Research Policy xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

The sociology of scientific validity: How professional networks shape judgement in peer review

Misha Teplitskiy[a,**], Daniel Acuna[b], Aïda Elamrani-Raoult[c], Konrad Körding[d], James Evans[e,*]

[a] Harvard University, United States
[b] Syracuse University, United States
[c] Ecole Normale Supérieure, France
[d] University of Pennsylvania, United States
[e] University of Chicago, United States

ARTICLE INFO

Keywords:
Peer review
Research evaluation
Bias
Social network
Co-authorship
Resource allocation

ABSTRACT

Professional connections between the creators and evaluators of scientific work are ubiquitous, and the possibility of bias ever-present. Although connections have been shown to bias predictions of uncertain future performance, it is unknown whether such biases occur in the more concrete task of assessing scientific validity for completed works, and if so, how. This study presents evidence that connections between authors and reviewers of neuroscience manuscripts are associated with biased judgments and explores the mechanisms driving that effect. Using reviews from 7981 neuroscience manuscripts submitted to the journal *PLOS ONE*, which instructs reviewers to evaluate manuscripts on scientific validity alone, we find that reviewers favored authors close in the co-authorship network by ~0.11 points on a 1.0–4.0 scale for each step of proximity. *PLOS ONE*'s validity-focused review and the substantial favoritism shown by distant vs. very distant reviewers, both of whom should have little to gain from nepotism, point to the central role of substantive disagreements between scientists in different professional networks ("schools of thought"). These results suggest that removing bias from peer review cannot be accomplished simply by recusing closely connected reviewers, and highlight the value of recruiting reviewers embedded in diverse professional networks.

1. Introduction

Around the globe, public and private organizations invest more than $2 trillion a year into research and development (Industrial Research Institute, 2017). Many of these organizations, including funders and publishers of scientific research, face the challenging task of allocating financial or reputational resources across scientific projects that require increasingly deep and varied domain expertise to evaluate (Jones, 2009; Wuchty et al., 2007). Because the relevant expertise is generally possessed only by professional peers of the projects' creators, their reviews are considered the gold standard of scientific evaluation. Despite its ubiquity, however, peer review faces persistent critiques of low reliability and bias. Reviewers of a particular scientific work disagree with each other's assessment notoriously often (Bornmann, 2011; Campanario, 1998; Cicchetti, 1991; Marsh et al., 2008). Indeed, agreement is often only marginally better than chance and comparable to agreement achieved for Rorschach inkblot tests (Lee, 2012). An even bigger concern is reviewers' bias for or against particular social and intellectual groups, particularly those to whom they are professionally connected. Given that scientists often work on highly specialized topics in small, dense clusters, the most relevant expert evaluators are typically peers of the research creators. As a result, evaluating organizations often rely on close, relevant connections to a focal work for input and many have suspected that connections between reviewers and creators are the locus of nepotism and associated bias.

Several studies of scientific evaluation have demonstrated that professional connections are, indeed, associated with biased review. For example, recent studies document that those who reviewers grant proposals and candidates for promotion favor the research of collaborators and coworkers (Bagues et al., 2016; Jang et al., 2016; Sandström and Hällsten, 2007; van den Besselaar, 2012). Other research reveals that higher ratings tend to be given to the research of men (Bagues et al., 2017; Wennerås and Wold, 1997). These patterns of widespread disagreement and bias in scientific evaluation greatly

* Corresponding author at: Knowledge Lab and Department of Sociology, University of Chicago, United States.
** Corresponding author at: Laboratory for Innovation Science, Harvard University, United States.
E-mail addresses: mteplitskiy@fas.harvard.edu (M. Teplitskiy), jevans@uchicago.edu (J. Evans).

https://doi.org/10.1016/j.respol.2018.06.014
Received 26 January 2018; Received in revised form 22 June 2018; Accepted 27 June 2018
0048-7333/

complicate selection of the most deserving research and generate new problems, such as "reviewing the reviewers" to identify which provides unbiased information. From the perspective of researchers, evaluation decisions that drive their careers and billions of research dollars are possibly unfair and, to a large extent, the "luck of the reviewer draw" (Cole et al., 1981, p. 885).

Despite the centrality of peer review to the scientific enterprise and the research attention devoted to it, important questions remain. First, existing studies of reviewer bias have focused on *prospective* judgments, like promotions and funding competitions. Administrators' and reviewers' task in these settings is to predict future performance. These prospective judgments are inherently uncertain and may hinge on information asymmetry, such that some reviewers have private information about the applicant that other reviewers lack (Bagues et al., 2016; Li, 2017). It is unknown whether professional connections also influence *retrospective* judgments, such as those in manuscript review, where the task is to evaluate completed work. In retrospective judgments uncertainty about the work should be much lower and, in principle, all reviewers should have equal access to the relevant information, presented explicitly in the manuscript. It is thus plausible that connections are not associated with any bias in retrospective evaluations.

Second, current studies do not distinguish among mechanisms driving bias. We consider three such mechanisms: (1) nepotism, (2) similar tastes on "soft" evaluation criteria like "significance" or "novelty," and (3) shared views on contested substantive matters – a view we call "schools of thought" to denote shared theoretical and methodological assumptions and commitments. Disambiguating these mechanisms is critical because the right policy to mitigate bias in peer review hinges on the mechanism(s) driving it. In the case of nepotism, the most effective policy may be to recuse reviewers closely connected with those reviewed or provide reviewer training on conscious and nonconscious biases in judgment. In the case of soft evaluation criteria, it may be important to separate the review process into components that are technical ("objective") and more subjective. With respect to schools of thought, it may be important to select reviewers embedded in diverse professional networks. In practice, these mechanisms are difficult to disentangle: professional networks overlap with individuals' scientific views, and evaluations typically collapse technical and soft criteria (Lamont, 2009; Lee, 2012; Travis and Collins, 1991).

This study addresses both aforementioned shortcomings of the literature on scientific evaluation. Our research moves beyond prospective judgments and estimates the effect that professional connections play in the concrete, retrospective context of manuscript review. We use the editorial files of 7981 neuroscience manuscripts submitted in 2011-2 to the journal *PLOS ONE*, which instructs reviewers to evaluate manuscripts' scientific validity alone[1]. We measure connections between reviewers and authors by their locations in the co-authorship network. Co-authorship distances are strongly associated with whom authors nominate as reviewers, suggesting that formal co-authorship is an informative signal of affinities between scientists. We find that reviewers give authors a ~0.11 point bonus (1.0 = Reject, 4.0 = Accept) for each step of proximity in the co-authorship network. We do not measure review or manuscript quality directly, so we cannot distinguish whether close reviewers *over*estimate the scientific validity of manuscripts or distant reviewers *under*estimate it. Nevertheless, if a single, true assessment of a manuscript's validity exists, the study reveals bias: reviewers' judgments systematically over- or under-shoot this value as a function of professional proximity.

To explore mechanisms driving reviewer bias, we exploit the uniqueness of *PLOS ONE*'s review process and patterns in reviewer decisions. Unlike conventional journals that evaluate work on both technical and "soft" criteria, such as "significance" or "novelty," *PLOS ONE* evaluates single-blinded[2] manuscripts only on whether they are scientifically valid[3]. Furthermore, *PLOS ONE* greatly limits conflicts of interest by accepting *all* manuscripts meeting standards of scientific validity (about 70% of submissions), regardless of how many related manuscripts are already published or under review. We find that reviewers disagree frequently even on this technical evaluation (inter-rater reliability = 0.19), which suggests that disagreement and biases cannot be attributed to soft evaluation criteria alone. Furthermore, the co-authorship bonus is *not* limited to the closest co-author connections only. Distant reviewers (co-authors of co-authors) give more favorable recommendations than *very* distant reviewers (co-authors of co-authors of co-authors and beyond), despite both types of reviewers having little to gain from nepotism. This pattern suggests that biases are unlikely to be driven by nepotism alone. Instead, we draw on literature from science and technology studies to argue that scientists' views on contested substantive matters overlap with their professional connections. Consequently, closely connected researchers are likely to belong to the same "school of thought" and favor each other's work because it matches their scientific views.

In sum, we find evidence of professional network bias in an unlikely context – judgments of scientific validity regarding completed work by reviewers whom editors choose (at least in principle) specifically for their fairness. The data are most consistent with scientists in a substantive "school of thought" favoring work by others who share their perspective. To the extent that this mechanism is primary, policies used by journals and funding agencies around the world to mitigate bias will be inadequate. Rather than simply recusing the most closely connected evaluators on the assumption of nepotism, our findings suggest that fair evaluations require evaluators from diverse professional networks.

2. Disagreement and biases in peer review

A voluminous literature has documented ways in which scientific evaluations do not necessarily converge on the underlying quality of the work or individual. Given the literature's long-standing focus on disagreement, we first compare levels of disagreement typical of conventional evaluation settings, which simultaneously value validity, significance and novelty, with *PLOS ONE*, which evaluates on validity alone. Next, this section reviews studies of biases in scientific evaluation associated with professional connections. We identify three mechanisms hypothesized to drive bias - nepotism, subjective review criteria, and schools of thought – and discuss contexts in which they are likely to be stronger or weaker. We return to these mechanisms in Section 4.5, which utilizes *PLOS ONE*'s unique review process to disentangle those mechanisms more unambiguously than previously possible.

2.1. Empirical patterns: low reliability and favoritism

Reviewers frequently disagree about which work or person merits publication or funding (Bornmann, 2011; Campanario, 1998; Cicchetti, 1991; Wessely, 1998). Although debate remains regarding whether peer review in multi-paradigm, low-consensus disciplines like sociology is less reliable than in high-consensus disciplines like physics (Hargens, 1988), disagreement is pervasive across disciplines (Bornmann et al., 2010; Cole et al., 1981; Marsh et al., 2008). In reviewing this literature, Cicchetti found that inter-rater reliabilities (0 = no agreement,

[1] We supplement these quantitative data with a small set of editor interviews. Selected editors were drawn randomly from the dataset.

[2] Single blind review is common in the natural and life sciences. How blinding may affect our results is discussed in Section 3.1.

[3] *PLOS ONE* also requires that manuscripts be clearly written and adhere to the journal's data policy. A blank reviewer form is available at the following address: http://journals.plos.org/plosone/s/file?id=t6Vo/plosone-reviewer-form.pdf. Accessed 2017-12-20.