Forthcoming in *Philosophy of Science* 2015.

Commensuration Bias in Peer Review

Carole J. Lee

Department of Philosophy, Box 353350, University of Washington, Seattle, WA 98195

c3@uw.edu

Abstract: To arrive at their final evaluation of a manuscript or grant proposal, reviewers must convert a submission's strengths and weaknesses for heterogeneous peer review criteria into a single metric of quality or merit. I identify this process of commensuration as the locus for a new kind of peer review bias. Commensuration bias illuminates how the systematic prioritization of some peer review criteria over others permits and facilitates problematic patterns of publication and funding in science. Commensuration bias also foregrounds a range of structural strategies for realigning peer review practices and institutions with the aims of science.

Acknowledgements: I am grateful for the Career Enhancement Fellowship (funded by the Mellon Foundation and administered by the Woodrow Wilson Foundation) and Royalty Research Fund Scholars grant #A79071 (from the University of Washington) that afforded research time to complete this article. For helpful references and/or comments, I would like to thank Catherine Elgin, Elena Erosheva, Chris Haufe, John Manchak, Miriam Solomon, Kyle Stanford, and audience members at my 2013 American Association for the Advancement of Science presentation of this material.

1. Introduction.

There is a troubling state of doublespeak in contemporary science. At least, that is what a stylized reading of the peer review bias literature suggests. In public, truth and innovation are valorized by individual scientists (Simmons, Nelson, and Simonsohn 2011; Nosek, Spies, and Motyl 2012; Neuliep and Crandall 1990, 1993), journals (Fontanarosa and DeAngelis 2004; The Lancet 2011), granting agencies (National Institutes of Health 2014b; National Science Foundation 2011a), decrees by Congress (U.S. Senate 2007) and the Executive Office of the President (Orszag and Holdren 2010). Behind closed doors, individual scientists and scientific institutions thwart the achievement of these goals through biases in peer review.

Science's primary aim – truth – has taken a hit. In the aggregate, reviewers and editors prefer to publish new, statistically significant effects (Olson et al. 2002; Dickersin 1990; Neuliep and Crandall 1990, 1993). The disproportionate publication (Fanelli 2010b) of statistically significant results has led to false and exaggerated meta-analysis measures (Palmer 2000; Simes 1986). And, authors, in chase of statistical significance, have increasingly turned to scientifically questionable (Fanelli 2009; John, Loewenstein, and Prelec 2012) and fraudulent (Fang, Steen, and Casadevall 2012) research practices, proliferating the publication of false positives (Simmons, Nelson, and Simonsohn 2011).

Innovation has also taken a hit. The perception that peer review is intellectually conservative is shared by grant applicants (Gillespie, Chubin, and Kurzon 1985; McCullough 1989; National Research Council 2007, 149), grant agency directors (Carter 1979; Kolata

2009), and granting institutions (National Science Foundation 2007a; 2). In response, applicants to the National Institutes of Health [NIH] and the National Science Foundation [NSF] downplay the more transformative aspects of their research proposals (McCullough 1989; Travis and Collins 1991) or (some worry) forgo submitting paradigm-challenging ideas altogether (National Science Foundation 2007a, 7).

Science, the stylized reading goes, is in a situation of doublespeak. In public, scientists and scientific institutions celebrate truth and innovation. In private, they perpetuate peer review biases that thwart these goals. How has science gotten into this situation? And, what can be done about it?

In this paper, I will introduce a new kind of peer review bias – namely, commensuration bias – to explain how this state of affairs has been permitted and facilitated by the under-characterization of peer review criteria taken as a heterogeneous collective. In peer review, reviewers, editors, and grant program officers must make interpretive decisions about how to weight the relative importance of qualitatively different peer review criteria – such as novelty, significance, and methodological soundness – in their assessments of a submission's final/overall value. Not all peer review criteria get equal weight (Rockey 2011); and, weightings can vary across reviewers and contexts even when reviewers are given identical instructions (Lamont 2009, 166; Langfeldt 2001, 828). I identify this process of commensuration – of converting scores along heterogeneous criteria into a single metric or scale of value (Espeland and Stevens 1998) – as the locus for commensuration bias.

Commensuration in peer review can be characterized as being "biased" in a number of ways. Traditionally, peer review "bias" has been understood as reviewer deviation from the impartial application of peer review criteria (Lee et al. 2013, 4-5). Extending this line of thought, commensuration bias can be conceptualized as reviewer deviation from the impartial weighting of peer review criteria in determinations of a submission's final value. Sources of partiality in commensuration can include reviewer idiosyncrasy (Lamont 2009) and social bias – such as subconscious racism – against authors/applicants (Lee and Erosheva 2014).

In this paper, I will explore yet another way of conceptualizing how commensuration may be biased in peer review: in particular, I will explore how reviewer judgments deviate from whatever commensuration practices would, in the aggregate, effectively and efficiently achieve science's goals. This way of conceptualizing commensuration bias evaluates the instrumental value of peer review criteria and commensuration practices towards broader epistemic ends.[1] Commensuration bias, understood in this way, illuminates how intellectual priorities in individual peer review judgments can collectively subvert the attainment of community-wide goals and foregrounds a shared possibility space of structural strategies –

[1] This new instrumental way of conceptualizing what could count as a "bias" in peer review comes apart from the traditional conception in interesting ways: for example, it may be most effective and efficient for a scientific community to achieve its aims by violating impartial peer review through some optimal distribution of partial peer review processes across the community (Lee and Schunn 2011; Lee 2012; Zollman 2009).

applicable to heterogeneous peer review domains – to ameliorate and counterbalance it. This analysis also animates a less Orwellian interpretation of the current state of peer review in science by revealing how commensuration bias can result from an uncoordinated confluence of individual decisions and structural conditions, with multiple loci for enacting change.

2. Commensuration in Peer Review.

Reviewers for manuscripts and grant proposals are asked to make evaluations that have normative and structural features in common. In both cases, reviewers are asked to evaluate submissions along multiple normative dimensions. Journal editors at the top 100 science journals ask reviewers to assess the novelty, significance, and methodological soundness of submitted manuscripts (Frank 1996). NIH asks reviewers to provide separate scores for a proposal's significance, innovation, and approach (a methodology component), as well as for the investigator's qualifications and institutional resources to carry out the proposed research (National Institutes of Health 2014b). NSF asks ad hoc reviewers to assess proposals for "intellectual merit" and "broader impacts," which collectively include sub-criteria for a proposal's potential for transformative research (an innovation component), rationale (a methodology component), whether the PI/team have sufficient qualifications and institutional resources to carry out the project successfully, and social impact/benefits (National Science Foundation 2012).

However, in the end, reviewers for manuscripts and grant proposals are asked to provide a recommendation about where a submission lies along a *single* dimension of evaluation. In the case of manuscript review, the ordinal scale of value includes ranked options like “accept,” “accept with minor revisions,” “revise and resubmit,” and “reject” (Hargens and Herting 1990). At NIH, reviewers are asked to provide an overall impact score on a 9 point rating scale, where “reviewers are instructed to weigh the different criteria as they see fit in deriving their overall scores” (National Institutes of Health 2014c). And, at NSF, ad hoc reviewers and panelists are asked to rate proposals as “excellent,” “very good,” “good,” “fair,” or “poor.” To make these final, global recommendations, reviewers in all of these cases must transform strengths and weaknesses for heterogeneous review criteria into a single scale of value.

This practice of commensuration – of converting distinct qualities into a single metric – is a ubiquitous practice in our social lives (Espeland and Stevens 1998). We assign prices to dissimilar objects/services and calculate disability-adjusted life years lost to radically different kinds of afflictions (Murray and Acharya 1997). By quantifying along a single scale of value, commensuration creates relationships that are easy to represent and compare across fundamentally different things. Its ability to simplify is powerful, but “inherently interpretive, deeply political, and too important to be left implicit” (Espeland and Stevens 1998, 315).

In the philosophy of science, commensuration plays a tacit role in central questions about theory choice. The practice of quantifying along a single scale of value to compare fundamentally different scientific theories is most overt in Bayesian models of science, which provide a calculus for measuring and comparing evidential support for competing scientific theories along a single scale of credence. Commensuration may, on the face of it, seem less obvious when theory choice is driven by epistemic values such as accuracy, scope, and fruitfulness; however, even these cases implicitly involve ranking scientific theories along a scale of epistemic and/or pragmatic “acceptability.”

As Kuhn observed (1977, 325-26), in the absence of a single, objective weight function determining the joint application of distinct epistemic values, one can raise historically, sociologically, and philosophically rich questions about how epistemic values are weighted across individuals, disciplines, and epochs.[2] Analogously, there are important historical, sociological, and philosophical questions to ask about how peer review criteria are weighted across individuals, disciplines, and institutional contexts in science. How do scientists weight heterogeneous peer review criteria in manuscript and grant review? Are there better and worse ways of weighting these criteria to promote truth and innovation in science? If so, how can we change current review processes, institutions, and communities to promote and sustain such improvements?

3. Commensuration Bias in Peer Review.

---

[2] Wesley Salmon (1990) argues that there may be a single Bayesian algorithm at work, where differences in how scientists emphasize epistemic values get expressed as different prior probability values. Here, commensuration takes place a step earlier in the calculation of priors.

To begin, consider manuscript review at top science journals, which involves evaluations of novelty, significance, and methodological soundness (Frank 1996): novelty promotes the discovery of new truths; methodological soundness assesses the likely truth of study conclusions by evaluating the reliability of data collection and analysis strategies; and, determinations of significance tell us which novel truths are most interesting or important (Lamont 2009).

To protect the truth-seeking aims of science, two classes of research should be understood as intellectually significant. Replications are needed to correct and increase confidence in already published results (Nosek, Spies, and Motyl 2012, 617). And, null results are needed to prevent publication bias – the disproportionate publication of studies reporting statistically significant rather than null results – which, in the aggregate, leads to false and exaggerated meta-analysis measures (Palmer 2000; Simes 1986).[3]

Yet, null results are vanishing wholesale from the publication record. A systematic study of 4,600 papers across the physical, social, and biomedical sciences found that the percentage of published papers reporting statistically significant rather than null results ranges from 70.2% in Space Science journals all the way up to 91.5% in Psychology and Psychiatry journals (Fanelli 2010b). Publication bias rates increased between 1990 and 2007 by 22% (Fanelli 2012).

---

[3] These classes of result are not mutually exclusive: for example, failed replications of previously published effects are also null results (Bornstein 1990, 73).

Publication bias appears to be caused primarily by authors who fail to submit null results in preemptive anticipation of their likely rejection (Dickersin, Min, and Meinert 1992; Ioannidis 1998) and perceived unimportance (Easterbrook et al. 1991). To make matters worse, authors increasingly game publication standards (Lee 2013) by adopting methodologically questionable (Fanelli 2009; John, Loewenstein, and Prelec 2012) and fraudulent (Fang, Steen, and Casadevall 2012) research practices that boost the likelihood of achieving statistically significant but false effects (Simmons, Nelson, and Simonsohn 2011).

Unfortunately, replications are also held in low esteem. 72% of social science editors view new effects as "*more important* for the advancement of the field and thus inclusion in their journal" than successful replications, while 58% viewed new effects as more important than failed replications (Neuliep and Crandall 1990, 88, italics mine). Social science reviewers, while having a slightly higher opinion of replication results, overwhelmingly thought that replications should take up 10% or less of journal space (Neuliep and Crandall 1993, 25). Natural science journal editors publish replications at lower rates than social science editors (Madden, Easley, and Dunn 1995, 78).

Overall, the systematic overweighting of intellectual significance qua statistical significance in manuscript review should be understood as a commensuration bias: in the aggregate, it distorts the publication record in ways that interfere with reporting true conclusions and archiving a broad, unbiased evidence base for accurate meta-analysis measures.

Commensuration bias in the evaluation of grant proposals involves a different kind of weighting problem. If we take a primary goal of grant agencies to be "to advance the frontiers of knowledge" (National Science Foundation 2013a) and "foster fundamental creative discoveries, innovative research strategies, and their applications" (National Institutes of Health 2013a), then novelty must be given sufficient weight to avoid commensuration bias in grant review. However, regression analyses of 32,546 NIH applications for fiscal year 2010 estimated that the independent contribution of innovation towards overall impact scores was relatively low (Rockey 2011): overall impact scores changed, on average, by only 1.4 points for each 1 point change in innovation score (holding scores for other review criteria constant); in contrast, overall impact scores changed, on average, by 6.7 points for each 1 point change in approach score (holding scores for other review criteria constant).[4]

Especially innovative projects face special challenges. A survey of 288 NIH reviewers for proposals to the Director's Pioneer Award Program found that, for reviewers, the most innovative projects involved methodological risk, including especially "the use of equipment or techniques that have not been proven or considered difficult" (Lal et al. 2012,

[4] This interpretation of the approach score sounds awkward because overall impact scores range only from 1 to 9 points. It may be that surveyed NIH reviewers are right that a disproportionate number of reviewer scores fall in a narrow area of the score range (National Institutes of Health 2013b).

82-83). Similarly, in-depth interviews of reviewers at the European Research Council found that "frontier" proposals were conceptualized in ways that made it more difficult for them to appear methodologically rigorous (Luukkonen 2012, 54). These results suggest that high innovation is conceptually tied to weaker approach/methodology evaluations.

To counteract conservativism and commensuration bias against innovative projects, program officers could fund highly innovative projects with lower overall scores. However, at NSF, program officers report committing only 1 to 3% of their resources to funding these kinds of proposals, despite having the authority to fund more (National Research Council 2007, 152).

To make matters worse, if applicants come to believe that novelty is too lightly weighted by reviewers and program officers, we would expect applicants to downplay the novel aspects of their proposed projects or fail to submit potentially transformative projects altogether (Travis and Collins 1991; McCullough 1989). Along these lines, a report found that the majority of surveyed NSF reviewers (>60%) evaluated submissions of which fewer than 10% were transformative in nature (National Science Foundation 2007b, 14).

Assessing the future impact and likely success of innovative projects is difficult to do prospectively (National Science Foundation 2007a, 1) and from the limited perspective of established scientific theory (Shatz 2004; Haufe 2013). Furthermore, scientists may resist the prioritization of highly innovative work over incremental science (National Science Foundation 2011b). However, giving sufficient weight to innovation is essential for achieving science's goals. More broadly, the notion of commensuration bias helps us to think directly about how such goals are or are not effectively and efficiently achieved by standing peer review criteria, norms, and practices.

4. Debiasing Commensuration.

Because commensuration practices are considered biased as a downstream consequence of evaluations made by individuals over multiple stages and levels of hierarchy, the notion of commensuration bias opens the door to a multi-layered range of strategies for ameliorating and counterbalancing it. Even though journals and grant agencies serve different functions in science, the shared normative and structural features of their peer review processes foreground a common possibility space of strategies for debiasing commensuration. In what follows, I will articulate a number of these strategies. I will not be able to evaluate them here (for lack of space); however, I will close by articulating the fruitful questions and future projects that this analysis animates.

At the Reviewer Level.

Institutions can debias commensuration directly by providing instructions to reviewers about how to interpret and commensurate peer review criteria. Alternatively, peer review criteria can be abridged to include only those directly relevant for achieving community-wide goals: for example, *PLOS ONE* and *PeerJ* – peer-reviewed, open-access, online journals – ask reviewers to evaluate manuscripts on the basis of technical soundness rather than intellectual significance (PLOS ONE 2013; PeerJ 2014) to promote the truth-seeking aims of science through the successful peer review and publication of replications and null results. Likewise, peer review criteria can be augmented, as when NSF added the requirement that reviewers assess proposals for their transformative potential (National Science Foundation 2007c).

At the Managerial Level.

Editors and program officers can fulfill distributional policy priorities by asking reviewers to score peer review criteria separately and then curating a portfolio of accepted submissions that include some percentage of submissions that score very highly for qualities typically underweighted by reviewers. Alternatively, journal editors and program officers can ask reviewers to use less fine-grained scoring systems, which leads to more ties between competing submissions and gives editors and grant program officers more leeway to use policies (such as innovation mandates at grant institutions) to serve as tie-breakers (Langfeldt 2001; National Institutes of Health 2008).

At the Intra-Institutional Level.

Institutions can increase overall acceptance rates to increase the acceptance of submissions that do not score most highly on reviewers' preferred criteria but excel on underweighted dimensions. Journals can increase acceptance rates by embracing online publication, which has turned page-limits into an anachronism (Nosek and Bar-Anan 2012). Grant agencies can increase acceptance rates by distributing fewer funds to each accepted proposal, pace current trends at NSF (National Science Foundation 2007b, 5) and NIH (National Institutes of Health 2014a). Empirical research suggests that increasing acceptance rates really can debias commensuration: a cross-agency comparison of grant programs for the Research Council of Norway found that reviewers and program officers gave novelty more consideration as proposal funding rates increased (Langfeldt 2001, 833).

Institutions can also counterbalance commensuration bias by housing compensatory, parallel systems under their "brand" (Young, Ioannidis, and Al-Ubaydli 2008, 1421). Traditional print journals can reserve online-only publication for studies deemed methodologically sound but not intellectually significant (qua statistically significant). Grant agencies can create separate funding mechanisms for innovative research (e.g., NIH Director's Transformative Research Awards, NIH's Pioneer Award Program, NSF's Early-Concept Grants for Exploratory Research).

At the Community Level.

New institutions can combat commensuration bias's harmful effects by accepting submissions that do not score most highly on reviewers' preferred criteria but excel on underweighted dimensions. *PLOS ONE* and *PeerJ* publish results not favored elsewhere. Data repositories (e.g., the Dataverse Network, the Dryad Digital Repository) archive and share data for future reanalysis and meta-analysis. And, the Center for Open Science archives and (if desired) publishes a research project's entire lifecycle. In the domain of funding, scientists are increasingly turning to alternative sources of support including crowdsourcing sites (e.g., Experiment.com) and philanthropic and innovation prizes (McKinsey & Company 2009; Knowledge Ecology International 2008).

Future Research.

In future work, I hope that researchers will use analytic techniques – such as simulation models and comparative studies – to address questions I have not been able to explore here. There is the obvious normative question of which commensuration debiasing strategy (or mix of strategies) would most benefit science's aims. Future normative work should also articulate a richer conception of science's aims – e.g., "true explanations, adequate predictions, useful policy suggestions," or something else (Zollman 2009, 198) – and think about how those aims should inform peer review determinations of a submission's "significance."

Additionally, there are feasibility questions about how to change individual and institutional incentives to motivate adoption and adherence to commensuration debiasing strategies. Any strategy that successfully changes peer review priorities is also one that will have to successfully change the culture of what is valued in scientific communities. As such, future feasibility-related work needs to think directly and fundamentally about how to change hypercompetitive scientific disciplines that perpetuate conservative values and practices around reputation-building (Harley and Acord 2011) – perhaps by enlisting other levers for cultural change, such as changes to tenure and promotion criteria (Franzoni, Scellato, and Stephan 2011; Fanelli 2010a) – to realign commensuration practices with the aims of science.

5. Conclusion.

In this paper, I located commensuration practices in peer review as a locus for a new kind of peer review bias, namely commensuration bias. Commensuration bias illuminates how the practice of prioritizing some peer review criteria over others can, in the aggregate, obstruct scientific aims. To address commensuration bias, I identified progressive strategies and institutions designed to decrease the looseness with which peer review criteria are jointly applied by moving toward more finely-tuned criteria, scoring instructions, and/or distributional/institutional policies. Future research will need to explore the normative desirability and practical feasibility of these strategies for realigning commensuration practices with the aims of science, with a special sensitivity to questions about how to re-engineer the reward structures and culture of science.

References

Bornstein, Robert F. 1990. “Publication Politics, Experimenter Bias and the Replication Process in Social Science Research.” *Journal of Social Behavior and Personality* 5:71-81.

Carter, Luther J. 1979. “A New and Searching Look at NSF.” *Science* 204:1064-65.

Dickersin, Kay. 1990. “The Existence of Publication Bias and Risk Factors for its Occurrence.” *Journal of the American Medical Association* 263:1385-89.

Dickersin, Kay, Yuan-I Min, and Curtis L. Meinert. 1992. “Factors Influencing Publication of Research Results: Follow-Up of Applications Submitted to Two Institutional Review Boards.” *Journal of the American Medical Association* 267:374-79.

Easterbrook, Philippa J., Jesse A. Berlin, Ramana Gopalan, and David R. Matthews. 1991. “Publication Bias in Clinical Research.” *The Lancet* 337:867-72.

Espeland, Wendy Nelson, and Mitchell L. Stevens. 1998. “Commensuration as a Social Process.” *Annual Review of Sociology* 24:313-43.

Fanelli, Daniele. 2009. “How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data.” *PLOS ONE* 4:1-11.

---. 2010a. “Do Pressures to Publish Increase Scientists’ Bias? An Empirical Support from US States Data.” *PLOS ONE* 5:1-7.

---. 2010b. ““Positive” Results Increase Down the Hierarchy of the Sciences.” *PLOS ONE* 5:1-10.

---. 2012. “Negative Results are Disappearing from Most Disciplines and Countries.” *Scientometrics* 90:891-904.

Fang, Ferric C., R. Grant Steen, and Arturo Casadevall. 2012. “Misconduct Accounts for the Majority of Retracted Scientific Publications.” *Proceedings of the National Academy of Sciences* 109:17028-33.

Fontanarosa, Phil B., and Catherine D. DeAngelis. 2004. “Information for *JAMA* Authors - 2004.” *Journal of the American Medical Association* 291:107-08.

Frank, Erica. 1996. “Editors’ Requests of Reviewers: A Study and a Proposal.” *Preventative Medicine* 25:102-04.

Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan. 2011. “Changing Incentives to Publish.” *Science* 333:702-03.

Gillespie, Gilbert W., Daryl E. Chubin, and George M. Kurzon. 1985. “Experience with NIH Peer Review: Researchers’ Cynicism and Desire for Change.” *Science, Technology, & Human Values* 10:44-54.

Hargens, Lowell L., and Jerald R. Herting. 1990. “A New Approach to Referees’ Assessments of Manuscripts.” *Social Science Research* 19:1-16.

Harley, Diane, and Sophia Krzys Acord. 2011. Peer Review in Academic Promotion and Publishing: Its Meaning, Locus, and Future. In *The Future of Scholarly Communication Project*. Berkeley: Center for Studies in Higher Education, UC Berkeley.

Haufe, Chris. 2013. “Why do Funding Agencies Favor Hypothesis Testing?” *Studies in History and Philosophy of Science* 44:363-74.

Ioannidis, John 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* 279:281-86.

John, Leslie K., George Loewenstein, and Drazen Prelec. 2012. “Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling.” *Perspectives on Psychological Science* 23:524-32.

Kitcher, Philip. 1990. “The Division of Cognitive Labor.” *The Journal of Philosophy* 87:5-22.

Knowledge Ecology International. 2008. Selected Innovation Prizes and Research Programs. Knowledge Ecology International.

Kolata, Gina. 2009. “Grant System Leads Cancer Researchers to Play it Safe.” *The New York Times*, June 27. Accessed February 14, 2014. http://www.nytimes.com/2009/06/28/health/research/28cancer.html?pagewanted=all&_r=0.

Kuhn, Thomas S. 1977. “Objectivity, Value Judgment, and Theory Choice.” In *The Essential Tension: Selected Studies in Scientific Tradition and Change*, 320-39. Chicago: University of Chicago Press.

Lal, Bhavya, Alyson Wilson, Seth Jonas, Elizabeth Lee, Amy Richards, and Vanessa Peña. 2012. An Outcome Evaluation of the National Institutes of Health (NIH) Director’s

Pioneer Award (NDPA) Program, FY 2004-2006. Alexandria, VA: Institute for Defense Analyses.

Lamont, Michèle. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.

Langfeldt, Liv. 2001. "The Decision-Making Constraints and Processes of Grant Peer Review, and their Effects on the Review Outcome." *Social Studies of Science* 31:820-41.

Lee, Carole J. 2012. "A Kuhnian Critique of Psychometric Research on Peer Review." *Philosophy of Science* 79 (Proceedings):859-70.

---. 2013. "The Limited Effectiveness of Prestige as an Intervention on the Health of Medical Journal Publications." *Episteme* 10:387-402.

Lee, Carole J., and Christian D. Schunn. 2011. "Social Biases and Solutions for Procedural Objectivity." *Hypatia: A Journal of Feminist Philosophy* 26:352-73.

Lee, Carole J., and Elena A. Erosheva. 2014. "Identifying Commensuration Bias in Grant Review." NIH Peer Review Challenge, First Prize: Most Creative Idea, submitted June 14, 2014. http://public.csr.nih.gov/Pages/challenge.aspx.

Lee, Carole J., Cassidy R. Sugimoto, Guo Zhang, and Blaise Cronin. 2013. "Bias in Peer Review." *Journal of the American Society for Information Science* 64:2-17.

Luukkonen, Terttu. 2012. "Conservativism and Risk-Taking in Peer Review: Emerging ERC Practices." *Research Evaluation* 21:48-60.

Madden, Charles S., Richard W. Easley, and Mark G. Dunn. 1995. "How Journal Editors View Replication Research." *Journal of Advertising* 24:77-87.

McCullough, Jim. 1989. "First Comprehensive Survey of NSF Applicants Focuses on their Concerns about Proposal Review." *Science, Technology, & Human Values* 14:78-88.

McKinsey & Company. 2009. "And the Winner is. . . " Capturing the Promise of Philanthropic Prizes. McKinsey & Company.

Murray, Christopher J. L., and Arnab K. Acharya. 1997. "Understanding DALYs." *Journal of Health Economics* 16:703-30.

National Institutes of Health. 2008. "Enhancing Peer Review: The NIH Announces New Scoring Procedures for Evaluation of Research Applications Received for Potential FY2010 Funding." NOT-OD-09-024. Available at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-09-024.html.

---. 2013a. "About NIH: Mission." Accessed December 19. http://nih.gov/about/mission.htm.

---. 2013b. *Enhancing Peer Review Survey Results Report*. Available at http://enhancing-peer-review.nih.gov/docs/Enhancing_Peer_Review_Report_2012.pdf.

---. 2014a. "NIH Data Book." Accessed February 14. http://report.nih.gov/nihdatabook/index.aspx.

---. 2014b. "NIH Grants Policy Statement." Accessed February 1. http://grants.nih.gov/grants/policy/nihgps_2012/nihgps_ch2.htm - nih_org.

---. 2014c. "NIH Reviewer Orientation." Accessed February 14, 2014. http://grants.nih.gov/grants/peer/guidelines_general/reviewer_orientation.pdf.

National Research Council. 2007. Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future. Washington: The National Academies Press.

National Science Foundation. 2007a. *Enhancing Support of Transformative Research at the National Science Foundation*. NSB-07-32. Available at http://www.nsf.gov/nsb/documents/2007/tr_report.pdf.

---. 2007b. "Impact of Proposal and Award Management Mechanisms." NSF 07-45. Available at http://www.nsf.gov/pubs/2007/nsf0745/nsf0745.pdf.

---. 2007c. "Important Notice No. 130: Transformative Research." Available at http://www.nsf.gov/pubs/2007/in130/in130.jsp.

---. 2011a. Empowering the Nation Through Discovery and Innovation: NSF Strategic Plan for Fiscal Years (FY) 2011-6. NSF 11-047. Available at http://www.nsf.gov/news/strategicplan/nsfstrategicplan_2011_2016.pdf.

---. 2011b. National Science Foundation's Merit Review Criteria: Review and Revisions. NSB/MR-11-22. Available at http://www.nsf.gov/nsb/publications/2011/meritreviewcriteria.pdf.

---. 2012. Proposal and Award Policies and Procedures Guide: Part I - Grant Proposal Guide. NSF 13-1. Available at: http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpgprint.pdf.

---. 2013a. "NSF at a Glance." Accessed December 19. http: //www.nsf.gov/about/glance.jsp.

Neuliep, James W., and Rick Crandall. 1990. "Editorial Bias Against Replication Research." *Journal of Social Behavior and Personality* 5:85-90.

---. 1993. "Reviewer Bias Against Replication Research." *Journal of Social Behavior and Personality* 8:21-29.

Nosek, Brian A., and Yoav Bar-Anan. 2012. "Scientific Utopia I: Opening Scientific Communication." *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory* 23:217-43.

Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl. 2012. "Scientific Utoptia II: Restructuring Incentives and Practices to Promote Truth over Publishability." *Perspectives on Psychological Science* 7:615-31.

Olson, Carin M., Drummond Rennie, Deborah Cook, Kay Dickersin, Annette Flanagin, Joseph W. Hogan, Qi Zhu, Jennifer Reiling, and Brian Pace. 2002. "Publication Bias in Editorial Decision Making." *Journal of the American Medical Association* 287:2825-28.

Orszag, Peter R., and John P. Holdren. 2010. Science and Technology Priorities for the FY 2012 Budget, July 21, 2010. Office of Management and Budget.

Palmer, A. Richard. 2000. "Quasireplication and the Contract of Error: Lessons from Sex Ratios, Heritabilities and Fluctuating Asymmetry." *Annual Review of Ecology and Systematics* 31:441-80.

PeerJ. 2014. "Aims & Scope." Accessed February 14. https://peerj.com/about/aims-and-scope/.

PLOS One. 2013. "*PLOS ONE* Journal Information." Accessed December 1. http://www.plosone.org/static/information.

Rockey, Sally. 2011."Correlation Between Overall Impact Scores and Criterion Scores." Last modified March 8. http://nexus.od.nih.gov/all/2011/03/08/overall-impact-and-criterion-scores/.

Salmon, Wesley C. 1990. "Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes." In *Scientific Theories*, ed. C. Wade Savage, 175-204. Minneapolis: University of Minnesota Press.

Shatz, David. 2004. *Peer Review: A Critical Inquiry*. New York: Rowman & Littlefield Publishers, Inc.

Simes, Robert John. 1986. "Publication Bias: The Case for an International Registry of Clinical Trials." *Journal of Clinical Oncology* 4:1529-41.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Findings in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22:1359-66.

Solomon, Miriam. 2001. *Social Empiricism*. Cambridge, MA: The MIT Press.

The Lancet. 2013. "Types of Article and Manuscript Requirements." Accessed December 1. http://www.thelancet.com/lancet-information-for-authors/article-types-manuscript-requirements.

Travis, G. D. L., and H. M. Collins. 1991. “New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System.” *Science, Technology, & Human Values* 16:322-41.

U.S. Senate. 2007. *America COMPETES Act*. 110th, H.R. 2272.

Young, Neal S., John P. A. Ioannidis, and Omar Al-Ubaydli. 2008. “Why Current Publication Practices May Distort science.” *PLOS Medicine* 5:1418-22.

Zollman, Kevin J. S. 2009. “Optimal Publishing Strategies.” *Episteme* 6:185-99.

---. 2010. “The Epistemic Benefit of Transient Diversity.” *Erkenntnis* 72:17-35.