# What hiding applicant names reveals about discrimination in evaluations

Haruka Uchida / 10 Dec 2025

'Blinding' – concealing candidate names during evaluations – is a commonly used policy to counter discrimination in evaluations. This column presents experimental evidence from a major international academic conference to assess whether blinding improves representation without reducing quality. Blinding eliminates the gap in conference acceptance rates between students and senior researchers and reduces score disparities linked to institution rank, while maintaining predictive power for future citations and publications. Blinding can shift representation by career stage and institutional prestige without sacrificing quality assessment.

*Editors' note: Vox has launched a new section – Job Market Papers 2025 – that features short Vox columns written by PhD candidates on their Job Market Papers. The main goal is to provide a platform for excellent research that will not appear in journals or major discussion paper series for years. It is also a means for established economists to more easily track the research of the youngest members of the profession.*

Disparities in evaluations are widespread – in hiring decisions, loan approvals, and even criminal justice proceedings (e.g. Bertrand and Duflo 2017, Lang and Spitzer 2020). A common policy proposed to remedy such disparities is to conceal candidate names during evaluations, known as 'blinding'. If evaluators are using candidate identity information to discriminate, then shouldn't hiding this information improve outcomes?

Whether blinding is a successful policy or not remains contested, with two fundamental questions unanswered. First, does blinding cause evaluators to change *who* receives favourable evaluations without sacrificing the ability to screen for *quality*? Past work has largely focused on how blinding affects representation (e.g. Goldin and Rouse 2000, Blank 1991), but whether blinding induces a representation-quality tradeoff is ambiguous because it depends on how evaluators use the information conveyed by names. If evaluators use the information to cater to biased preferences (Becker 1957), then blinding can simultaneously increase representation and improve the selection of high-quality candidates. On the other hand, if evaluators use names as informative signals of quality (Phelps 1972, Arrow 1973, Aigner and Cain 1977), then blinding can shift representation at the expense of quality. The possibility of these different mechanisms leads to the second question: why do disparities in evaluations exist in the first place?

In my paper (Uchida 2025), I answer these questions by running two field experiments in the review process of a major international academic conference in computational neuroscience. The setting is policy-relevant: whether blinding should be used in academia continues to be debated, and gaps persist in important outcomes such as opportunities to present work, journal acceptances, and tenure decisions across gender, institution prestige, and career stage (e.g. Blank 1991, Sarsons 2017, Doleac et al. 2021).

I begin by asking how blinding changes reviewer decisions. The main challenge in answering this question is that in most settings, reviewers and applicants choose whether to participate in a blind process, so that differences across blind and non-blind regimes may reflect differences in who chooses to take part rather than the effect of blinding itself. I overcome this by using two stages of randomisation. First, each of the 245 reviewers was randomly assigned to see author lists ('non-blind') or not ('blind'). This ensures that blind and non-blind

reviewers have similar characteristics on average. Second, each of the 657 submitted papers was randomly assigned to two blind and two non-blind reviewers. This allows me to compare how the same paper is judged under blind versus non-blind review.

## Hiding author names helps early-career applicants and applicants from non-top-20 ranked institutions

For each assigned submission, reviewers receive a title, 300-word abstract, and 2-page summary, and give a score from 1 to 10. To understand *who* benefits from blinding, I test how scores of blind versus non-blind reviewers differ by applicants' student status (students, post-PhD), affiliated institution rank (top-20, non-top-20), and gender (male, female). The applicant is the individual who submits the work and would present it if accepted. Because co-authorship is common in computational neuroscience, I also show the effects by co-author characteristics in my paper.

Figure 1 shows the average scores by applicant traits and whether the reviewer received author lists or not. Among non-blind reviewers (i.e. those who see author names), applicants who are post-PhD and from top-20 ranked universities receive significantly higher scores than applicants who are students or affiliated with lower-ranked institutions. When the same submissions are scored by reviewers who do not receive author lists, score gaps by student status and institution rank shrink.

**Figure 1** Reviewer scores and effects of blinding

These changes in score gaps lead to changes in acceptances by student status. Blinding essentially eliminates the difference in acceptance rate between students and more senior applicants. Acceptance-rate gaps by institution rank are significantly reduced when the conference accepts a small fraction of papers. Interestingly, I do not find a significant change by gender. Overall, these effects of blinding on score gaps show that reviewers do use author names when the information is provided to them.

## Hiding author names changes representation while preserving the evaluation's ability to screen on quality

A key policy concern is that blinding shifts representation at the cost of quality. If author names convey information that is predictive of underlying paper quality, then hiding them may remove useful information. For instance, names could convey career stage, and applicants further along in their careers may be more likely to produce high-quality research. The main challenge in answering this question is that in most evaluation settings, underlying candidate quality is not observed by the researcher.

To test whether blinding interferes with the ability to select high-quality submissions, I track each submission for five years after the conference, collecting proxy measures of its quality: citation and publication outcomes.

Figure 2 plots the relationship between a paper's percentile rank in quality and its percentile rank in blind and non-blind scores. I find that a paper's blind score is as good a predictor of its citation and publication outcomes as its non-blind score. Papers that would be accepted under blind review have comparable citation and publication outcomes as those that would be accepted under non-blind review.

**Figure 2** Effects of blinding on quality

## The nature and extent of discrimination can determine how blinding affects representation and quality

How can blinding change representation of selected applicants without changing quality? And what underlying forces drive the gaps in evaluations in the first place?

To answer these questions, I build on past work examining sources of disparities (e.g. Canay et al. 2024) and develop a model of how non-blind reviewers assign scores using the content of the submission and author traits. I use my model to decompose disparities into four distinct forms of discrimination:

1. Accurate statistical discrimination (Phelps 1972, Arrow 1973, Aigner and Cain 1977), where reviewers use author group memberships to accurately update beliefs on underlying paper quality.

2. Inaccurate statistical discrimination (Bordalo et al. 2019, Bohren et al. 2019, Coffman et al. 2021), where reviewers rely on inaccurate beliefs about paper quality

3. Pursuit of alternative objectives beyond paper quality, such as favouring applicants whose acceptance would benefit others the most.

4. All other determinants of disparities, including taste-based discrimination and animus (Becker 1957).

The question is then: across submissions with comparable content, to what extent can differences in reviewer scores across author traits be attributed to differences in true quality, differences in reviewers' misbeliefs about quality, or reviewers' beliefs over alternative objectives?

Estimating this model requires overcoming two main challenges. The first is that distinguishing between actual and perceived quality differences requires observing reviewer beliefs. I therefore run a second experiment with the same conference that directly elicits

reviewers' beliefs during the review process about submission outcomes (future citation and publication status) and alternative objectives (for example, how much the applicant's acceptance would benefit others).

The second challenge is accounting for submission content. This is a longstanding issue because oftentimes evaluators may base decisions on aspects of a submission's content that researchers do not have data on. Failing to account for submission content implies that differences in outcomes, such as quality beliefs across groups, may be driven by differences in submission content rather than discrimination based on author traits. I address this comparability issue by using blind scores as a proxy for submission content, given that blind reviewers assign scores using only submission content without receiving author names.

Figure 3 presents the decomposition results. I find that the underlying forms of discrimination driving disparities in reviewer scores differ across traits. The entirety of the score gap by student status can be explained by two channels: reviewers hold overly pessimistic beliefs about the quality of papers submitted by students and value alternative objectives such as talk quality, which they believe is worse for students than for more senior applicants. In contrast, the score gap by institution rank is not explained by these channels and is instead consistent with a preference for applicants from top-ranked institutions (or animus against those from non-top-20 ranked institutions).

**Figure 3** Decomposing non-blind score gaps

In sum, the efficacy of blinding depends on why disparities exist in the first place. My experiments show that blinding can shift representation, particularly by career stage and institution rank, without compromising on the ability to screen on quality. My model decomposition helps explain why: the mechanisms that generate changes in representation can offset each other in their effects on quality. More broadly, the decomposition demonstrates how data from blind evaluations can be leveraged to learn more about the sources of disparities that exist in the absence of blinding.

These insights, and the methodology developed in my paper, extend beyond academic review processes. Many policy-relevant evaluation settings, including job hiring, grant allocation (Li 2017), and social insurance receipt (Low and Pistaferri 2025), face potential trade-offs between information, representation, and quality. Understanding the mechanisms driving disparities therefore remains essential for designing fair and effective evaluation systems.

# References

Aigner, D J, and G G Cain (1977), "Statistical theories of discrimination in labor markets", *ILR Review* 30(2): 175–187.

Arrow, K J (1973), *The theory of discrimination*, Princeton University Press.

Becker, G S (1957), *The economics of discrimination,* University of Chicago Press.

Bertrand, M, and E Duflo (2004), "Field experiments on discrimination", in *Handbook of economic field experiments*, Volume 1, Elsevier.

Blank, R M (1991), "The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review", *American Economic Review* 1041–67.

Bohren, J A, A Imas, and M Rosenberg (2019), "The dynamics of discrimination: Theory and evidence", *American Economic Review* 109(10): 3395–436.

Bordalo, P, K Coffman, N Gennaioli, and A Shleifer (2019), "Beliefs about gender", *American Economic Review* 109(3): 739–73.

Canay, I A, M Mogstad, and J Mountjoy (2024), "On the use of outcome tests for detecting bias in decision making", *Review of Economic Studies* 91(4): 2135–67.

Coffman, K B, C L Exley, and M Niederle (2021), "The role of beliefs in driving gender discrimination", *Management Science* 67(6): 3551–69.

Doleac, J L, E Hengel, and E Pancotti (2021), "Diversity in economics seminars: Who gives invited talks?", *AEA Papers and Proceedings* (111): 55–59.

Goldin, C, and C Rouse (2000), "Orchestrating impartiality: The impact of 'blind' auditions on female musicians", *American Economic Review* 90(4): 715–41.

Lang, K, and A K L Spitzer (2020), "Race discrimination: An economic perspective", *Journal of Economic Perspectives* 34(2): 68–89.

Li, D (2017), "Expertise versus bias in evaluation: Evidence from the NIH", *American Economic Journal: Applied Economics* 9(2): 60–92.

Low, H, and L Pistaferri (2025), "Disability insurance: Error rates and gender differences", *Journal of Political Economy* 133(9): 2962–3018.

Phelps, E S (1972), "The statistical theory of racism and sexism", *American Economic Review* 62(4): 659–61.

Sarsons, H (2017), "Recognition for group work: Gender differences in academia", *American Economic Review* 107(5): 141–5.

Uchida, H (2025) "What do blind evaluations reveal? How discrimination shapes representation and quality", SSRN Working Paper.

**AUTHORS**

Haruka Uchida

SHARE

**AUTHORS**

Haruka Uchida

SHARE