

The Benefits of Double-Blind Review

Hanna Wallach

June 20, 2013

wallach@cs.umass.edu

This essay constitutes a (near) transcript of a talk that I gave at the ICML 2013 Workshop on Peer Review and Publishing Models.

Double-Blind Review

So, I'm sure everyone here knows what double-blind review is, but nonetheless, I'm going to define it anyway—just to make sure we're all on the same page. Double-blind peer review is where the identities of a paper's authors and reviewers are concealed from each other. This is in contrast to single-blind review, in which reviewer identities are concealed from authors but not *vice versa*, and open peer review, where neither the identities of authors or reviewers are concealed. The primary motivation behind double-blind review is to eliminate bias in the reviewing process by preventing factors other than scientific quality from influencing the perceived merit of the work under review. At this point in time, double-blind review is the *de facto* standard for machine learning conferences.

Disclaimers and Context

I want to start with a couple of disclaimers and some context.

First, I want to remind everyone that although I've read a lot about double-blind review, this isn't my research area and I'm not presenting my own research today. As a result, I probably can't answer super detailed questions about the studies that I'll be talking about.

I also want to note that I'm not opposed to open peer review—I was a free and open source software developer for over ten years and I care a great deal about openness and transparency. Rather, my motivation in giving this talk today is simply to create awareness of and initiate discussion about the benefits of double-blind review.

Lastly, and most importantly, I think it's essential to acknowledge that there's a lot of research on double-blind review out there. Not all of this research is in agreement, in part because it's hard to control for all the variables involved and in part because most studies involve a single journal or discipline. And, because these studies arise from different disciplines, they can be difficult to track down—to my knowledge at least, there's no “Journal of Double-Blind Review Research.” These factors make for a hard landscape to navigate. My goal today is therefore to draw your attention to some of the key benefits of double-blind review so that we don't lose sight of them when considering alternative reviewing models.

How Blind Is It?

Before I discuss these benefits, however, I'd like to address one of the most commonly heard criticisms of double-blind review: “But it's possible to infer author identity from content!”—i.e., that double-blind review isn't really blind, so therefore there's no point in implementing it. It turns out that there's some truth to this statement, but there's also a lot of untruth too. There are several studies that directly test this assertion by asking reviewers whether authors or institutions are identifiable and, if so, to record their identities and describe the clues that led to their identification.

The results are pretty interesting: when asked to guess the identities of authors or institutions, reviewers are correct only 25–42% of the time [1]. The most common identification clues are self-referencing and authors' initials or institution identities in the manuscript, followed by reviewers' personal knowledge [2, 3]. Furthermore, higher identification percentages correspond to journals in which papers are required to explicitly state the source of the data being studied [2]. This indicates that journals, not just authors, bear some responsibility for the degree of identification clues present

and can therefore influence the extent to which review is truly double-blind.

Is It Necessary?

Another commonly heard criticism of double-blind review is “But I’m not biased!”—i.e., that double-blind review isn’t needed because factors other than scientific quality do not affect reviewers’ opinions anyway. It’s this statement that I’ll mostly be focusing on today. There are many studies that address this assertion by testing the extent to which peer review can be biased against new ideas, women, junior researchers, and researchers from less prestigious universities or countries other than the US. In the remainder of this talk, I’m therefore going to give a brief overview of these studies’ findings. But before I do that, I want to talk a bit more about bias.

Implicit Bias

I think it’s important to talk about bias because I want to make it very clear that the kind of bias I’m talking about is *not* necessarily ill-intentioned, explicit, or even conscious. To quote the AAUW’s report [4] on the underrepresentation of women in science, “Even individuals who consciously refute gender and science stereotypes can still hold that belief at an unconscious level. These unconscious beliefs or implicit biases may be more powerful than explicitly held beliefs and values simply because we are not aware of them.” Chapters 8 and 9 of this report provide a really great overview of recent research on implicit bias and negative stereotypes in the workplace. I highly recommend reading them—and the rest of the report for that matter—but for the purpose of this talk, I just want you to remember that “Less-conscious beliefs underlying negative stereotypes continue to influence assumptions about people and behavior. [Even] good people end up unintentionally making decisions that violate [...] their own sense of what’s correct [and] what’s good.”

Prestige and Familiarity

Perhaps the most well studied form of bias is the “Matthew effect,” originally introduced by Robert Merton in 1968 [5]. This term refers to the “rich-get-richer” phenomenon whereby well known, eminent researchers get more credit for their contributions than unknown researchers. Since 1968, there’s been a considerable amount of follow-on research investigating the extent to which the Matthew effect exists in science. In the context of peer review, reviewers may be more likely to recommend acceptance of incomplete or inferior papers if they are authored by more prestigious researchers.

Country of Origin

It’s also important to consider country of origin and international bias. Here there’s been research [6] showing that reviewers from within the United States and reviewers from outside the United States evaluate US papers more favorably, with US reviewers showing a stronger preference for US papers than non-US reviewers. In contrast, US and non-US reviewers behaved near identically for non-US papers.

Gender

One of the most widely discussed pieces of recent work on double-blind review and gender is that of Budden et al. [1], whose research demonstrated that following the introduction of double-blind review by the journal *Behavioral Ecology*, there was a significant increase in papers authored by women. This pattern was not observed in a similar journal that instead reveals author information to reviewers. Although there’s been some controversy surrounding this work [7], mostly questioning whether the observed increase was indeed to do with the policy change or a more widely observed phenomenon, the original authors reanalyzed their data and again found that double-blind review favors increased representation of female authors [8].

Race

Race has also been demonstrated to influence reviewers' recommendations, albeit in the context of grant funding rather than publications. Even after controlling for factors such as educational background, country of origin, training, previous research awards, publication record, and employer characteristics, African-American applicants for National Institutes of Health R01 grants are 10% less likely than white applicants to be awarded research funding [9].

Stereotype Threat

I also want to talk briefly about stereotype threat. Stereotype threat is a phenomenon in which performance in academic contexts can be harmed by the awareness that one's behavior might be viewed through the lens of a negative stereotype about one's social group [10]. For example, studies have demonstrated that African-American students enrolled in college and female students enrolled in math and science courses score much lower on tests when they are reminded beforehand of their race or gender [10, 11]. In the case of female science students, simply having a larger ratio of men to women present in the testing situation can lower women's test scores [4]. Several factors may contribute to this decreased performance, including the anxiety, reduced attention, and self-consciousness associated with worrying about whether or not one is confirming the stereotype. One idea that hasn't yet been explored in the context of peer review, but might be worth investigating, is whether requiring authors to reveal their identities during peer review induces a stereotype threat scenario.

Reviewers' Identities

Lastly, I want to talk briefly about the identification of reviewers. Although there's much less research on this side of the equation, it's definitely worth considering the effects of revealing reviewer identities as well—especially for more junior reviewers. To quote Mainguy et al.'s article [12] in PLoS Biol-

ogy, “Reviewers, and especially newcomers, may feel pressured into accepting a mediocre paper from a more established lab in fear of future reprisals.”

Summary

I want to conclude by reminding you that my goal today was to create awareness about the benefits of double-blind review—and I hope I’ve succeeded in this goal. There’s a great deal of research on double-blind review and although it can be a hard landscape to navigate—in part because there are many factors involved, not all of which can be trivially controlled in experimental conditions—I hope I’ve convinced you that there are studies out there that demonstrate benefits of double-blind review. Perhaps more importantly though, I hope I’ve convinced you that double-blind review promotes the *perception* of fairness. To again quote Mainguy et al., “[Double-blind review] bears symbolic power that will go a long way to quell fears and frustrations, thereby generating a better perception of fairness and equality in global scientific funding and publishing.”

References

- [1] Budden, Tregenza, Aarssen, Koricheva, Leimu, Lortie. “Double-blind review favours increased representation of female authors.” 2008.
- [2] Yankauer. “How blind is blind review?” 1991.
- [3] Katz, Proto, Olmsted. “Incidence and nature of unblinding by authors: our experience at two radiology journals with double-blinded peer review policies.” 2002.
- [4] Hill, Corbett, St, Rose. “Why so few? Women in science, technology, engineering, and mathematics.” 2010.
- [5] Merton. “The Matthew effect in science.” 1968.
- [6] Link. “US and non-US submissions: an analysis of reviewer bias.” 1998.

- [7] Webb, O'Hara, Freckleton. "Does double-blind review benefit female authors?" 2008.
- [8] Budden, Lortie, Tregenza, Aarssen, Koricheva, Leimu. "Response to Webb et al.: Double-blind review: accept with minor revisions." 2008.
- [9] Ginther, Schaffer, Schnell, Masimore, Liu, Haak, Kington. "Race, ethnicity, and NIH research awards." 2011.
- [10] Steele, Aronson. "Stereotype threat and the intellectual test performance of African Americans." 1995.
- [11] Dar-Nimrod, Heine. "Exposure to scientific theories affects women's math performance." 2006.
- [12] Mainguy, Motamedi, Mietchen. "Peer review—the newcomers' perspective." 2005.