university, research council, academic journal, etc.) nor too closely associated with the person, unit, or institution being reviewed; and peers submit their reviews without, initially at least, knowledge of other reviewers’ comments and recommendations. In some cases, however, peers will be known to one another, as with in vivo review, and may even be able to confer and compare their evaluations (e.g., members of a National Science Foundation [NSF] review panel).

Peer review, broadly construed, covers a wide spectrum of activities, including but not limited to observation of peers’ clinical practice; assessment of colleagues’ classroom teaching abilities; evaluation by experts of research grant and fellowship applications submitted to federal and other funding agencies; review by both editors and external referees of articles submitted to scholarly journals; rating of papers and posters submitted to conferences by program committee chairs and members; evaluation of book proposals submitted to university and commercial presses by in-house editors and external readers; and assessments of the quality, applicability, and interpretability of data sets (Lawrence, Jones, Matthews, Pepler, & Callaghan, 2011; Parsons, Duerr, & Minster, 2010). To this list one might add promotion and tenure decisions in higher education for which an individual’s institutional peers and select outside experts determine that person’s suitability for tenure and/or promotion in rank, and also the procedures whereby candidates are admitted to national academies, elected fellows of learned societies, or awarded honors such as the Fields Medal or Nobel Prize.

Received May 10, 2012; revised July 12, 2012; accepted July 13, 2012

© 2012 ASIS&T • Published online 6 December 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.22784

In many ideal depictions, peer review processes are understood as providing “a system of institutionalized vigilance” (Merton, 1973, p. 339) in the self-regulation of knowledge communities. Peer expertise is coordinated to vet the quality and feasibility of submitted work. Authors, in the anticipation of the peer evaluation of their work, aim to conform to shared standards of excellence out of expediency and in accordance with an internalized ethos (Merton, 1973). The norms and values to which peers hold each other are conceived as being universally and consistently applied to all members, where these norms and values pertain to the content of authors’ evidence and arguments independently of their social caste or positional authority (Merton, 1973). When these norms and values are impartially interpreted and applied, peer evaluations are understood as being fair. It is the impartial interpretation and application of shared norms and standards that make for a fair process, which—psychologically (Tyler, 2006) and epistemologically—legitimizes peer review outcomes, content, and institutions.

This is why critics’ charge of bias in peer review is so troubling: Threats to the impartiality of review appear to threaten peer review’s psychological and epistemic legitimacy. Although there are a few exceptions (Lamont, 2009; Lee, in press; Mallard, Lamont, & Guetzkow, 2009), variations in the interpretation and application of epistemic norms and values are almost always conceived of as problematic. Failures in impartiality lead to outcomes that result from the “luck of the reviewer draw” (Cole, Cole, & Simon, 1981, p. 885), fail to uphold the meritocratic image of knowledge communities (Lee & Schunn, 2011; Merton, 1973), protect orthodox theories and approaches (Travis & Collins, 1991), insulate “old boy” networks (Gillespie, Chubin, & Kurzon, 1985; McCullough, 1989), encourage authors to “chase” disputable standards (Ioannidis, 2005, p. 696), and mask bad faith efforts by reviewers who also serve as competitors (Campanario & Acedo, 2005). Perceived partiality leads to dissatisfaction among those whose professional success or failure is determined by review outcomes (Gillespie, Chubin, & Kurzon, 1985; McCullough, 1989; Ware & Monkman, 2008).

The charge of bias also threatens the social legitimacy of peer review. Peer review signals to the body politic that the world of science and scholarship takes seriously its social responsibilities as a self-regulating, normatively driven community. The enormity and complexity of contemporary science and its ramified institutional arrangements are such that peer review has, in the words of Biagioli (2002, p. 34), been “elevated to a ‘principle’ — a unifying principle for a remarkably fragmented field.” As a consequence, the system is held to almost impossibly strict standards and routinely exposed to intense scrutiny by insiders and outsiders alike, including elected politicians (Gustafson, 1975; Walsh, 1975).

Does the “mundane reality of peer review” depart radically from its “mythology” (Biagioli, 2002, p. 13)? Given that human fallibility and venality are inescapable facts of life, it seems unreasonable to imagine that “the flywheel

of science” (Chubin & Hackett, 1990, p. 5) could function flawlessly. Throughout the literature, charges of systematic bias—not just isolated incidents—are repeatedly aired. Such concerns need to be addressed in an open and thoroughgoing fashion to ensure that trust in the integrity of peer review is maintained. In this spirit, our review seeks to articulate notions of impartiality and bias that are faithful to concerns raised by quantitative research on peer review; characterize major genres of research on bias by their methods, assumptions, and concerns; report their results; and indicate how alternative forms of peer review might ameliorate various forms of bias.

Our discussion will draw on literature on the origins, purpose, and mechanics of scientific peer review across multiple genres including journal articles, grant proposals, and fellowship applications (e.g., Bornmann & Daniel, 2007; Bornmann, Mutz, & Daniel, 2007, 2008, 2009; Campanario, 1998a, 1998b; Chubin & Hackett, 1990; Daniel, Mittag, & Bornmann, 2007; Hames, 2007; Holbrook & Frodeman, 2011; Kronick, 1990; Marsh, Bornmann, Mutz, Daniel, & O’Mara, 2009; Shatz, 2004; Spier, 2002); the growing body of empirical and meta-analytic research on the reliability and predictive validity of the peer review process (e.g., Bornmann, 2011a, 2011b; Peters & Ceci, 1982a), including the various kinds of presumptive biases (e.g., institutional, gender, cognitive) associated with different types of review systems (e.g., Alam et al., 2011; Blank, 1991; Budden et al., 2008); survey research on scientists’ attitudes towards peer review (e.g., Sense About Science, 2009; Ware & Monkman, 2008); and debates surrounding the relative merits of open peer review in light of new experimental, web-based systems (e.g., Delamothe & Smith, 2002).

History of Peer Review

The origins of scholarly peer review are commonly associated with the formation of national academies in 17th-century Europe, although some have found foreshadowing of the practice. Biagioli (2002, p. 31) has described in detail “the slow differentiation of peer review from book censorship” and the role state licensing and censorship systems played in 16th-century Europe. A few years after the Royal Society of London (1662) and the Académie Royale des Sciences of Paris (1699) were established, both bodies created in-house journals, the *Philosophical Transactions* and *Journal des Sçavans*, respectively. These prototypical scientific journals gradually replaced the exchange of experimental reports and findings via correspondence, formalizing a process that up until then had been essentially personal, informal, and nonassured in nature. In London, Henry Oldenburg was appointed Secretary to the Royal Society and became the journal’s first editor, gathering, reporting, and editing the work of others (Manten, 1980). From these early efforts gradually emerged the process of independent review of scientific reports by acknowledged experts that persists to this day. Indeed, as early as 1731 the Royal Society of Edinburgh had adopted a review process in

which materials sent to it for publication were vetted and evaluated by knowledgeable members (Spier, 2002, p. 357).

This was the era of the amateur scientist and armchair philosopher who “produced reliable knowledge in and through a moral economy patterned upon the conventions of gentlemanly conversation” (Shapin, 1995, p. 290). But professional science is not conducted by “logically omniscient lone knowers” (Kitcher, 1993, p. 59), and mechanisms thus evolved to formalize the ways in which the trustworthiness of scientific findings could be verified and promulgated to a wider audience. Over time, three principal forms of journal peer review evolved: single-blind, double-blind, and open. Of these, single-blind (the author’s identity is known to the reviewer while the reviewer’s is concealed from the author) is the most widely used, not least because it is a less onerous and less expensive system to operate than double-blind, for which considerable (often unsuccessful) effort is required in order to remove all traces of the author’s identity from all parts of the manuscript/proposal under review (e.g., Blank, 1991; *Nature*, 2008).

Commitment to and Dependence on Peer Review

Today there are literally thousands (estimates vary considerably) of peer-reviewed journals in existence, although the stringency and consistency with which peer review procedures are applied across this population are variable (Mabe, 2003). In any given year these journals publish, at a conservative estimate, a million articles (Björk, Roos, & Lauri, 2009). Each one of those articles will, in all likelihood, have been read by at least one, often two, and sometimes three or more reviewers, selected by journal editors, and most of those submissions will have undergone multiple rounds of review prior to eventual publication in a journal of record. In addition to the million or so published articles there will be at any given moment a very sizeable pool of rejected articles moving through the system, as many (but not all) leading journals have high rejection rates (Schultz, 2010a, 2010b). These rejected papers will also have consumed a great deal of reviewer time (Hamermesh, 1994; Vines, Rieseberg, & Smith, 2010). Moreover, at least some of those rejected papers will be resubmitted to a different journal (possibly more than one) in an effort to see the light of day (Cronin & McKenzie, 1992). As Kravitz and Baker (2011, para. 1) put it: “each submission of a rejected manuscript requires the entire machinery of peer review to creak to life anew,” creating, in effect, “a *journal loop* bounded only by the number of journals available and the dignity of the Authors.”

But that is only part of the story. Research councils, foundations, universities, and other grant-awarding bodies also need to call upon the services of peer experts to review the millions of research proposals, intra- and extramural, seeking funding in any given year. In the U.S. alone, the National Institutes of Health (NIH) and the NSF, the two principal funding agencies, together receive nearly 90,000 research proposals each year and fund less than one quarter

of these (National Institutes of Health, 2011; National Science Foundation, 2011). Many of those who review NSF and NIH research proposals are probably at the very least also regular reviewers of papers for a range of academic journals and conferences and also occasional reviewers of promotion and tenure dossiers, tens of thousands of which require careful scrutiny by multiple reviewers every academic year. It is not hard to grasp the enormity of the burden placed upon members of the scientific community, both junior and senior (Vines, 2010), by a system that, with very few exceptions (see Engers & Gans, 1998), operates on a voluntary, unremunerated basis (Kravitz & Baker, 2011).

With advances in technology, scientific research has become highly sophisticated, collaborative, distributed, and capital intensive in recent years: as a result many manuscripts are now accompanied by large amounts of supplementary materials that require careful scrutiny, placing an even greater burden on conscientious reviewers. As the commercial and career stakes rise, in what Ziman (2000, p. 211) has termed the age of “post-academic science,” so does the burden placed on the shoulders of those individuals refereeing for the world’s leading scientific journals.

The competition for both pecuniary resources and attention in the marketplace of ideas has intensified to such an extent that reviewers need to be ever alert to the possibility of fraud (e.g., data fabrication, data trimming), credit misallocation (e.g., unearned/gift authorship), and potential conflicts of interest (e.g., undeclared commercial or consulting ties) in the publications they evaluate. Although unethical practices have been documented repeatedly in the medical and biomedical fields (e.g., Biagioli, 1998; Cronin, 2002; Sismondo, 2009), there is also suggestive evidence that chicanery and corner-cutting may be on the rise in some of the social sciences (Shea, 2011).

More than ever, we need to rely on peer review in the efficient and effective evaluation of knowledge claims. Research on bias in peer review seeks to identify ways in which it fails to do so. However, as Bornmann (2008) notes, the focal concept of bias has not been defined unambiguously in the literature, perhaps because there is presumed to exist a shared, albeit tacit, understanding of this term. In what follows, we will articulate a general notion of bias, defined as the violation of impartiality in peer evaluation, that draws the empirical literature’s normative concerns together. We will then identify different categories of bias research by their hypothesized source of partiality and (in some cases) by the methods and assumptions adopted to study that type of bias.

Bias in the Peer Review Process

In the context of quantitative research on bias in peer review, reviewer bias is understood as the violation of impartiality in the evaluation of a submission. We define impartiality in peer evaluations as the ability for any reviewer to interpret and apply evaluative criteria in the same way in the assessment of a submission. That is, impartial reviewers

arrive at identical evaluations of a submission in relation to evaluative criteria because they see the relationship of the criteria to the submission in the same ways. And, so long as the evaluative criteria have to do with the cognitive content of the submission and its relationship to the literature, impartiality ensures evaluations are independent of the author's and reviewer's social identities and independent of the reviewer's theoretical biases and tolerance for risk.

There are many reasons to challenge this ideal notion of impartiality in peer review. Lamont (2009) and Mallard et al. (2009) argue that evaluative criteria should not be subject to unifying, transdisciplinary interpretations (Lamont, 2009; Mallard et al., 2009). Lee (in press) argues that impartiality in peer evaluations may not be possible since definitions of evaluative criteria underdetermine their interpretation and application in both multidisciplinary and disciplinary contexts. Lamont (2009, p. 19) argues that the cognitive value of submissions cannot and should not be assessed in ways that are dissociated from the reviewer's "sense of self and relative positioning" with respect to the submission's content. Likewise, it is not clear that a reviewer's theoretical or methodological orientations should be looked upon as normatively problematic. However, we articulate this notion of impartiality in an effort to identify an underlying ideal that aligns different genres of quantitative research on bias in peer review.

These genres of quantitative research can be categorized by differences in their conception of the primary source of bias: (a) error in assessing a submission's "true quality," (b) social characteristics of the author, (c) social characteristics of the reviewer, and/or (d) content of the submission.¹ In what follows we will characterize these genres of work (and their subgenres), identify assumptions and methods adopted to undertake this quantitative research, and provide a selective review of their findings. In all these genres and subgenres, bias is deemed problematic qua partiality. However, when critics implicitly or explicitly express additional grounds for normative complaint, we identify them throughout.

Bias as Deviation From "True Quality" Value

Some quantitative research conceives of bias as a kind of error in identifying "the true quality of the object being rated" (Blackburn & Hakal, 2006, p. 378). Errors in identifying the true quality of submissions violate the ideal of impartiality in peer review by demonstrating that reviewers—in succumbing to error—can fail to interpret and apply evaluative criteria in consistent ways. The assumption that there exists such a value along a single dimension is commonplace within psychometric research, which mea-

¹From a conceptual point of view, social characteristics of editors may also bias peer review. However, we do not expand on this type of bias due to the paucity of empirical work on the topic. Whatever studies are available are mentioned in our discussion of bias as a function of "social characteristics of the reviewer."

sures single-dimension constructs such as intelligence and creativity (Hargens & Herting, 1990; Rust & Golombok, 2009). Improvement in peer review practices, from this perspective, involves improving the reliability with which reviewers identify the true quality value of submissions.

Bias as deviation from proxy measures for true quality. In one subgenre of this research, studies seek to assess the construct validity of peer review as a test/process by comparing its outcomes to proxy measures for manuscript quality. Proxy measures include reviewers' pooled mean rating (Goodman, Berlin, Fletcher, & Fletcher, 1994), ratings by super experts (Gardner & Bond, 1990), editor/panel decisions (Marsh, Jayasinghe, & Bond, 2008; van Rooyen, Godlee, Evans, Smith, & Black, 2010), ratings by readers of a journal (Justice, Cho, Winker, & Berlin, 1998), citation counts (Hagstrom, 1971; Campanario, 1995; Daniel, 2005; Bornmann & Daniel, 2009; Gottfredson, 1978), and subsequent publication (Bornmann & Daniel, 2008b; Bornmann, Mutz, Marx, Schier, & Daniel, 2011). Reliance on proxies for quality is especially common in research on peer review in medicine, which seeks to carry out randomized controlled trials to identify practices that improve peer review processes and outcomes (Godlee, Gale, & Martyn, 1998; Justice et al. 1998; van Rooyen, Godlee, Evans, Smith, et al., 2010).

For example, studies have investigated the citation patterns of "rejected-then-published-elsewhere" articles. A high subsequent citation rate is used to indicate error in the original decision to reject a manuscript (Bornmann & Daniel, 2008b, 2009; Bornmann, Mutz, Marx, et al., 2011). Bornmann and Daniel (2009) found, when comparing citation counts, that 15% of accepted papers and 15% of rejected papers (that were subsequently published elsewhere) should not have been accepted/rejected at a top chemistry journal. Bornmann, Mutz, Marx, et al. (2011) found that acceptance by the original journal was a good predictor of later citation success and that rejection was a good predictor of limited citation, thereby validating editorial decisions.

In some studies, subsequent publication of a rejected manuscript in a more prestigious journal is also used as an indication of error in the original publication decision (Bornmann, Mutz, Marx, et al., 2011). However, Cronin and McKenzie (1992) note relatively few instances of "upward migration" and challenge the notion that such cases may reflect error on the part of the original publication decision: upward migration may sometimes result from the manuscript's finding a better fit—in terms of "focus, scope or style"—with a more prestigious journal (p. 316).

Bias as low inter-rater reliability. From a psychometric perspective, in order for peer review to be a valid test of submission quality, reviewer judgments must be reliable with respect to each other (Hargens & Herting, 1990; Rust & Golombok, 2009, p. 72). Some researchers have suggested

that inter-rater reliability for two reviewers on a single submission should be about 0.8–0.9 (Marsh et al., 2008, p. 162), which is similar to the rate found for intelligence and personality tests (Rust & Golombok, 2009). Unfortunately, agreement between reviewers is very low (e.g., Bornmann & Daniel, 2008a; Ernst & Resch, 1999; Jackson, Srinivasan, Rea, Fletcher, & Kravitz, 2011; Rothwell & Martyn, 2000), with agreement “barely beyond chance” (Kravitz et al., 2010, p. 1) and comparable to rates found for Rorschach inkblot tests (Lee, in press).

Research has demonstrated that inter-reviewer agreement is improved when reviewers evaluate more rather than fewer grant applications, suggesting improvement via learning/training (Jayasinghe, Marsh, & Bond, 2003). Research has also shown improvements in inter-rater reliability with the addition of more reviewers per grant application (Jayasinghe et al., 2003). Such improvements are important to psychometrically oriented researchers since they decrease the chance that review outcomes vary dramatically as a function of which reviewers are chosen (Cole, Cole, & Simon, 1981). However, empirical study suggests that increasing the number of reviewers per journal manuscript does not significantly affect final decisions (Schultz, 2010a).

Inter-rater reliability research focuses on recommendation outcomes without studying other qualities of reviews, such as their length, tone, and presence of references. Without considering the nature and language of the review, it is difficult to assess whether systematic bias is present and what type of bias it may be (epistemic, language, etc.). When “we shift focus away from the numerical representation of a reviewer’s assessment to the content upon which such assessments are grounded, we can identify” (Lee, in press, p. 5) ways in which inter-rater disagreement might reflect normatively appropriate disagreements. Editors and grant program officers may seek reviewers who can evaluate different aspects of a submission according to their own subspecialization and expertise (Bailar, 1991), and we would not expect high inter-rater reliability in cases where quality along these different aspects diverges (Hargens & Herting, 1990). These considerations suggest that “diversity of opinion among referees may be desirable and beneficial” (Chubin & Hackett, 1990, p. 102)—disagreement is thus sometimes normatively desirable and appropriate (Harnad, 1982; Hirschauer, 2009; James, Demaree, & Wolf, 1984; Lee, in press).

Philosophical and qualitative sociological research challenges the psychometric assumption that peer review involves assessments along a single dimension of evaluation. Peer review criteria—such as novelty, soundness, and significance—may be open to different, normatively appropriate interpretations (Lamont, 2009; Lee, in press) and fail to reduce to a single dimension of evaluation. If this is the case, then we would expect normatively appropriate disagreement between reviewers. That normative credibility is conceptually different from high inter-rater reliability is also demonstrated by Bornmann, Mutz, and Daniel (2010), who found that studies with high levels of inter-rater agreement

turned out to be less statistically credible than those with low levels of agreement.

Bias as a Function of Author Characteristics

Among Merton’s classical norms of science is universalism, the ideal that knowledge claims be evaluated according to “preestablished impersonal criteria” that assess the excellence or originality of a person’s ideas, rather than on particular facts about their social identity and status (Merton, 1973, p. 269). As expressed by Peters and Ceci (1982b), universalism in the context of peer review requires that an author’s research be “judged on the merit of [his/her] ideas, not on the basis of academic rank, sex, place of work, publication record, and so on” (p. 252). Social bias is the differential evaluation of an author’s submission as a result of her/his perceived membership in a particular social category. Social bias challenges the thesis of impartiality by suggesting that reviewers do not evaluate submissions—their content and relationship to the literature— independently of the author’s (perceived) identity.

Some view this type of bias as malicious in nature. For example, acknowledging the problem of ad hominem bias, *Nature*’s review policies warn that reviewer anonymity cannot be protected “in the face of a successful legal action to disclose identity in the event of a reviewer having written personally derogatory [review] comments about the authors” (*Nature*, 2012, para. 41). However, bias that violates the norm of universalism need not be ill-intended or conscious at all (Lee & Schunn, 2011). An individual may sincerely espouse norms of equality and invoke normatively appropriate criteria to justify biased evaluations. However, implicit biases in evaluation—resulting from automatic and subconscious processes—are not usually blocked by the conscious, deliberative processes by which egalitarian beliefs are formed and sustained (Bargh & Williams, 2006; Chaiken & Trope, 1999). For example, hiring studies demonstrate that, despite identical curricula vitae, male applicants are deemed as having superior qualities than female applicants (Steinpreis, Anders, & Ritzke, 1999). Ironically, evaluators given the opportunity to disagree with blatantly sexist statements were more likely to reject women for stereotypically male jobs (Monin & Miller, 2001).

Much (but not all) research in this genre assumes that the quality of work by individuals across different social groups (e.g., prestigious vs. not, male vs. female) is, in the aggregate, roughly comparable. As a result, we should expect the rate with which members of less powerful social groups enjoy successful peer review outcomes to be proportionate to their representation in submission rates. Researchers infer the existence of bias when a difference is discovered and infer the lack of bias when no difference is discovered. Very few studies are able to demonstrate that their submission pools are similar to or representative of the larger population of researchers (for an exception, see RAND, 2005). Some models refine comparisons across groups by controlling for

additional factors that might correlate with submission quality. For example, some studies control for factors such as type of institution (Blank, 1991; Xie & Shauman, 1998), experience (RAND, 2005), and rank (Ley & Hamilton, 2008) since these are acknowledged as affecting the resources and expertise needed to do quality work. Studies that distribute for review submissions that are identical in all respects except for the perceived social category to which the stated author belongs control for quality most adequately (Borsuk et al., 2009; Peters & Ceci, 1982b). However, which author attributes should and do correlate with indicators of manuscript quality are questions that deserve further theorizing and testing.

Prestige bias. As Merton observed, prestige-based bias calls attention to a “class structure” in science, where those rich in prestige disproportionately accumulate limited resources (e.g., grant monies, publication space, awards), which allows them to garner yet more prestige in a process of cumulative advantage (Merton, 1973, p. 443; Price, 1976). The preferential evaluation of contributions by the prestigious versus the nonprestigious has been dubbed “the Matthew effect” (Merton, 1968). Some researchers perceive that prestige-bias affects peer review: surveys report that applicants to the NSF and NIH are concerned about “old boy” networks (McCullough, 1989, p. 82; Gillespie et al., 1985, p. 49) and bias against researchers in nonmajor universities (Gillespie et al., 1985, p. 49).

A study of peer-reviewed grant decisions for awarding long-term fellowships to postgraduate researchers in biomedicine discovered that funding rates decreased correlatively with institutional prestige; however, the effect was small and not statistically significant (Bornmann & Daniel, 2006). In a much discussed and highly cited study, Peters and Ceci (1982b) investigated whether “researchers affiliated with prestigious institutions will tend to fare better than colleagues at less prestigious ones” (p. 748). To control for the quality of submissions, they resubmitted published articles by prestigious individuals from prestigious institutions under fictitious names associated with less prestigious institutions. They found that resubmitted manuscripts were rejected 89% of the time (higher than the journal’s 80% rejection rate) on the grounds that the studies contained “serious methodological flaws” (p. 187).

Reviewers do not necessarily use the prestige of an author as direct grounds for their recommendations. For example, Bornmann, Weymuth, and Daniel (2010) investigated the content of reviews of rejected articles to identify which negative comments were the best predictors of future success at subsequent high- and low-ranked journals. They found that reviewers cite relevance and design of research rather than social factors (such as affiliation and institution).

Affiliation bias. Affiliation bias occurs when reviewers and authors/applicants enjoy formal or informal relationships. This bias may be classified as a kind of bias that varies as a

function of reviewer characteristics, since affiliation is shared between authors and reviewers. Affiliation bias may be a form of prestige bias in cases where reviewers and authors enjoy formal or information relationships due to shared, prestige-marked characteristics (e.g., institutional affiliation). Wennerås and Wold (1997) discovered that post-doctoral fellowship applicants with personal ties to reviewers were assessed as more competent than those who were not affiliated but equally productive. A replication of this work found a 15% affiliation bonus for both male and female applicants (Sandström & Hällsten, 2008). However, affiliation does not always result in favorable outcomes for authors and applicants: Oswald (2008) found that two journals housed at top economics departments did not favor or even discriminated against authors from the journal’s parent institution.

Nationality bias. Many studies have found that journals favor authors located in the same country as the journal (e.g., Daniel, 1993; Ernst & Kienbacher, 1991; Link, 1998), some highlighting a particularly strong degree of preferential attachment in the U.S. (Ernst & Kienbacher, 1991). Yet other studies suggest that American authors are more critical of their compatriots and more lenient when assessing grant applications of non-American authors (Marsh et al., 2008). These studies use current author address as a proxy for nationality; however, doing so conflates current affiliation or address with country of origin and ethnicity. Others worry that nationality bias may reflect prose quality and not nationality per se (Cronin, 2009). However, it may also be that language and writing style are cited as problems for manuscripts written by non-native speakers even when there is nothing problematic about the prose (Herrera, 1999).

Language bias. The potential for language bias has been examined both in terms of acceptance rates and as a dependent variable when blinding reviews. An examination of reviews in medical research demonstrated a significant difference in the acceptance rates for abstracts written by authors from English- and non-English-speaking countries (Ross et al., 2006). The difference diminished when the editors instituted blind review: “Blinding significantly attenuated the association between language and likelihood of abstract acceptance” (p. 1679). Tregenza (2002) found that acceptance rates at ecology and evolution journals were higher for first authors living in wealthy English-speaking nations versus wealthy non-English-speaking nations. However, another study found that language was not an important criterion for acceptance or rejection, noting no significant difference in acceptance rates for “linguistically criticized” manuscripts compared with those that did not receive such criticism (Loonen, Hage, & Kon, 2005, p. 1,469).

Gender bias. In light of the gender gap in STEM (science, technology, education, and medicine) fields (Budden et al., 2008; Wennerås & Wold, 1997), the prevailing assumption has been that men are overall more favorably treated than

women in peer review. Although empirical research on gender bias in publication and grant outcomes has produced “data and interpretations which at times are contradictory” (Rees, 2011, p. 140), recent meta-analysis suggests that claims of gender bias in peer review “are no longer valid” (Ceci & Williams, 2011, p. 3,157).

For example, if there is gender bias in review, we would expect double-blind conditions to increase acceptance rates for female authors. However, this is not the case (Blank, 1991). Nor are manuscripts by female authors disproportionately rejected at single-blind review journals such as *Journal of Biogeography* (Whittaker, 2008), *Journal of the American Medical Association* (Gilbert, Williams, & Lundberg, 1994), *Nature Neuroscience* (Nature Neuroscience, 2006), and *Cortex* (Valkonen & Brooks, 2011). Even when the quality of submissions is controlled for, manuscripts authored by women do not appear to be rejected at a higher rate than those authored by men (Borsuk et al., 2009).

Wennerås and Wold (1997) found that female biomedical postdoctoral fellowship applicants had to be 2.5 times more productive than a male applicant to receive the same competence score. However, replications of the study at comparable institutions in the U.K. (Grant, Burden, & Breen, 1997), Canada (Friesen, 1998), and Germany (Bornmann & Daniel, 2006) failed to discover statistically significant gender bias in the awarding of the same type of postdoctoral fellowship. A later replication at the same institution found that gender-based allotments had reversed (Sandström & Hällsten, 2008).

Meta-analyses (Marsh et al., 2009) and large-scale studies (Marsh, Jayasinghe, & Bond, 2011; RAND, 2005) of grant outcomes found no gender differences after adjusting for factors such as discipline, country, institution, experience, and past research output. One study found that female applicants received only 63% of the funding that their male colleagues received from the NIH (RAND, 2005). However, a later study found funding success rates were nearly equal for men and women at NIH when controlling for research rank/stage (Ley & Hamilton, 2008).

Bias as a Function of Reviewer Characteristics

Bias as a function of reviewer characteristics challenges the impartiality of peer review by demonstrating that reviewers fail to evaluate a submission’s content and relationship to the literature independently of reviewer characteristics. Such bias is demonstrated by showing that specific classes of reviewers are systematically tougher or softer on identical submissions (e.g., Jayasinghe et al., 2003) or across multiple submissions (e.g., Gilbert et al., 1994).

Evaluative strictness or leniency can be idiosyncratic to individual reviewers (Casati, Marchese, Ragone, & Turrini, 2009; Marsh & Ball, 1989; Thurner & Hanel, 2010). Strictness or leniency can also vary systematically as a function of the social categories to which reviewers belong. Studies show significant differences in the patterns of reviewing by gender, with female reviewers being stricter than their

male colleagues (e.g., Borsuk et al., 2009; Jayasinghe et al., 2003; Lane & Linden, 2009; Wing, Benner, Petersen, Newcomb, & Scott, 2010). Female editors have been found to reject more submissions than their male colleagues (Gilbert et al., 1994; Lane & Linden, 2009), although the reverse phenomenon has also been discovered (Wing et al., 2010). Toughness may also vary by disciplinary affiliation: Lee and Schunn (2011) found philosophers’ reviews were more negative in tone and more likely to lead to rejection than those written by psychologists. Wood (1997) found American reviewers to be more lenient than their colleagues from the U.K. or Germany. Marsh et al. (2008) found American reviewers to be more lenient than Australians and suggested that the leniency of American reviewers results from “a culture that is comfortable being generous in their evaluations” (p. 163).

This observation raises interesting normative questions about the role that cultural differences play in reviewer style and strictness. Because Marsh et al. (2008) work within a psychometric framework, they conceive of cultural differences in peer evaluations as sources of contamination or error in assessments of a submission’s true quality value. However, when evaluative cultures are specific to disciplines, it is less clear whether such differences should be understood as a form of problematic bias. Lamont (2009) and Mallard et al. (2009) argue that discipline-specific evaluative cultures articulate appropriate ways to approach theory/method and provide the proper epistemic grounds for fairly evaluating grant proposals.

Content-Based Bias

Content-based bias involves partiality for or against a submission by virtue of the content (e.g., methods, theoretical orientation, results) of the work.² Since different types of content-based bias challenge the thesis of impartiality in different ways, we will save analysis of these challenges to discussion of the subtypes. Content-based bias is primarily studied in the context of scientific disciplines. This is because the overarching concern motivating research on content-based bias is whether peer review is capable of the kind of self-regulation that encourages scientific progress and the achievement of other scientific goals. Most studies attempt to demonstrate content-based bias by showing that review outcomes vary as a function of the submission’s content. However, when such studies are not available, surveys or anecdotal evidence from researchers or grant program managers are appealed to instead.

Many hypothesize that reviewers will evaluate more favorably the submissions of authors who belong to similar “schools of thought,” a form of “cognitive cronyism” (Travis & Collins, 1991, p. 323). The perception that cognitive cronyism is at play in peer review contexts is evidenced by

²Content-based bias may also include a form of “ego bias,” where reviewers and editors prefer submissions that cite their own work (e.g., Sugimoto & Cronin, in press).

conversations among grant committee members at the U.K. Science and Engineering Research Council, which reveal attempts to contextualize reviewer recommendations by identifying theoretical and subdisciplinary affiliations between reviewers and proposal authors (Travis & Collins, 1991). Sandström (2009) operationalized cognitive cronyism in reviews by examining the relationships between key noun phrases appearing in the titles and abstracts of papers being reviewed and papers written by the reviewers, hypothesizing that reviewers would favor work that was similar to their own. The data did not support the hypothesis.

At what point does cognitive difference become discrimination? Travis and Collins (1991) contrast cognitive cronyism with bias based on social status. For Travis and Collins, cognitive cronyism is not pernicious like social status bias so long as the boundaries of cognitive communities and social hierarchies do not coincide. However, in cases where they do coincide, outsiders may find “old-boy networks” that control journal and conference content (Hull, 1988, p. 156) and citation networks (Ferber, 1986) difficult to penetrate for social reasons disguised as purely cognitive ones (Lee & Schunn, 2011).

If reviewers prefer research that is similar in cognitive orientation and content to their own, then we would expect that, on the whole, reviewers disfavor research inconsistent with their theoretical orientation as well as research falling outside the mainstream, including interdisciplinary and transformative research.

Confirmation bias. In the psychological literature, confirmation bias is the tendency to gather, interpret, and remember evidence in ways that affirm rather than challenge one’s already held beliefs (Nickerson, 1998). Historical and philosophical analyses have demonstrated the obstructive and constructive role that confirmation bias has played in the course of scientific inquiry, theorizing, and debate (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Solomon, 2001). In the context of peer review, confirmation bias is understood as reviewer bias against manuscripts describing results inconsistent with the theoretical perspective of the reviewer (Jelicic & Merckelbach, 2002). As such, confirmation bias can also be classified as a type of bias that varies as a function of reviewer characteristics. Confirmation bias challenges the impartiality of peer review by questioning whether reviewers evaluate submissions on the basis of their content and relationship to the literature, independently of their own theoretical/methodological preferences and commitments. Confirmation bias also challenges the impartiality of scientists qua scientists by questioning their ability to evaluate scientific hypotheses on the basis of the evidence independently of their “desires, value perspectives, cultural and institutional norms and presuppositions, expedient alliances and their interests” (Lacey, 1999, p. 6).

Empirical study suggests reviewers are vulnerable to confirmation bias. Ernst, Resch, and Uher (1992) found that referees who had published work in favor of a contro-

versial clinical intervention judged a manuscript whose data supported the use of that intervention more favorably than those who had published work against it. Confirmation bias for or against manuscripts may be rooted in biased assessments along more specific dimensions of evaluation. For example, Mahoney (1977) found that reviewers judged the methodological soundness, data presentation, scientific contribution, and publishability of a manuscript to be of higher quality when its data were consistent with the reviewer’s theoretical orientation. However, consistency between a reviewer’s theoretical orientation and a manuscript’s reported results does not automatically lead to confirmation bias. Hull’s (1988) analysis of reviewer recommendations for *Systematic Zoology* demonstrates that, during a time of warring schools of taxonomy, confirmation bias among reviewers was “far from total” (p. 333) since allies can disagree on fundamental tenets and wish to prevent the publication of weak papers that could become easy targets for rivals.

Conservatism. Peer review is often censured for its conservatism, that is, bias against groundbreaking and innovative research (Braben, 2004; Chubin & Hackett, 1990; Wesseley, 1998). Conservatism violates the impartiality of peer review by suggesting that reviewers do not interpret and apply evaluative criteria in identical ways since what count as the proper criteria of evaluation—and their relative weightings—are disputed. Although some challenge the suggestion that conservatism is epistemically problematic (Shatz, 2004), most argue that conservatism threatens scientific progress by stifling the funding and public articulation of alternative and revolutionary scientific theories (Stanford, 2012). More locally, conservatism violates explicit mandates, articulated by journals and granting institutions, to fund and publish innovative research (Frank, 1996; Horrobin, 1990; Luukkonen, 2012).

Many have voiced concern about conservatism in peer review, including past directors at the NSF and NIH (Carter, 1979; Kolata, 2009) and applicants to these institutions (Gillespie et al., 1985, p. 49; McCullough 1989, p. 83). Research suggests that authors proposing unorthodox as opposed to orthodox claims must meet a higher burden of proof: Resch, Ernst, and Garrow (2000) demonstrated that studies supporting unorthodox medical treatments were rated less highly even though the supporting data were equally strong. Qualitative research reveals another possible source for conservatism: for many grant panelists, “frontier” research is understood as “paradigm-shifting” and “revolutionary” (Luukkonen, 2012, p. 54), while “excellent” research is understood as involving “methodological rigour and solid quality of the research” (Luukkonen, 2012, p. 54). Because of the uncertainty surrounding the pursuit of novel methods and theories—and the need for multiple contingency plans should a new experiment or project not go as planned—it may be more difficult for frontier research to appear excellent qua methodologically rigorous or solid.

There is a paucity of quantitative work on whether and where conservatism arises in peer review. This gap indicates a crucial area for future research—one facing methodological and conceptual challenges. Since all manuscripts and grant proposals aim to be novel in some respect, studies on conservatism must find ways to measure degrees of novelty and/or parse out how different types of novelty (e.g., in methods, theory, application context, research question, or statistical analyses) impact peer evaluations.

Bias against interdisciplinary research. Some researchers have expressed concerns of bias against interdisciplinary research since, it is thought, disciplinary reviewers prefer mainstream research (Travis & Collins, 1991). Bias against interdisciplinary research, if discovered, would violate the impartiality of peer review by suggesting that reviewers do not interpret and apply evaluative criteria in identical ways because what count as the proper criteria of evaluation—and their relative weightings—are disputed. Bias against interdisciplinary research would also be problematic since many of the most important social and scientific problems require multiple disciplinary perspectives to address (Metzger & Zare, 1999, p. 642).

Efforts to demonstrate interdisciplinary bias have been mixed. Porter and Rossini (1985) found that interdisciplinary proposals at the NSF received lower ratings. However, no difference in peer rating for interdisciplinary research was found by the Finnish Research Council (Bruun, Hukkinen, Huutoniemi, & Klein, 2005) or by the International Review Committee for Physics (Rinia, van Leeuwen, van Vuren, & van Raan, 2001). The perception of bias remains, however: grant panelists at the European Research Council gave their favorite interdisciplinary projects the highest rating in a strategic effort to counterbalance anticipated bias (Luukkonen, 2012, p. 56). Rather than endorse this kind of gaming behavior by reviewers, the Public Library of Science (PLoS) and the U.K. Research Integrity Office recommend seeking the expertise of a larger number of reviewers, a practice undertaken by the Royal Society (Science and Technology Committee, 2011), to ensure that interdisciplinary work is evaluated by individuals with the appropriate skills and expertise.

Publication bias. Publication bias is the tendency for journals to publish research demonstrating positive rather than negative outcomes, where “positive outcomes” include results that have a positive direction (Bardy, 1998), are statistically significant irrespective of the direction of result (Dickersin, Min, & Meinert, 1992), or both (Fanelli, 2010; Ioannidis, 1998). The controversy surrounding publication bias demonstrates that scientists disagree about the evaluative merits of research reporting negative outcomes (Ioannidis, 2005; Palmer, 2000). More commonly, publication bias is understood as normatively problematic because it leads to exaggerated effect size measurements in later meta-analyses (Ioannidis, 2005; Palmer, 2000), creates publication patterns that conflict with overall disciplinary goals

(Lee, 2012), and encourages the practice of “burying” or “redressing” negatives as positives in distorting ways (Chan, Hróbjartsson, Haahr, Gøtzsche, & Altman, 2004; Gerber & Malhotra, 2008). There is work suggesting that publication bias is the result of reviewer and editor preferences for positive outcomes: for example, the *Journal of the American Medical Association* was more likely to accept statistically significant results on the primary outcome (Olson et al., 2002). However, other work suggests it is authors, anticipating the rejection of negative outcomes, who are primarily responsible for the disproportionate publication of positive outcomes (Dickersin, 1990; Easterbrook, Berlin, Gopalan, & Matthews, 1991) as well as for the increased time lag in the publication of negative results (Ioannidis, 1998).

Forms of Peer Review

Despite concerns about bias, researchers still believe peer review is necessary for the vetting of knowledge claims. One of the most comprehensive surveys of perception of peer review to date found that 93% disagree with the claim that peer review is unnecessary; 85% believe peer review benefits scientific communication; and 83% believe that “without peer review there would be no control” (Ware & Monkman, 2008, p. 1). This suggests that, for researchers, “the most important question with peer review is not whether to abandon it, but how to improve it” (Smith, 2006, p. 180). Many scholars and editorial staff are advocating alternative models of peer review in the hope of accelerating the publication process (reducing “time to market”) and making the review process itself more transparent and less susceptible to bias of different kinds (e.g., Smith, 1999). Both double-blind and open peer review are heavily supported at present. Double-blind review is more commonly found in the humanities and social sciences, and sometimes used in the clinical medical and nursing fields (Ware & Monkman, 2008, p. 8). Single-blind review remains the norm in life sciences, physical science, and engineering (Ware & Monkman, 2008, p. 8). It is widely believed that double-blind review ensures greater fairness for authors while continuing to protect reviewer identities to promote frank commentary. In open peer review authors and reviewers know one another’s identities, reviews may be open to the public, and, in some cases, members of the public can self-select as reviewers. Variations exist on the theme of open review, although most such systems disclose the identity of the reviewer to the public. Hybrid systems also exist that combine elements of open review with public commentary and traditional anonymous peer review.

Double-blind peer review. Single-blind is the most widely used model for peer review; yet in Ware and Monkman’s (2008) survey, 56% of respondents indicated that they would prefer double-blind review (only 25% prefer single-blind): “[d]ouble-blind review was primarily supported because of its perceived objectivity and fairness” (p. 2), a perception that might have an impact on submission behavior by

authors who perceive themselves to be vulnerable to social bias (Budden et al., 2008). Studies in fields where double-blind is the norm have shown high levels of satisfaction (Baggs, Broome, Dougherty, Freda, & Kearney, 2008).

Others have suggested that double-blind review could protect against bias, but noted the difficulty in truly “blinding” a submitted manuscript (Brown, 2007; Science and Technology Committee, 2011). There is far from unanimous support for double-blind review: in a survey of nearly 1,500 editors in chemistry, a plurality of respondents stated that double-blind was “pointless, because content and references give away identity” (Brown, 2007, p. 133). This assumption has been tested in a number of empirical studies, which showed that reviewers can successfully identify authors 25%–40% of the time (Baggs et al., 2008; Ceci & Peters, 1984; Justice et al., 1998; Yankauer, 1991). It has been suggested that these numbers might be significantly higher in more specialized fields (Lane, 2008).

Several studies have attempted to show that double-blind review reduces social bias against authors. Budden et al. (2008) found an increase in female first-authored manuscripts in a single journal following a policy change to double-blind review. However, in a reexamination of this work, Engqvist and Frommen (2008) demonstrated that the data were inconclusive, given the growth in submissions by female authors across similar journals within the field before the policy change; Webb, O’Hara, and Freckleton (2008) demonstrated that alternative statistical techniques suggest no change in publication outcomes; and Whittaker (2008) argued that the number of authors of unknown gender in the original study may have been large enough to cast doubt on the effect.

Other studies have tested for social bias against authors by randomizing double- and single-blind reviewing across manuscripts, but found no significant differences in outcomes (Alam et al., 2011; Blank, 1991; Smith, Nixon, Bueschen, Venable, & Henry, 2002) and no difference in the quality of the resultant reviews (Alam et al., 2011; Justice et al., 1998). One study did demonstrate that more-favorable reviews were written when the reviewers correctly guessed the author of the anonymized manuscripts (Isenberg, Sanchez, & Zafran, 2009). However, given that senior and well-known authors are more likely to be unmasked (Justice et al., 1998), it is unclear whether this is attributable to bias or to demonstrable differences in quality.

Open peer review. Open peer review, that is, a form of review in which authors and reviewers are both known to one another, is seen by proponents as a way to induce transparency in the scholarly communication process and speed up the process of vetting new work. One of the driving motivations behind the move for new forms of peer review is to reduce bias by opening up the “deliberative chambers” (Lamont, 2009, p. 2) to much closer scrutiny. Those who believe that submissions have a true quality value predict that transparency incentivizes better quality reviews and recommendations; however, empirical work suggests this might

not be the case (Godlee et al., 1998; van Rooyen, Godlee, Evans, Black, & Smith, 1999). Those concerned about social- and content-based biases may see increased transparency as a way for reviewers to be accountable and sensitive to their own forms of partiality. In addition, transparency about reviewer identities would allow authors and members of the community to contextualize review content with knowledge about the particular theoretical, methodological, disciplinary, and cultural perspective from which the review is written. This contextualized knowledge would enable the community to use differences along these axes as a resource in the articulation of each other’s background assumptions and the development of more refined critiques and responses—a process of transformative criticism that uses the diversity of individuals’ content-based partiality to improve the objectivity of the community as a whole (Longino, 1990, 2002).

Despite the potential advantages of open peer review, researchers and scholars seem somewhat reticent to adopt it. In Ware and Monkman’s (2008) survey, only 13% preferred open review to other models and only 27% thought it could be an effective form of review compared with 17% in the Melero and Lopez-Santovenia (2001) study. Nearly half of all Ware and Monkman’s (2008) respondents said that open peer review would make them less likely to review. Other studies have noted that disclosing the reviewer’s name would act as a disincentive and lead to a decline in the potential pool of willing reviewers (Baggs et al., 2008; van Rooyen, Delamothe, & Evans, 2010). Some scholars note that reviewer anonymity protects the social cohesion of research groups by allowing same-group reviewers to “play down their areas of disagreement” in public (Hull, 1988, p. 334). More generally, scholars feel that “anonymity protects younger, less powerful reviewers from possible retribution on the part of the rejected author” (Peters & Ceci, 1982b, p. 251). Of course, these statistics may be a reflection of generational differences, specifically (to some degree at least) the attachment established scholars have to the “legacy system” (Kravitz & Baker, 2011, para. 1).

Controlled studies have found that open review is associated with a higher refusal rate (on the part of reviewers) and an increase in the amount of time taken to write reviews (van Rooyen, Delamothe, et al., 2010; Walsh, Rooney, Appleby, & Wilkinson, 2000). Studies are inconclusive on the effect on quality, with some finding no difference (van Rooyen, Delamothe, et al., 2010; van Rooyen et al., 1999) and others an increase in quality (Walsh et al., 2000). It has been suggested that open peer review may increase levels of inter-rater reliability; however, recent studies have found no difference in agreement levels between open and closed peer review (e.g., Bornmann & Daniel, 2010b).

Despite the lack of empirical evidence in favor of open reviewing, many journals are implementing, or experimenting with, the approach. *Shakespeare Quarterly* tried it successfully (Cohen, 2010); it has been used by *BMJ* for more than a decade (Science and Technology Committee, 2011; Ware & Monkman, 2008); MIT Press, *PLoS Medicine*, and

Nature have had less successful attempts due to lack of engagement by authors and reviewers (*Nature*, 2006; Science and Technology Committee, 2011; Timmer, 2008; Ware & Monkman, 2008).

There are, however, multiple ways in which open review can be conducted. In the cases just cited, the editor instigated and oversaw the review process. While author-suggested reviewers are also employed by a few journals, these reviewers tend to be more favorable than editor-nominated reviewers (Bornmann & Daniel, 2010a). A more radical approach involves “crowdsourcing”: let the readers themselves decide what should be admitted into the scholarly literature. Such an approach has been used by the book publishing company *unbound*, which allows authors to post their work and garner support from readers to have it published commercially (<http://unbound.co.uk/>). In the scholarly realm, it has been proposed that arXiv.org, a popular preprint repository, use open peer review to evaluate and certify the preprints it houses, the idea being that editors could invite referees to write and sign reviews of preprints, the result of which would be the ability to label certain articles as published articles, rather than preprints, thereby, ultimately, bypassing the world of commercial scholarly publishing (Boldt, 2011).

Hybrid peer review. Hybrid systems, as noted above, combine elements of openness (often in the form of public commentary) with traditional (i.e., blind) review. Discussions of hybrid systems are not new: More than 30 years ago Harnad (1979, p. 18) argued presciently for a “complementary mechanism . . . in which scientists could solicit open peer commentary on their work.” He claimed that this “would not only provide a medium and an incentive for efforts and initiatives that might otherwise have been buried in the evanescence and anonymity of closed peer review, but it would also focus and preserve the creative disagreement, in a disciplined form, answerable to print and posterity” (Harnad, 1979, p. 19). The ubiquity of online computing has made this possibility much more of a reality.

Implicit in Harnad’s proposal for hybrid review is the idea that it serves as a supplement to, not a replacement for, traditional closed peer review. Many hybrid options have been developed based on the same assumption. Harnad (1982) classifies these as either a priori systems, where comments are invited before formal peer review, or a posteriori systems, where comments are invited after the work has been reviewed.

A priori peer review. The journal *Atmospheric Chemistry and Physics* experimented with an a priori open review system in which articles were posted to the site (after initial editorial screening) and select reviewers as well as the public were invited to comment over an 8-week period. The author then prepared a final draft, which was reviewed and assessed for suitability for publication by the editor (Science and Technology Committee, 2011). Other journals have also explored ways of providing a platform for discussion preceding formal peer review (e.g., Copernicus Publications).

In the case of *Atmospheric Chemistry and Physics*, the original draft, the comments, and the final version are all made accessible online. Some journals have sought to capture and disseminate the “prepublication history” and anonymized reviews along with the final version of the paper (e.g., *EMBO Journal*, BioMed Central medical journals, *Electronic Transactions on Artificial Intelligence*, *Hydrology and Earth Systems Science Discussions*, and *BMC Medicine* [Science and Technology Committee, 2011]).

The *Journal of Medical Internet Research* has taken a slightly different route, providing a list of articles that are awaiting peer review. A reviewer can choose to sign up and review the article. If the article is accepted, the name of the reviewer will accompany the article. If it is rejected, the reviewer will remain anonymous. The goal of the procedure is to “give constructive feedback to the authors and/or to prevent publication of uninteresting or fatally flawed articles” (*Journal of Medical Internet Research*, 2012). There have been other experiments in publishing reviews alongside the reviewed manuscript. One such system works as a social networking site for reviewing—allowing authors to submit work that is distributed anonymously to registered reviewers on the site. The reviews are then “scored” by the authors, providing an evaluation/ranking mechanism for reviewers. If reviewers agree, their signed reviews are published in an online journal (de Vrieze, 2012).

A posteriori peer review. Postpublication comments are seen by some as “a useful supplement to formal peer review, rather than a replacement for it” (Ware & Monkman, 2008, p. 2). Kravitz and Baker (2011, para. 1) have outlined a radical model that combines prepublication (anonymous) peer review with “the development of a marketplace where the priority of a paper rises and falls based on its reception from the field.” Systems where stable documents are provided and readers are allowed to post comments are the most common form of postpublication review (e.g., PLoS journals: see <http://www.plos.org/>), but other novel implementations have been tried; for example, *Open Medicine*, which posts reviewed articles on a wiki, allows real-time changes by readers. This is probably the most radical of implementations, as it allows collaborative authoring of the peer-reviewed work, rather than mere comments on it. Other proposals have included “rebound” systems, which allow the author to engage in postreview debate over a rejected manuscript (Sen, 2012). Post-peer-review commentary on, and rating of, published papers in thousands of journals by domain experts across a range of fields is the approach used by Faculty of 1000 (see <http://f1000.com/>). These various types of open review have been referred to as “fear review” as it is suggested that this exposure of, and the ability for peers to publicly comment on, one’s published work may “well be more nerve-racking than having it read discretely by only two or three peers” (Cronin, 2003, p. 15).

Although a number of proposals have been suggested, very few studies have investigated how these various approaches might actually mitigate bias or how bias might

be quantified with precision. Despite the many calls for an overhaul of the status quo, there is little evidence that alternative forms of review either reduce bias or have been enthusiastically embraced by the scientific community at large. In fact, studies have found no significant difference in perceived bias or review quality in these systems but, rather, a refusal on the part of reviewers to participate and protracted turnaround times. Given that the most commonly voiced criticism of the current system is the time taken to produce reviews, as noted by 38% of survey respondents (Ware & Monkman, 2008)—and the finding that there is “no crisis in supply of peer reviewers” (Vines et al., 2010, p. 1,041)—a wholesale shift seems unlikely in the near term.

Conclusions and Future Research

Impartiality ensures both the consistency and meritocracy of peer review. Research on bias in peer review—predicated on the ideal of impartiality—raises not just local hypotheses about specific sources of partiality, but much broader questions about whether the processes by which knowledge communities regulate themselves are epistemically and socially corrupt. Contra impartiality, the evidence suggests that peer evaluations vary as a function of author nationality and prestige of institutional affiliation; reviewer nationality, gender, and discipline; author affiliation with reviewers; reviewer agreement with submission hypotheses (confirmation bias); and submission demonstration of positive outcomes (publication bias).

However, a closer look at the empirical and methodological limitations of research on bias raises questions about the existence, extent, and normative status of many hypothesized forms of bias. Psychometrically oriented research is predicated on the questionable assumption that disagreement among reviewers is not normatively appropriate or desirable. Research on bias as a function of author characteristics adopts the untested assumption that authors belonging to different social categories submit manuscripts and grant proposals of comparable quality. Despite vocal concerns about conservatism in science, there is no empirical evidence (beyond anecdote) to buttress or belie such worries. And the evidence for bias against interdisciplinary research is mixed, as is the evidence for bias against female authors and authors living in non-English-speaking countries.

Research on bias in peer review also suggests that peer review is social in ways that go beyond the social categories to which authors and reviewers belong: Relationships between individuals in the process impact outcomes (e.g., affiliation bias), and individuals make decisions conditioned on beliefs about what others value (e.g., publication bias). Future research might usefully investigate these complex and dynamic social relations. Consider, for example, how the editor’s relationships and beliefs about other actors may have an impact on his/her decisions. On the basis of previous experience with reviewers, the editor may differentially value and preferentially assign reviewers to manuscripts,

which may alter final recommendations. Frequent or highly sought authors to the journal may develop a privileged relationship with the editor and with potential reviewers. Editors may feel peer pressure when evaluating manuscripts submitted by frequent reviewers and editorial board members (Lipworth, Kerridge, Carter, & Little, 2011). The readership may function as an invisible hand in the selection of authors and manuscript content, since the editor will need to be cognizant of the needs and wants of the marketplace. An editor may also be influenced by her/his relationship with the editorial board and/or publisher (commercial, academic, or society). The editor’s strategy or vision for the journal may have a bearing on which manuscripts are reviewed and ultimately accepted for publication. As Chubin and Hackett (1990, p. 92) note, “[t]he journal editor occupies a delicate position between the author and reviewers, alternating among the roles of wordsmith and gatekeeper, caretaker and networker, literary agent and judge.”

Not all of these sources of social influence impact peer review in problematic ways. For example, the ways in which authors, reviewers, and editors anticipate each others’ scrutiny and judgment may serve to improve the quality of each of their contributions (Bailar, 1991; Hirschauer, 2009), and editors’ personal connections allow them to learn about and capture high-impact papers for publication (Laband & Piette, 1994). These examples suggest that the sociality of peer review can be structured “to enrich, rather than threaten” the well-being of peer review (Lipworth et al., 2011, p. 1,056). A natural direction for future research includes articulating and assessing alternative normative models that acknowledge reviewer partiality, with a focus on the epistemic and cultural bases for reviewer disagreement; the ways editors and grant program managers anticipate, capitalize on, and manage reviewer disagreement; and the ways publication venues and funding opportunities should be structured to accommodate reviewer differences (Hargens & Herting, 1990; Lee, 2012). Finally, the inescapable sociality and partiality of peer evaluation raise questions about whether impartiality can or should be upheld as the ideal for peer review.

Acknowledgments

The authors acknowledge the contributions of the three reviewers who provided thoughtful commentary on an earlier version of this article. Blaise Cronin, being Editor-in-Chief of *JASIST*, was uninvolved in the review process, which was handled entirely by Jonathan Furner, an associate editor of the Journal and editor for the Advances in Information Science section.

References

- Alam, M., Kim, N.A., Havey, J., Rademaker, A., Ratner, D., Tregre, B., . . . Coleman, W.P. 3rd. (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.

- Baggs, H.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.
- Bailar, J.C. (1991). Reliability, fairness, objectivity and other inappropriate goals in peer review. *Behavioral and Brain Sciences*, 14, 137–138.
- Bardy, A.H. (1998). Bias in reporting clinical trials. *British Journal of Clinical Pharmacology*, 46, 147–150.
- Bargh, J.A., & Williams, E.L. (2006). The automaticity of social life. *Current Directions in Psychological Science*, 15(1), 1–4.
- Biagioli, M. (1998). The instability of authorship: Credit and responsibility in contemporary medicine. *The FASB Journal*, 12(1), 3–16.
- Biagioli, M. (2002). From book censorship to academic peer review. *Emergences*, 12(1), 11–45.
- Björk, B.-C., Roos, A., & Lauri, M. (2009). Scientific journal publishing: Yearly volume and open access availability. *Information Research*, 14(1). Retrieved from <http://informationr.net/ir/14-1/paper391.html>
- Blackburn, J.L., & Hakel, M.D. (2006). An examination of sources of peer-review bias. *Psychological Science*, 17(5), 378–382.
- Blank, R.M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review. *American Economic Review*, 81(5), 1041–1068.
- Boldt, A. (2011). Extending ArXiv.org to achieve open peer review. *Journal of Scholarly Publishing*, 42(2), 238–242.
- Bornmann, L. (2008). Scientific peer review: An analysis of the peer review process from the perspective of sociology of science theories. *Human Architecture — Journal of the Sociology of Self-Knowledge*, 6(2), 23–38.
- Bornmann, L. (2011a). Peer review and bibliometrics: Potentials and problems. In J.C. Shin, R.K. Toutkoushian, & U. Teichler (Eds.), *University rankings: Theoretical basis, methodology and impacts on global higher education* (pp. 145–164). Berlin, Germany: Springer.
- Bornmann, L. (2011b). Scientific peer review. In B. Cronin (Ed.), *Annual review of information science and technology*, 45 (pp. 199–245). Medford, NJ: Information Today/American Society for Information Science & Technology.
- Bornmann, L., & Daniel, H.-D. (2006). Potential sources of bias in research fellowship assessments. Effects of university prestige and field of study on approval and rejection of fellowship applications. *Research Evaluation*, 15(3), 209–219.
- Bornmann, L., & Daniel, H.-D. (2007). Gatekeepers of science: Effects of external reviewers' attributes on the assessments of fellowship applications. *Journal of Informetrics*, 1(1), 83–91.
- Bornmann, L., & Daniel, H.-D. (2008a). The effectiveness of the peer review process: Inter-referee agreement and predictive validity of manuscript refereeing at *Angewandte Chemie International Edition*, 47(38), 7173–7178.
- Bornmann, L., & Daniel, H.-D. (2008b). Selecting manuscripts for a high impact journal through peer review: A citation analysis of communications that were accepted by *Angewandte Chemie International Edition*, or rejected but published elsewhere. *Journal of the American Society for Information Science and Technology*, 59(11), 1841–1852.
- Bornmann, L., & Daniel, H.-D. (2009). Extent of type I and type II errors in editorial decisions: A case study on *Angewandte Chemie International Edition*. *Journal of Informetrics*, 3(4), 348–352.
- Bornmann, L., & Daniel, H.-D. (2010a). Do author-suggested reviewers rate submissions more favorably than editor-suggested reviewers? A study on *Atmospheric Chemistry and Physics*. *PLoS ONE*, 5(10), e13345.
- Bornmann, L., & Daniel, H.-D. (2010b). Reliability of reviewers' ratings when using public peer review: A case study. *Learned Publishing*, 23, 124–131.
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2007). Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, 1, 226–238.
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). How to detect indications of potential sources of bias in peer review: A generalized latent variable modeling approach exemplified by a gender study. *Journal of Informetrics*, 2, 280–287.
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2009). The influence of the applicants' gender on the modeling of a peer review process by using latent Markov models. *Scientometrics*, 81(2), 407–411.
- Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PLoS One*, 5(12), e14331.
- Bornmann, L., Mutz, R., Marx, W., Schier, H., & Daniel, H.-D. (2011). A multilevel modeling approach to investigating the predictive validity of editorial decisions: Do the editors of a high profile journal select manuscripts that are highly cited after publication? *Journal of the Royal Statistical Society*, 174, 857–879.
- Bornmann, L., Weymuth, C., & Daniel, H.-D. (2010). A content analysis of referees' comments: How do comments on manuscripts rejected by a high-impact journal and later published in either a low- or high-impact journal differ? *Scientometrics*, 83(2), 493–506.
- Borsuk, R.M., Aarssen, L.W., Budden, A.E., Koricheva, J., Leimu, R., Tregenza, T., Lortie, C.J., (2009). To name or not to name: The effect of changing author gender on peer review. *Bioscience*, 59(11), 985–989.
- Braben, D.W. (2004). *Pioneering research: A risk worth taking*. Hoboken, NJ: Wiley-Interscience.
- Brown, R.J.C. (2007). Double anonymity in peer review within the chemistry periodicals community. *Learned Publishing*, 20(2), 131–137.
- Bruun, H., Hukkinen, J., Huutoniemi, K., & Klein, J.T. (2005). Promoting interdisciplinary research: The case of the Academy of Finland. Helsinki: Academy of Finland.
- Budden, A.E., Tregenza, T., Aarssen, L.W., Koricheva, J., Leimu, R., . . . (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution*, 23, 4–6.
- Campanario, J.M. (1995). Commentary: On influential books and journal articles initially rejected because of negative referees' evaluations. *Science Communication*, 16, 304–325.
- Campanario, J.M. (1998a). Peer review for journals as it stands today — Part I. *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.
- Campanario, J.M., & Acedo, E. (2005). Rejecting highly cited papers: The views of scientists who encounter resistance to their discoveries from other scientists. *Journal of the American Society for Information Science and Technology*, 58(5), 734–743.
- Carter, L.J. (1979). A new and searching look at NSF. *Science*, 204, 1064–1065.
- Casati, F., Marchese, M., Ragone, A., & Turrini, M. (2009). Is peer review any good? A quantitative analysis of peer review (Technical Report No. DISI-09-045). Retrieved from <http://eprints.biblio.unitn.it/1654>
- Ceci, S.J., & Peters, D. (1984). How blind is blind review? *American Psychologist*, 39(12), 1491–1494.
- Ceci, S.J., & Williams, W.M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8), 3157–3162.
- Chaiken, S. & Trope, T. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Chan, A.-W., Hróbjartsson, A., Haahr, M.T., Gøtzsche, P.C., & Altman, D.G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *Journal of the American Medical Association*, 291, 2457–2465.
- Chubin, D.E., & Hackett, E.J. (1990). *Peerless science: Peer review and U.S. science policy*. Stony Brook, NY: State University of New York Press.
- Cohen, P. (2010). Scholars test web alternative to peer review. *The New York Times*. Retrieved from <http://www.nytimes.com/2010/08/24/arts/24peer.html?pagewanted=all>
- Cole, S., Cole, J.R., & Simon, G. (1981). Chance and consensus in peer review. *Science*, 214, 881–886.
- Cronin, B. (2002). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, 52(7), 558–569.

- Cronin, B. (2003). Scholarly communication and epistemic cultures. *New Review of Academic Librarianship*, 9, 1–24.
- Cronin, B. (2009). Editorial. Vernacular and vehicular language. *Journal of the American Society for Information Science and Technology*, 60(3), 433.
- Cronin, B., & McKenzie, G. (1992). The trajectory of rejection. *Journal of Documentation*, 48(3), 310–317.
- Daniel, H.-D. (1993). Guardians of science: Fairness and reliability of peer review. Weinheim, Germany: Wiley-VCH.
- Daniel, H.-D. (2005). Publications as a measure of scientific achievement and scientists' productivity. *Learned Publishing*, 18, 143–148.
- Daniel, H.-D., Mittag, S., & Bornmann, L. (2007). The potential and problems of peer evaluation in higher education and research. In A. Cavalli (Ed.), *Quality assessment for higher education in Europe*. London: Portland Press, 71–82.
- Delamothe, T., & Smith, R. (2002). Twenty thousand conversations: rapid responses suggest new models of knowledge creation. *British Medical Journal*, 324. Retrieved from <http://www.bmj.com/content/324/7347/1171.full>
- Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association*, 263, 1385–1389.
- Dickersin, K., Min, Y.-I., & Meinert, C.L. (1992). Factors influencing publication of research results: Follow-up of applications submitted to two institutional review boards. *Journal of the American Medical Association*, 267, 374–379.
- de Vrieze, J. (2012). Online social network seeks to overhaul peer review in scientific publishing. *ScienceInsider*. Retrieved from <http://news.sciencemag.org/scienceinsider/2012/01/online-social-network-seeks-to.html>
- Easterbrook, P.J., Berlin, J.A., Gopalan, R., & Matthews, D.R. (1991). Publication bias in clinical research. *Lancet*, 337(8746), 867–872.
- Engers, M., & Gans, J.S. (1998). Why referees are not paid (enough). *American Economic Review*, 88(5), 1341–1350.
- Engqvist, L., & Frommen, J.G. (2008). Double-blind peer review and gender publication bias. *Animal Behaviour*, 76, E1–E2.
- Ernst, E., & Kienbacher, T. (1991). Chauvinism. *Nature*, 352, 560.
- Ernst, E., & Resch, K.L. (1999). Reviewer bias against the unconventional? A randomized double-blind study of peer review. *Complementary Therapies in Medicine*, 7(1), 19–23.
- Ernst, E., Resch, K.L., Uher, E.M. (1992). Reviewer bias. *Annals of Internal Medicine*, 116(11), 958.
- Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States data. *PLoS ONE*, 5(4), e10271.
- Ferber, M.A. (1986). Citations: Are they an objective measure of scholarly merit? *Signs*, 11(2), 381–389.
- Frank, E. (1996). Editors' requests of reviewers: A study and a proposal. *Preventative Medicine*, 25, 102–104.
- Friesen, H.G. (1998). Equal opportunities in Canada. *Nature* 391, 326.
- Gardner, M.J., & Bond, J. (1990). An exploratory study of statistical assessment of papers published in the *British Medical Journal*. *Journal of the American Medical Association*, 263, 1355–1357.
- Gerber, A.S., & Malhotra, N. (2008). Publication bias in empirical sociological research — Do arbitrary significance levels distort published results? *Sociological Methods & Research*, 37(1), 3–30.
- Gilbert, J.R., Williams, E.S., & Lundberg, G.D. (1994). Is there gender bias in JAMA's peer-review process? *Journal of the American Medical Association*, 272(2), 139–142.
- Gillespie, G.W., Chubin, D.E., & Kurzon, G.M. (1985). Experience with NIH peer review: Researchers' cynicism and desire for change. *Science, Technology, & Human Values*, 10(3), 44–54.
- 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, 237–240.
- Goodman, S.N., Berlin, J.A., Fletcher, S.W., & Fletcher, R.H. (1994). Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Annals of Internal Medicine*, 121, 11–21.
- Gottfredson, S.D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist*, 33, 920–934.
- Grant, J., Burden, S., & Breen, G. (1997). No evidence of sexism in peer review. *Nature*, 390, 438.
- Greenwald, A.G., Pratkanis, A.R., Leippe, M.R., & Baumgardner, M.H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, 93(2), 216–229.
- Gustafson, T. (1975). The controversy over peer review. *Science*, 190(4219), 1060–1066.
- Hagstrom, W.O. (1971). Inputs, outputs, and the prestige of university science departments. *Sociology of Education*, 44, 375–397.
- Hamermesh, D.S. (1994). Facts and myths about refereeing. *Journal of Economic Perspectives*, 8(1), 153–164.
- Hames, I. (2007). *Peer review and manuscript management in scientific journals*. Oxford, UK: Blackwell.
- Hargens, L.L., & Herting, J.R. (1990). Neglected considerations in the analysis of agreement among journal referees. *Scientometrics*, 19, 91–106.
- Harnad, S. (1979). Creative disagreement. *The Sciences*, 19, 18–20.
- Harnad, S. (Ed.). (1982). *Peer commentary on peer review: A case study in scientific quality control*. Cambridge, UK: Cambridge University Press.
- Herrera, A.J. (1999). Language bias discredits the peer-review system. *Nature*, 397(6719), 467.
- Hirschauer, S. (2009). Editorial judgments: A praxeology of “voting” in peer review author(s). *Social Studies of Science*, 40(1), 71–103.
- Holbrook, J.B., & Frodeman, R. (2011). Peer review and the ex ante assessment of societal impacts. *Research Evaluation*, 20(3), 239–246.
- Horrobin, D.F. (1990). The philosophical basis of peer review and the suppression of innovation. *Journal of the American Medical Association*, 263, 1438–1441.
- Hull, D. (1988). *Science as a process*. Chicago: University of Chicago Press.
- Ioannidis, J.P.A. (1998). Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *Journal of the American Medical Association*, 279, 281–286.
- Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124.
- 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.
- Jackson, J.L., Srinivasan, M., Rea, J., Fletcher, K.E., & Kravitz, R.L. (2011). The validity of peer review in a general medicine journal. *PLoS ONE*, 6(7), e22475.
- James, L.R., Demaree, R.G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69(1), 85–98.
- Jayasinghe, U.W., Marsh, H.W., & Bond, N. (2003). A multilevel cross-classified modelling approach to peer review of grant proposals: The effects of assessor and researcher attributes on assessor ratings. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 166(3), 279–300.
- Jelicic, M., & Merckelbach, H. (2002). Peer-review: Let's imitate the lawyers! *Cortex*, 38, 406–407.
- Journal of Medical Internet Research. (2012). Open peer review articles. Retrieved from <http://www.jmir.org/reviewer/openReview/abstracts>
- Justice, A.C., Cho, M.K., Winker, M.A., & Berlin, J.A. (1998). Does masking author identity improve peer review quality? A randomized controlled trial. *Journal of the American Medical Association*, 280(3), 240–242.
- Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusion*. New York: Oxford University Press.
- Kolata, G. (2009, June 27). System leads cancer researchers to play it safe. *New York Times*. Retrieved from <http://www.nytimes.com/2009/06/28/health/research/28cancer.html>
- Kravitz, D.J., & Baker, C.I. (2011). Toward a new model of scientific publishing: discussion and a proposal. *Frontiers in Computational*

- Neuroscience, 5(55). Retrieved from http://www.frontiersin.org/computational_neuroscience/10.3389/fncom.2011.00055/full
- Kravitz, R.L., Franks, P., Feldman, M.D., Gerrity, M., Byrne, C., & Tierney, W.M. (2010). Editorial peer reviewers' recommendations at a general medical journal: Are they reliable and do editors care? *PLoS ONE*, 5(4), e10072.
- Kronick, D.A. (1990). Peer review in 18th-century scientific journalism. *Journal of the American Medical Association*, 263, 1321–1322.
- Laband, D.N., & Piette, M.J. (1994). Favoritism versus search for good papers: Empirical evidence regarding the behavior of journal editors. *Journal of Political Economy*, 102, 194–203.
- Lacey, H. (1999). *Is science value free? Values and scientific understanding*. New York: Routledge.
- Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Cambridge, MA: Harvard University Press.
- Lane, D. (2008). Double-blind review: Easy to guess in specialist fields. *Nature*, 452, 28.
- Lane, J.A., & Linden, D.J. (2009). Is there gender bias in the peer review process at *Journal of Neurophysiology*? *Journal of Neurophysiology*, 101(5), 2195–2196.
- Lawrence, B., Jones, C., Matthews, B., Pepler, S., & Callaghan, S. (2011). Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation*, 6(2), 4–37.
- Lee, C.J. (in press). A Kuhnian critique of psychometric research on peer review. *Philosophy of Science*.
- Lee, C.J. (2012). Incentivizing procedural objectivity: Community response to publication bias. Unpublished manuscript.
- Lee, C.J., & Schunn, C.D. (2011). Social biases and solutions for procedural objectivity. *Hypatia: A Journal of Feminist Philosophy*, 26(2), 352–373.
- Ley, T.J., & Hamilton, B.H. (2008). The gender gap in NIH grant applications. *Science*, 322, 1472–1474.
- Link, A.M. (1998). US and non-US submissions. *Journal of the American Medical Association*, 280(3), 246–247.
- Lipworth, W.L., Kerridge, I.H., Carter, S.M., & Little, M. (2011). Journal peer review in context: A qualitative study of the social and subjective dimensions of manuscript review in biomedical publishing. *Social Science & Medicine*, 72, 1056–1063.
- Longino, H. (1990). *Science as social knowledge*. Princeton, NJ: Princeton University Press.
- Longino, H. (2002). *The fate of knowledge*. Princeton, NJ: Princeton University Press.
- Loonen, M.P.J., Hage, J.J., & Kon, M. (2005). Who benefits from peer review? An analysis of the outcome of 100 requests for review by Plastic and Reconstructive Surgery. *Plastic and Reconstructive Surgery*, 116(5), 1461–1472.
- Luukkonen, T. (2012). Conservatism and risk-taking in peer review: Emerging ERC practices. *Research Evaluation*, 21, 48–60.
- Mabe, M. (2003). The growth and number of journals. *Serials*, 16(2), 191–197.
- Mahoney, M.J. (1977). Publication preferences: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2), 161–175.
- Mallard, G., Lamont, M., & Guetzkow, J. (2009). Fairness as appropriateness: Negotiating epistemological differences in peer review. *Science, Technology, & Human Values*, 34(5), 573–606.
- Manten, A.A. (1980). The growth of European scientific journal publishing before 1850. In A.J. Meadows (Ed.), *Development of scientific publishing in Europe* (pp. 1–22). New York: Elsevier.
- Marsh, H.W., & Ball, S. (1989). The peer review process used to evaluate manuscripts submitted to academic journals: Interjudgmental reliability. *Journal of Experimental Education*, 57, 151–169.
- Marsh, H.W., Jayasinghe, U.W., & Bond, N.W. (2008). Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist*, 63(3), 160–168.
- Marsh, H.W., Jayasinghe, U.W., & Bond, N.W. (2011). Gender differences in peer reviews of grant applications: A substantive-methodological synergy in support of the null hypothesis model. *Journal of Informetrics*, 5, 167–180.
- Marsh, H.W., Bornmann, L., Mutz, R., Daniel, H.-D., & O'Mara, A. (2009). Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Educational Research*, 79, 1290–1326.
- McCullough, J. (1989). First comprehensive survey of NSF applicants focuses on their concerns about proposal review. *Science, Technology, & Human Values*, 14(1), 78–88.
- Melero, R., & Lopez-Santovenia, F. (2001). Referees' attitudes toward open peer review and electronic transmission of papers. *Food Science and Technology International*, 7(6), 521–527.
- Merton, R.K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63.
- Merton, R.K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.
- Metzger, N., & Zare, R.N. (1999). Interdisciplinary research: From belief to reality. *Science*, 283(5402), 642–643.
- Monin, B., & Miller, D.T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33–43.
- National Institutes of Health. (2011). NIH databook. Retrieved from <http://report.nih.gov/nihdatabook/>
- National Science Foundation. (2011). Proposal and award policies and procedures guide. Retrieved from <http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/index.jsp>
- Nature. (2006). Overview: Nature's peer review trial. Retrieved from <http://www.nature.com/nature/peerreview/debate/nature05535.html>
- Nature. (2008). Editorial. Working double-blind. Should there be author anonymity in peer review? Retrieved from <http://www.nature.com/nature/journal/v451/n7179/full/451605b.html>
- Nature. (2012). Peer-review policy. Retrieved from http://www.nature.com/authors/policies/peer_review.html
- Nature Neuroscience. (2006). Women in neuroscience: A numbers game. *Nature Neuroscience*, 9(7), 853.
- Nickerson, R.S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- Olson, C.M., Rennie, D., Cook, D., Kickersin, K., Flanagan, A., Hogan, J.W., . . . Pace, B. (2002). Publication bias in editorial decision making. *Journal of the American Medical Association*, 287, 2825–2828.
- Oswald, A.J. (2008). Can we test for bias in scientific peer-review. IZA discussion paper 3665. Bonn, Germany: Institute for the Study of Labor.
- Palmer, A.R. (2000). Quasi-replication and the contract of error: Lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annual Review of Ecology and Systematics*, 31, 441–480.
- Parsons, M.A., Duerr, R., & Minster, J.-B. (2010). Data citation and peer-review. *Eos, Transactions, American Geophysical Union*, 91(34), 297–298.
- Peters, D.P., & Ceci, S.J. (1982a). Peer-review practices of psychological journals: The fate of accepted, published articles, submitted again. *Behavioral and Brain Sciences*, 5(2), 187–195.
- Peters, D.P., & Ceci, S.J. (1982b). Peer-review research: Objections and obligations. *Behavioral and Brain Sciences*, 5(2), 246–252.
- Porter, A.L., & Rossini, F.A. (1985). Peer review of interdisciplinary research proposals. *Science, Technology, & Human Values*, 10(3), 33–38.
- Price, D.S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5), 292–306.
- RAND. (2005). Is there gender bias in federal grant programs? RAND: Infrastructure, Safety, and the Environment Brief RB-9147-NSF. Retrieved from http://www.rand.org/pubs/research_briefs/RB9147/RAND_RB9147.pdf
- Rees, T. (2011). The gendered construction of scientific excellence. *Interdisciplinary Science Reviews*, 36(2), 133–145.
- Resch, K.I., Ernst, E., & Garrow, J. (2000). A randomized controlled study of reviewer bias against an unconventional therapy. *Journal of the Royal Society of Medicine*, 93, 164–167.

- Rinia, E.J., van Leeuwen, Th.N., van Vuren, H.G., & van Raan, A.F.J. (2001). Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. *Research Policy*, 30, 357–361.
- Ross, J.S., Gross, C.P., Desia, M.M., Hong, Y.L., Grant, A.O., Daniels, S.R., . . . Krumholz, H.M. (2006). Effect of blinded peer review on abstract acceptance. *Journal of the American Medical Association*, 295(14), 1675–1680.
- Rothwell, P.M., & Martyn, C.N. (2000). Reproducibility of peer review in clinical neuroscience — Is agreement between reviewers any greater than would be expected by chance alone? *Brain*, 123, 1964–1969.
- Rust, J., & Golombok, S. (2009). *Modern psychometrics: The science of psychological assessment* (3rd ed.) New York: Routledge.
- Sandström, U. (2009). Cognitive bias in peer review: A new approach. In *Proceedings of 12th International Conference on Scientometrics and Informetrics*, July 28–31, Rio de Janeiro, Brazil, 1–5.
- Sandström, U., & Hallsten, M. (2008). Persistent nepotism in peer-review. *Scientometrics*, 74(2), 175–189.
- Schultz, D.M. (2010a). Are three heads better than two? How the number of reviewers and editor behavior affect the rejection rate. *Scientometrics*, 84(2), 277–292.
- Schultz, D.M. (2010b). Rejection rates for journals publishing in the atmospheric sciences. *Bulletin of the American Meteorological Society*, 91, 231–243.
- Science and Technology Committee. (2011). Eighth report: Peer review in scientific publications. Retrieved from <http://www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/856/85602.htm>
- Sen, C.K. (2012). Rebound peer review: A viable recourse for aggrieved authors? *Antioxidants & Redox Signaling*, 16(4), 293–296.
- Sense About Science (2009). *Peer review survey 2009: The full report*. London: Sense About Science. Retrieved from http://www.senseaboutscience.org/data/files/Peer_Review/Peer_Review_Survey_Final_3.pdf
- Shapin, S. (1995). *A social history of truth: Civility and science in seventeenth-century England*. Chicago: University of Chicago Press.
- Shatz, D. (2004). *Peer review: A critical inquiry*. Lanham, MD: Rowman & Littlefield.
- Shea, C. (2011, November 13). Fraud scandal fuels debate over practices of social psychology. *Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/As-Dutch-Research-Scandal/129746/>
- Sismondo, S. (2009). Ghosts in the machine: Publication planning in the medical sciences. *Social Studies of Science*, 39(2), 171–198.
- Smith, J.A., Nixon, R., Bueschen, A.J., Venable, D.D., & Henry, H.H. (2002). The impact of blinded versus unblinded abstract review on scientific program content. *Journal of Urology*, 168(5), 2123–2125.
- Smith, R. (1999). Opening up BMJ peer review: A beginning that should lead to complete transparency. *British Medical Journal*, 318, 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, 178–182.
- Solomon, M. (2001). *Social empiricism*. Cambridge, MA: MIT Press.
- Spier, R. (2002). The history of the peer-review process. *Trends in Biotechnology*, 20(8), 357–358.
- Stanford, P.K. (2012). Getting what we pay for: Unconceived alternatives and historical changes in scientific inquiry. Unpublished manuscript, Irvine, CA: University of California, Irvine.
- Steinpreis, R.E., Anders, K.A., & Ritzke D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles*, 41(7/8), 509–528.
- Sugimoto, C.R., & Cronin, B. (in press). Citation gamesmanship: Testing for evidence of ego bias in peer review. *Scientometrics*.
- Thurner, S., & Hanel, R. (2010). Peer-review in a world with rational scientists: Toward selection of the average. *European Physical Journal B*, 84(4), 707–711.
- Timmer, J. (2008). Online peer review supplements, doesn't replace real thing. *Ars Technica*. Retrieved from <http://arstechnica.com/old/content/2008/04/online-peer-review-supplements-doesnt-replace-real-thing.ars>
- Travis, G.D.L., & Collins, H.M. (1991). New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology, & Human Values*, 16(3), 322–341.
- Tregenza, T. (2002). Gender bias in the refereeing process? *Trends in Ecology & Evolution*, 17(8), 349–350.
- Tyler, T.R. (2006). Psychological perspectives on legitimacy and legitimization. *Annual Review of Psychology*, 57, 375–400.
- Valkonen, L., & Brooks, J. (2011). Gender balance in Cortex acceptance rates. *Cortex*, 47, 763–770.
- van Rooyen, S., Delamothe, T., & Evans, S.J.W. (2010). Effect on peer review of telling reviewers that their signed reviews might be posted on the web: Randomized controlled trial. *British Medical Journal*, 341, c5729.
- 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 randomized trial. *British Medical Journal*, 318(7175), 23–27.
- van Rooyen, S., Godlee, F., Evans, S.J.W., Smith, R., & Black, N. (2010). Effect of blinding and unmasking on the quality of peer review. *Journal of the American Medical Association*, 280, 234–237.
- Vines, T. (2010). The famous grouse — Do prominent scientists have biased perceptions of peer review? *Scholarly Kitchen: What's hot and cooking in scholarly publishing*. Retrieved from <http://scholarlykitchen.ssnet.org/2012/02/01/the-famous-grouse-do-prominent-scientists-have-biased-perceptions-of-peer-review/>
- Vines, T., Rieseberg, L., & Smith, H. (2010). No crisis in supply of peer reviewers. *Nature*, 468(7327), 1041.
- Walsh, J. (1975). NSF and its critics in Congress: New pressure on peer review. *Science*, 188(4192), 999–1001.
- Walsh, E., Rooney, M., Appleby, L., & Wilkinson, G. (2000). Open peer review: A randomized controlled trial. *British Journal of Psychiatry*, 176, 47–51.
- Ware, M., & Monkman, M. (2008). Peer review in scholarly journals: Perspective of the scholarly community — an international study. *Publishing Research Consortium*. Retrieved from <http://www.publishingresearch.net/documents/PRCsummary4Warefinal.pdf>
- Webb, T.J., O'Hara, B., & Freckleton, R.P. (2008). Does double-blind review benefit female authors? *Trends in Ecology and Evolution*, 23(7), 351–353.
- Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, 387, 341–343.
- Wesseley, S. (1998). Peer review of grant applications: What do we know? *Lancet*, 352(9124), 301–305.
- Whittaker, R.J. (2008). Journal review and gender equality: A critical comment on Budden et al. *Trends in Ecology and Evolution*, 23(9), 478–479.
- Wing, D.A., Benner, R.S., Petersen, R., Newcomb, R., & Scott, J.R. (2010). Differences in editorial board reviewer behavior based on gender. *Journal of Women's Health*, 19(10), 1919–1923.
- Wood, F.Q. (1997). *The peer review process*. Canberra, Australia: National Board of Employment, Education and Training.
- Xie, Y., & Shauman, K.A. (1998). Sex differences in research productivity: New evidence about an old puzzle. *American Sociological Review*, 63(6), 847–870.
- Yankauer, A. (1991). How blind is blind review? *American Journal of Public Health*, 81(7), 843–845.
- Ziman, J. (2000). *Real science: What it is, and what it means*. Cambridge, UK: Cambridge University Press.