The bias-reducing effect of voluntary anonymization of authors' identities: Evidence from peer review

Inna Smirnova[1*], Daniel M. Romero[1, 2, 3], Misha Teplitskiy[1,4*]

[1] School of Information, University of Michigan, Ann Arbor, MI

[2] Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI

[3] Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI

[4] Laboratory for Innovation Science at Harvard, Harvard University, Cambridge, MA

[*] To whom correspondence should be addressed. Email: innas@umich.edu; tepl@umich.edu.

This version: January 27, 2023.

Abstract

A persistent concern in formal evaluation processes like scientific peer review is that evaluators favor projects by prestigious producers. While anonymizing producers' identities can reduce prestige bias, achieving anonymization can be difficult because explicit and subtle signals of identities can appear throughout projects, and enforcing compliance can be prohibitively expensive. An attractive alternative is to encourage—"nudge"—producers to anonymize their submissions without checking for compliance. However, encouraging anonymization could fail to reduce bias if low-prestige producers do not anonymize because they do not anticipate bias ("information frictions") or high-prestige producers seek to benefit from the bias ("strategic non-anonymization"). We test these arguments in a quasi-experiment with IOP Publishing, one of the largest academic publishers, which adopted a policy encouraging anonymization and rolled it out across its portfolio of 57 physical sciences journals. Examining 156,015 submissions and measuring first author prestige with citations, we identify causal effects of voluntary anonymization. We find that neither information frictions nor strategy fully explain anonymization rates: highest-prestige authors anonymized less often but still substantially (19%), and the rate varied greatly by geography and gender. More importantly, the nudging policy increased positive peer reviews of low-prestige authors by 2.4% and acceptance by 5.6%. For middle- and high-prestige authors, the policy decreased positive reviews (1.8% and 1% [*n.s.*]) and final acceptance (4.6% and 2.2% [*n.s.*]). Overall, simply nudging producers to anonymize their submissions can help reduce prestige bias and should be considered by organizations for which enforcing full anonymization is impractical.

Keywords: evaluation processes, peer review, prestige bias, voluntary anonymization, causal evidence.

We are grateful for help and suggestions from Hao Peng, Kim Eggleton, Sharon Gordon, Charles Ayoubi, Michael Menietti as well as seminar participants at the Laboratory for Innovation Science at Harvard and IOP Publishing. We thank Schmidt Futures (grant number 2022-5-23-9) for financial support. All errors are our own.

Introduction

Many innovation-oriented organizations and individual consumers allocate their money and attention across projects and products based on the judgments of experts in the field. Examples of expert evaluations can be found across industries (Sharkey, Kovacs, and Hsu 2022), from scientific peer review (Li 2017) to literature and the arts (Chong 2020; Goldin and Rouse 2000; Reimers and Waldfogel 2021; Reinstein and Snyder 2005) to project assessment in the professional services (Criscuolo et al. 2017) and pharmaceutical industries (Edmondson and Gulati 2021). A persistent concern in such evaluations is that experts' assessments may be biased by knowing the identities of the individuals behind the idea or project (Snodgrass 2006; Vollaard and van Ours 2022). A voluminous literature has examined biases related to prestige (Merton 1968; Blank 1991; Tomkins, Zhang, and Heavlin 2017; Guo, Xin, and Barnes 2022), gender (Ceci and Williams 2011; Coffman 2014; Fox and Paine 2019; Rivera and Tilcsik 2019; Goldin and Rouse 2000), and race and ethnicity (Ernst and Kienbacher 1991; Harris et al. 2017; Murray et al. 2019; Nakamura et al. 2021; Quillian et al. 2019; Subbaraman 2021; Chen et al. 2022). Double-anonymous (i.e., "double-blind") evaluation process, in which evaluators are not shown the identities of the idea producers, has been long proposed as a method to reduce bias (Blank 1991; Tomkins, Zhang, and Heavlin 2017; Ho et al. 2013; Mulligan, Hall, and Raphael 2013; Goues et al. 2018; Jamali et al. 2020; Goldin and Rouse 2000; Snodgrass 2006; Huber et al. 2022), as well as better filter impactful work from flops (Sun, Barry Danfa, and Teplitskiy 2022). The strength of evidence for how well anonymization reduces bias varies across different producer characteristics, with arguably the strongest evidence for bias related to producer prestige (Blank 1991; Tomkins, Zhang, and Heavlin 2017; Simcoe and Waguespack 2011).

If the evidence shows that double-anonymous review reduces prestige bias in evaluations, and organizations wish to limit bias, why is it not universally adopted? One critical challenge is monitoring compliance. Truly double-anonymous evaluation requires that the projects or ideas being evaluated do not have clear signals of producers' identities. However, in projects like research papers, explicit or subtle identity signals can be present throughout, and ensuring that producers remove all such signals can be prohibitively costly. For example, one team found that removing identity signals from the U.S. NIH grant proposals took about 2-8 hours per proposal, and quality assurance took another 2-4 hours (Nakamura et al. 2021). In another case, the publishing company Institute of Physics Publishing (IOP) observed that about half of manuscripts submitted to journals with mandatory anonymization were not fully anonymized (personal communication, July 4, 2022). Non-compliant papers then needed to be sent back to the authors, leading to multiple time-consuming iterations of editing and reading. The company estimated that enforcing full anonymization would cost an additional 1,878 hours of editorial time per year. Another challenge is that organizations may be reluctant to expend so much effort on anonymization “in the age of Google,” when evaluators may be able to identify producers through simple searches anyway (Guo, Xin, and Barnes 2022). Why require anonymization if it will not work? Lastly, in science, less prestigious journals that lack “monopoly power” may worry that mandating anonymization would add extra burdens on authors and discourage submissions. Together, these constraints make it difficult for organizations to simply adopt truly double-anonymous evaluations. Many organizations continue using the conventional single-anonymous evaluation, in which the evaluators’ identities are hidden from the producers, but the identity of the producers is revealed to the evaluators. For example, the publishing company Wiley reports that 62% of all of their 1,600 journals are single-anonymous, with the ones in Physical and Life Sciences showing a rate of over 90% (Wiley 2022).

In this paper, we consider an alternative way to reduce prestige bias that is simple and cost-effective: encouraging but not forcing producers to anonymize their submissions (i.e., voluntary anonymization). Voluntary anonymization does not require checking for compliance and thus removes a large bottleneck for organizations. However, there are competing reasons to expect this method to reduce bias or fail to do so. A key question that determines whether voluntary anonymization succeeds is who, if anyone, chooses to anonymize. Different producers may have different motivations. For high-prestige producers, “strategic non-anonymization” might be the dominant choice, as producers may seek to use their reputation to secure better evaluations, hence keeping prestige bias in the system intact. Low-prestige producers, on the other hand, might also choose not to anonymize their submissions because they simply do not anticipate bias. Lack of anticipation might be caused by disparities between producer’s perceived and actual prestige, or simply a lack of knowledge about prestige bias. These “information frictions” may affect high-prestige producers as well, who may anonymize more often than they “should.”

Assessing which mechanism dominates in practice, and whether voluntary anonymization reduces bias has proven difficult. Ideally, an organization would be willing to 1) change its evaluation process and, 2) implement the change as an experiment or stagger the changes to the process over time to prevent confounding by common causes. Both requirements are necessary to identify causal effects in the field but challenging for most organizations. For example, between March 2015 and February 2017 *Nature Publishing Group* experimented with introducing two concurrent submission options—double- and single-anonymous—on a set of their 25 journals but failed to roll out the policy changes in a staggered fashion that did not allow them to identify any causal effects beyond the descriptive information (McGillivray and De Ranieri 2018), leaving stakeholders unsure about the true success of switching to such evaluation model and abandoning the double-anonymous option. Consequently, there are no existing studies establishing whether this attractive method does or does not work in settings that rely on expert evaluations.

Here, we address these challenges directly and estimate the causal effect of voluntary anonymization in a particularly important and high-stakes setting—scientific peer review. To this end, we partnered with the scientific publishing company IOP, which accounts for about 10% of all submissions in the field of physical sciences (Larivière, Haustein, and Mongeon 2015) (see *Data and Methods*). The company adopted a new policy encouraging authors to anonymize their manuscripts, without enforcing compliance, and rolled the policy out in a staggered timeline between November 2020 and December 2021. We conceptualize this form of voluntary anonymization as a nudge. Nudges are non-coercive and non-monetary policy interventions aimed at steering individuals' choices, such as the change of language or a format in which the decision options are presented (Thaler and Sunstein 2008; Hertwig and Grüne-Yanoff 2017; Beshears and Kosowsky 2020; Mertens et al. 2022). We use data on 156,015 submissions to IOP's 57 peer-reviewed journals received over the period of January 2018-February 2022. Through a difference-in-difference analysis, we estimate the causal effects of the IOP's new policy on first authors of various prestige groups at the three major review stages—editor's desk, external peer review, and final decision. We conceptualize author prestige as deference received from others (Sauder, Lynn, and Podolny 2012) and measure it with citations accrued up to submission year (Sun, Barry Danfa, and Teplitskiy 2022), with 0 citations defined as "low prestige," 1-100 citations as "middle prestige", and 101+ citations[1] as "high prestige." In the *Supplementary Appendix*, we consider alternative specifications, varying both the focus to corresponding authors and the operationalization of prestige. Using these data, we ask two research questions (RQs): First, when nudged, which authors choose to anonymize (RQ1)? Second, does the voluntary anonymization policy help reduce prestige bias in scientific evaluations (RQ2)? Additionally, we consider whether anonymization nudges have (unintended) effects on the quantity or characteristics of authors (RQ3) or reviewers (RQ4) who submit or review papers.

Our analyses lead to two key contributions. First, we contribute to the literature on de-biasing expert evaluations. While previous literature shows that enforced anonymization can be effective for de-biasing, we add to the managers' toolkit another, simpler alternative: voluntary anonymization. In answering RQ2, we find that nudging authors to anonymize generally improves the outcomes for low-prestige authors, while marginally hurting those for middle- and high-prestige ones. This method reduces bias by a smaller amount than enforced anonymization, but at essentially no cost, making it an attractive choice for resource-constrained organizations. Second, we contribute to the literature on anticipated discrimination. How individuals anticipate discrimination is very important in practice because they make strategic

[1] Corresponds to the top 25% of the citations distribution.

choices of what information to conceal or where not to allocate their time and effort based on it. Previous research on inequalities in labor markets (Kang et al. 2016) shows that disparity between anticipated and actual discrimination can have some pathologies, increasing bias and economic inequality overall. In answering RQ1, we provide empirical evidence that the authors' choice to anonymize is only partially explained by strategic non-anonymization and there are substantial information frictions present. Information frictions are usually viewed negatively, but in this setting, we argue, they lead to at least some high-prestige individuals' anonymizing their submissions, which in turn enables voluntary anonymization to improve fairness in peer review, at least in the short term.

Anonymization formats and prestige bias in evaluations

Previous work on anonymization in evaluations has found the double-anonymous review process to be an effective method to increase the success of authors with marginalized identities and reduce prestige bias that the evaluation process often suffers from (Snodgrass 2006; Simcoe and Waguespack 2011; Tomkins, Zhang, and Heavlin 2017; Huber et al. 2022; Ucci, D'Antonio, and Berghella 2022). Simcoe and Waguespack (2011) identified a unique natural experiment: the names of authors submitting proposals to the Internet Engineering Task Force, a community responsible for curating the technical standards used in the Internet, were sometimes obscured with "*et al*." Using this variation in name visibility the authors found that status signals explain up to three-quarters of the difference in publication rates among authors—when *et al*. obscured the name of a high-status author, the publication probability of a proposal dropped significantly relative to when it hid the name of a low-status author (Simcoe and Waguespack 2011). More recent experiments from scholarly peer review further support these conclusions. Huber and colleagues (2022) found that 58.8% of the reviewers recommended accepting the paper when the Nobel laureate was shown as the author vs. only 9.9% when the junior scholar's name was revealed to the reviewers, illuminating the existence of a strong prestige bias. Similarly, Tomkins, Zhang, and Heavlin (2017) found that when reviewing papers submitted to a computer science academic conference, single-anonymous reviewers were more likely than double-anonymous reviewers to provide an acceptance recommendation for papers of famous authors.

Although the evidence shows that double-anonymous review consistently reduces prestige bias in evaluations, it can be costly to fully enforce it due to both the burden on editors and authors' resistance to change. The American Economic Association, one of the oldest and leading scholarly venues dedicated to the discussion and publication of economics research, dropped the double-anonymous review process completely in July 2011 after a short period of experimentation, citing the administrative costs of anonymization as one of the reasons behind this decision (Fischman 2011).

An alternative way to overcome such enforcement-related challenges is to offer authors the option of submitting fully anonymized vs. non-anonymized works via just offering or encouraging the anonymization option. The effectiveness of this method has been much less studied, however. One existing attempt is when *Nature Publishing Group* temporarily introduced two submission tracks for a set of their 25 journals: McGillivray and De Ranieri (2018) report that the initiative was not popular, however, and only 12% of submitters during the pilot's timeframe opted for double-anonymous review after the new review option was offered. Only 25% of the papers under double-anonymous track were eventually accepted vs. 44% for papers that were submitted to the single-anonymous route. While this evidence is informative, the implementation was not designed to produce credible causal estimates.

Here, unlike previous inquiries, we aim to estimate causal effects of voluntary anonymization. Specifically, we study whether the bias-reducing benefits of enforced double-anonymization, which is often impractical for organizations to adopt, can be achieved at a much lower cost by nudging authors to anonymize their works, without enforcing it.

Voluntary anonymization and self-selection

The key factor affecting whether anonymization-via-nudging can reduce bias is who chooses to anonymize. Selection into anonymization can be driven by a variety of factors, including differential interpretation of the nudge, cultural norms, and so on. Here we focus on two key mechanisms, information frictions and strategic non-anonymization.

On the one hand, the choice to voluntarily anonymize one's submission is likely to depend on whether authors anticipate being discriminated against, or towards. For example, a series of resume audit experiments found that racial minorities conceal race signals on their resumes more often when applying to jobs from employers they anticipate to discriminate based on race (Kang et al. 2016). However, the same study shows that the level of discrimination individuals anticipate may not match reality, an inconsistency we call "information frictions." Information frictions are especially likely to arise in settings with status inconsistency (Stryker and Macke 1978; Bacharach, Bamberger, and Mundell 1993). Specifically, an individual who holds high status in one hierarchy may not anticipate discrimination that occurs due to low status in another hierarchy (Lenski 1967). In the Hollywood film industry, for instance, performers are often positioned in both the artistic and commercial status hierarchies and decide which films to pursue to (potentially) boost either their box office success or critical acclaim depending on how they perceive the prestige of these two status hierarchies (Han and Pollock 2021). In the context of science, such status inconsistency might occur when an author holds a high-status scientific role in a country that is viewed as low-status by editors and reviewers (Harris et al. 2017; Shukla 2021). If information frictions between anticipated and real discrimination exist, their effect is likely to vary by whether there is discrepancy between self-perceived and externally-perceived prestige, or lack of knowledge about discrimination in general. In the former case, low-prestige authors who do not perceive themselves as such may decrease anonymization, while high-prestige authors may do the opposite. If authors are unaware of discrimination, however, they may not anonymize regardless of their characteristics.

On the other hand, nudging may also fail to reduce bias due to "strategic non-anonymization" by high-status authors. High-status authors may choose not to anonymize because the existing prestige bias is likely to benefit them. Consistent with this view, when *Nature Publishing Group* offered authors an option of double-anonymous review across some of its journals, authors from the least prestigious institutions chose to fully anonymize their works at a significantly higher rate (13%) compared to authors from the most prestigious institutions (4%) and authors from the mid-range institutions (8%) (McGillivray and De Ranieri 2018). If high-status authors do not anonymize, their decisions could eventually undermine the decisions of low-status authors. Indeed, the "full-disclosure principle" suggests that individuals sometimes divulge certain information about themselves (politicians disclosing their tax records, for example) to prevent audiences from making unflattering conclusions if they remained silent (Frank 2008). In the context of peer review, this principle would predict that over time no author should anonymize because it would give away that they hold low status. Consequently, unenforced anonymization has the potential to recreate, if not exacerbate, the original prestige bias. This possibility is supported by research on the effect of the "ban-the-box" policies, which aimed to reduce the disparity between employment opportunities of black and white job seekers by hiding information about individuals' criminal records (Pager 2003). Recent work found that banning the box indeed encouraged stronger statistical discrimination and even increased the race gap in employment (Agan and Starr 2018).

In sum, how much nudging producers to anonymize their works reduces prestige bias depends on selection into anonymization. Information frictions and strategic non-anonymization are two key factors that may undermine bias reduction. Consequently, we pose two research questions:

1. When nudged to anonymize submissions, who complies?
2. What is the causal effect of anonymization-via-nudging on reducing prestige bias?

Lastly, it is important to consider whether anonymization—enforced or voluntary—affects the recruitment of submitting authors or reviewers, which is a key consideration for organizations like publishing companies. First, some authors may see changes to a journal's peer review policy as unnecessary, burdensome, or removing advantages they typically enjoy. Such authors may decide to submit their manuscripts to other venues. Similarly, the quantity and quality of reviewers recruited by editorial staff may suffer since reviewing papers of prestigious authors may be an incentive for some reviewers to accept a review assignment (Tomkins, Zhang, and Heavlin 2017). As a result, there might be a risk for journal editors that high-quality reviewers will agree to review anonymized papers at a lower rate (Huber et al. 2022). Relatedly, the types of reviewers recruited to review papers of higher- vs. lower-status authors when identities are not visible might change, leading to potentially lower quality reviews for one or the other group. This discussion leads to the following research questions:

3. Do anonymization nudges hurt author recruitment, in terms of quantity or quality?
4. Do anonymization nudges hurt reviewer recruitment, in terms of quantity or quality?

Data and Methods

Institutional context: IOP switches to a new policy

In November 2020, IOP announced that "it will be phasing in double-blind peer review for all of its journals" (Anderson 2020). Given the historical prevalence of single-anonymous review in physical sciences, this was a substantial change, aimed at engendering more fairness in evaluations. The policy, however, did not imply that anonymization would be enforced. Instead, IOP encouraged (nudged) authors to anonymize. Specifically, in journals adopting the policy, the submission guidelines were changed to include a "checklist for anonymising your manuscript for double-anonymous peer review" and an optional Word template for double-anonymous submissions (IOP Publishing 2022). However, the guidelines also included the following text: *"On journals currently operating double-anonymous peer review* ***you may include author identifying information on your manuscript****, but please be aware that we do not edit manuscripts before sending them out for review, therefore* ***you include author information at your own risk and accept that this will be visible to reviewers****"* (emphasis added) (IOP Publishing 2022). The policy had a staggered rollout over the course of 12 months starting from November 2020 and finishing up in December 2021. Each month during that period a specific set of journals on related topics from the IOP's entire journal portfolio was transitioned into the new policy.

Data sources and sample

Our data include 156,015 submissions received by the IOP over a period from January 2018 up until February 2022. Our submissions range across 57 journals in the broader field of Physics, including manuscripts on materials sciences, theoretical physics, nanotechnology, biomedical research, and ecology and environmental sciences. All journals use a similar peer review process. Editors are the first evaluators who receive submitted manuscripts and make a decision about whether to send them for review or simply reject them at the "desk". Once decided to process a manuscript for peer review, editors send out invitations to potential experts in the field to act as peer referees. Reviewers receive fully anonymized invitations (in both pre- and post-policy time periods) with instructions similar across journals, and all journals use a standard commercial editorial software. In our data, 63,154 submissions were desk-rejected and 89,418 went through the peer review process. A manuscript can be reviewed by several referees and over several evaluation rounds. Our data provide detailed information on peer review recommendations for all manuscripts that were sent out for external review.

At the individual level, our data consists of information on the authors of each manuscript, including their names, institutional affiliations, optionally reported gender, and ORCID IDs. On average, each submission has four authors, with a total of about 390,000 authors present in our panel. Similarly, we are able to observe reviewer-identifying information for the entire reviewer sample of over 168,000 scientists, including both the ones who were invited but did not respond or agree and the ones who agreed to perform the evaluation.

For both authors and reviewers we matched these person-specific data with bibliometric data from *OpenAlex* (accessed in April 2022)—one of the largest and most complete open catalogs with scientific documents and researcher profiles (Singh Chawla 2022). We used ORCID IDs as well as exact matches for the "researcher name and institution" database searches. We were able to collect publication and citation data for 54% of all lead authors and 33% of reviewers in our sample. In the *OpenAlex* data, 88% of records for lead authors and 83% of records for reviewers came from using self-provided ORCID information.

Measures

We operationalize our treatment variable *new policy* as an indicator that takes a value of 1 on months when the anonymization policy became available to authors submitting to a particular journal, and 0 on all pre-policy months in that journal.

We measure authors' *anonymization* compliance as a binary variable that takes a value of 1 if a paper that they are submitting was anonymized, and 0 otherwise.

We use three dependent variables to capture evaluation decisions at various stages of the paper review process. For each submission we computed a binary variable *passed the desk* that takes a value of 1 if a manuscript was sent out for external review, and 0 if it was desk-rejected by the editor. We excluded 1,329 withdrawn by authors, 968 accepted without review and 1,146 unidentified papers from our analysis. 57% of manuscripts out of all submissions in our sample moved to a review stage.

For each submission that passed the editor's "desk", we computed a binary *reviewer recommendation* variable for the first round of reviews the manuscript received. We coded this variable as 0 if the reviewer decision was a rejection, and 1 if the decision was positive—i.e., offering a revision recommendation or acceptance. 71% of first reviewer reactions to the original submissions are positive in our data.

For all reviewed papers we also computed a *final decision* variable that was coded as 1 if the manuscript was eventually accepted for publication, and 0 otherwise. This final evaluation outcome captures both reviewer reactions as well as the editor's final decision. We excluded 791 submissions withdrawn by the authors and 6,390 submissions that were still under review at the end of our observation period. 60% of final evaluation outcomes in our sample are acceptances.

Lead author characteristics

Following previous literature (Sun, Barry Danfa, and Teplitskiy 2022; Teplitskiy et al. 2022), we defined author *prestige* based on their total number of citations (collected from the *OpenAlex* database) accumulated up to a manuscript's submission year. We were able to link the lead authors of 85,027 manuscripts to their *OpenAlex* records. As expected, the distribution of citations was very heavy-tailed. We divided all lead authors into three groups—low, middle, and high prestige—based on the total number of citations they have accumulated up to a manuscript's submission year. Low-prestige lead authors (n = 29,450) had no citations, middle-prestige authors (n = 32,079) had 1 to 100 citations, and high-prestige authors (n = 23,498) had over 100 citations.

Lead author's gender is an indicator measure coded as 0 for male and 1 for female lead authors. We enriched our self-reported gender data with the information collected using a gender classifying software *Ethnea* (Torvik and Agarwal 2016), which yielded us with the lead author's gender information for 124,591 manuscripts in our sample. Majority of lead authors are male with only 23% of submissions coming from females.

Lead author's country was inferred from their institutional affiliations. The lead author's country information was missing for 798 submissions. Over half of all submissions came from lead authors located in five countries: China (30%), India (11%), USA (8.9%), Iran (4.3%), and Germany (3%).

Reviewer characteristics

We also obtained *OpenAlex* data on two reviewer-related characteristics: the number of previous publications and citations accumulated up to a manuscript's submission year. These characteristics were used in testing the hypothesis that the policy affected selection into reviewing.

Results

Out of IOP's 57 journal portfolio, 49 journals switched to the new policy (accounting for 94% of the total submission volume), while 8 journals never adopted double-anonymization in the timeframe of our observation. In our sample, 129,917 submissions originated prior to policy change and 26,098 were submitted after the new policy went into effect at the respective journal. Among the sample of post-policy submissions, 5,723 papers were anonymized, yielding a double-anonymization uptake rate of 22%. Journals in the environmental and engineering subfields of physics had the largest rates of uptake (over 30%), while those in theoretical subfields had relatively low rates (below 10%).

Who chooses to anonymize?

First, we explore the take-up of anonymization. We estimated the likelihood of anonymization for various author prestige groups using a logistic regression model with anonymization (= 1 if anonymized, = 0 if not) as the outcome.

In our models, we control for the journal to which the manuscript was submitted to account for stable journal-specific factors such as research sub-domain or impact factor. We also control for time trends by including submission year-month indicator variables. Figure 1 displays the odds ratios between lead author characteristics and take-up of anonymization (corresponds to Table S1 Model 3).

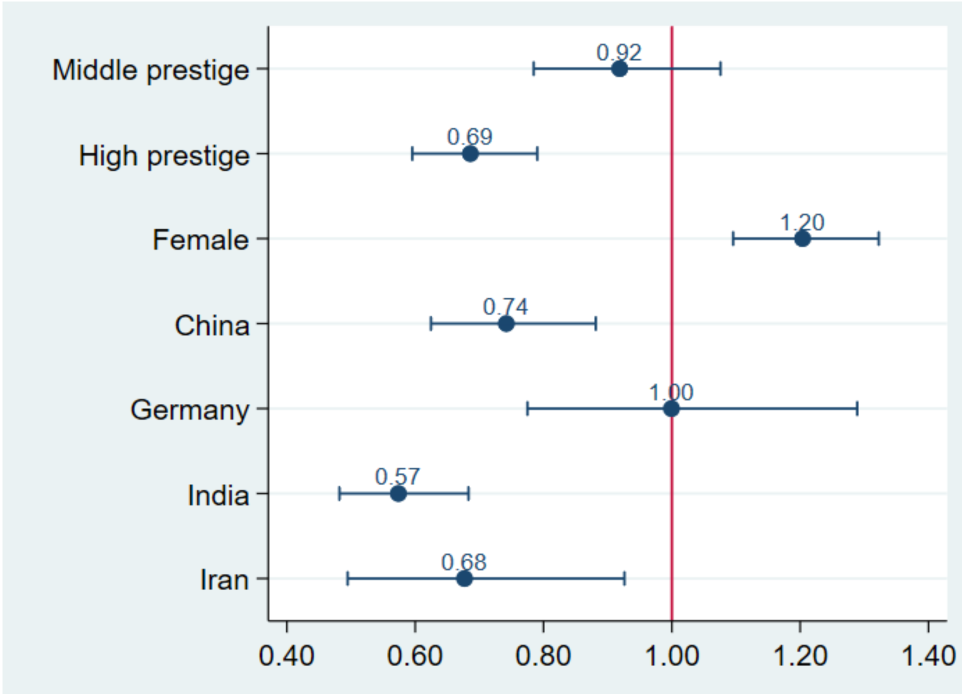

Figure 1. *Estimated logistic regression coefficients (odds ratios) with 95% CIs for the relationship between lead author characteristics and their likelihood of anonymizing the manuscript. Low prestige group is a reference point. USA is a baseline for country-level comparisons, only coefficients from the top submitting countries are presented in this graph.*

Controlling for the covariates, we see that the odds of high-prestige authors (with over 100 citations) to anonymize are 31 percentage points lower (p<0.001) relative to lowest prestige ones (no citations). Authors from the middle prestige category (1-100 citations) anonymize at a statistically similar rate to the lowest group. These findings corroborate previous research reporting that famous scientists often strategically reveal their identity before or during the review process (Guo, Xin, and Barnes 2022). We also find that female first authors are 20% more likely (p<0.001) to anonymize their papers compared to male authors, which is in line with broader conclusions from research on gender bias in evaluations (e.g., Goldin and Rouse 2000). In contrast, take-up of anonymization by country does not follow a simple "core-periphery" prediction that those from core countries (i.e., USA or Germany), who may benefit most from visible identities, anonymize less. The unexpected pattern suggests that additional factors play key roles in the decision.

The causal effect of the anonymization-via-nudging on evaluations

We modeled the effects of the new anonymization nudge policy on the outcomes of various stages of the review process using a difference-in-differences approach. We used linear probability models with the following general specification:

$$Pr(y_i = 1) = \beta_1 * \ new\,policy_j \ + \beta_2 * \ author\,prestige_i + \beta_3 * \ new\,policy_j * \\ * \ author\,prestige_i + \beta_4 * X_i \ + \beta_5 * Y_i + \ \alpha_j \ + \gamma_k + \epsilon_{ijk} \ (1)$$

where $y_i = \{0, 1\}$ is an indicator for the outcome for manuscript i, with 1 denoting success (e.g., passing the editor's desk, positive review, or acceptance). $New\,policy_j$ is an indicator that equals 1 after the journal j has adopted an unenforced double-anonymization review policy. IOP journals introduced the new policy on different dates during the November 2020-December 2021 period—for example, as the new policy became available at *Biomedical Materials* in November 2020, the $new\,policy_{Biomedical\,Materials}$ became 1 on all subsequent submission months. $Author\,prestige_i$ denotes the several indicators that equal 1 when the first author's citations up to submission fall into one of three groups—low, middle, or high (see Table S3 for the analogous analysis based on alternative citation binning and Table S4 for analysis based on corresponding authors). We include journal-level fixed effects α_j to control for time-invariant stable journal-specific factors that affect acceptance and year-month dummies γ_k to capture any time-related differences. We also control for the lead author's gender and affiliation country with the vectors X_i and Y_i, respectively.

We disaggregate the review process into three stages: desk review, external peer review (if not desk-rejected), and final decision stage. We ran separate regressions of the form (1) for each of the stages.

Desk review (placebo test). We start our analysis from the premise that double-anonymization should not affect the desk-review process since in both single- and double-anonymous conditions editors see author identities. Consequently, we estimate the effects of the policy on desk rejection as a placebo test for our theoretical claims.

Figure 2 displays average marginal effects of the new policy on desk review decisions: panel A presents the disaggregated marginal effects pre- and post-policy for different author prestige levels, and panel B illustrates the predicted differences that new policy adoption brought to various author groups. The full regression is presented in Table S2 (Models 1-2).

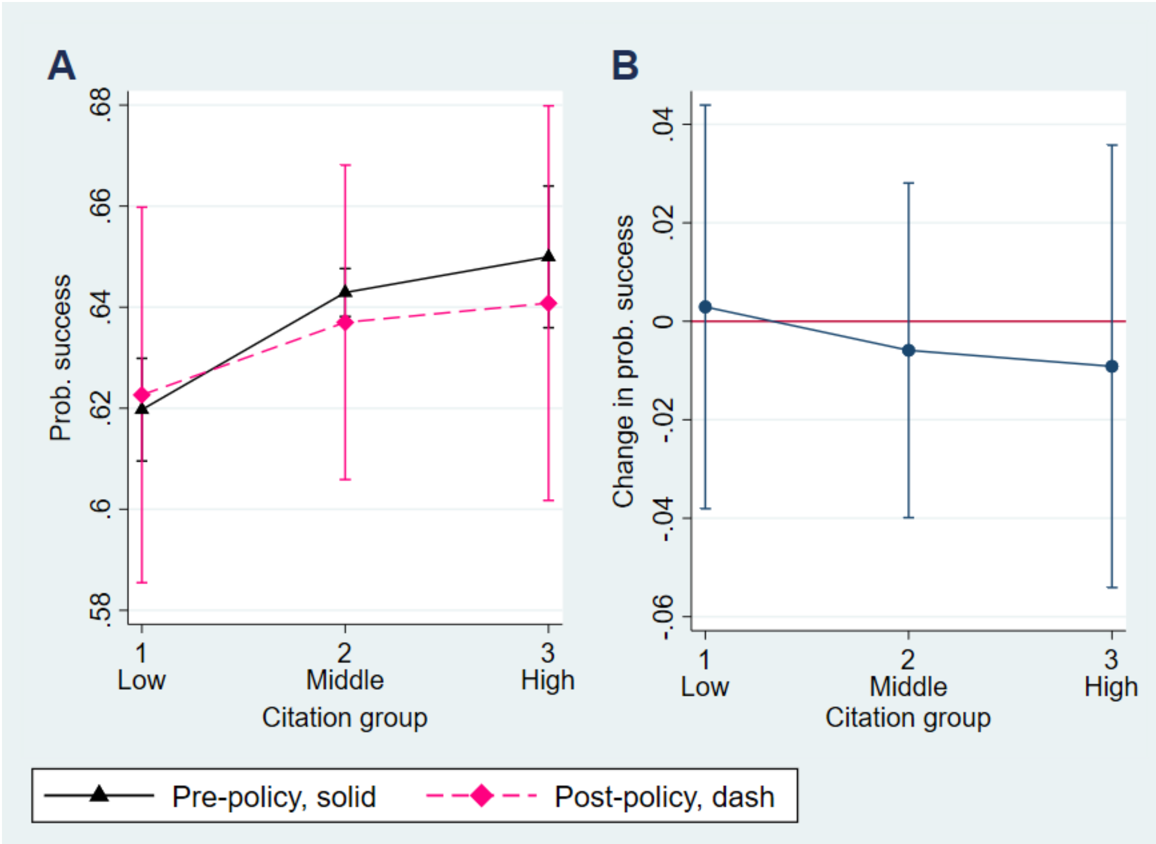

Figure 2. *Predicted effects of the new policy on desk review outcome for various lead author prestige groups.* ***A:*** *Marginal effects.* ***B:*** *Difference in evaluation post- vs. pre-policy. Error bars represent 95% CIs.*

The estimates show that the new policy does not have a statistically significant effect on desk rejection for all authors: although the point estimates for the effect on all three author prestige groups are negative, they are small (< 2%) and statistically close to 0 (p>0.10 in each case, Fig. 2B). These results are consistent with the prediction of no effect.

Peer review. Figure 3 displays estimates from a regression where the outcome variable is an indicator for receiving a positive reviewer recommendation in the first round of reviews, defined as {0 = *Reject*, 1 = *Accept, Minor revision*, or *Major revision*}, and standard errors are clustered at both the journal and the submission levels (Table S2 Models 3-4).

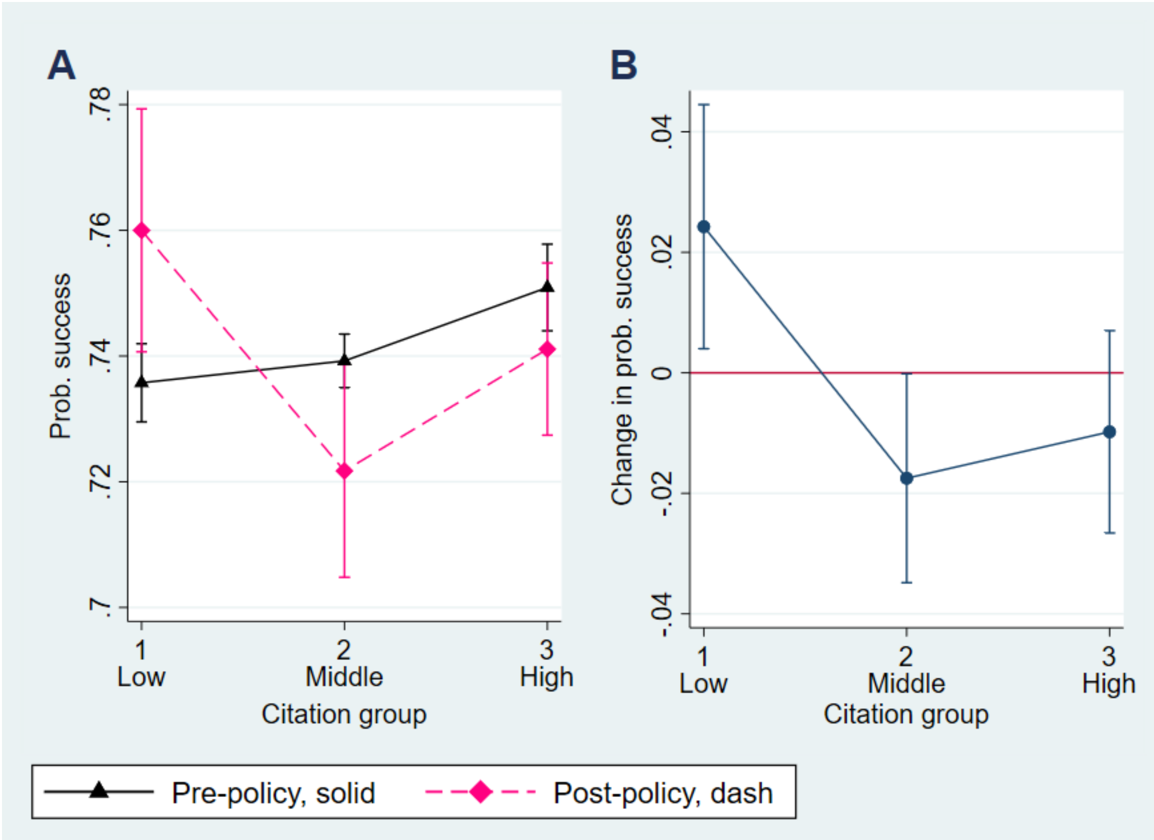

Figure 3. *Predicted effects of the new policy on peer review outcome for various lead author prestige groups.* ***A:*** *Marginal effects.* ***B:*** *Difference in evaluation post- vs. pre-policy. Error bars represent 95% CIs.*

The policy increased the probability that low-prestige authors receive a positive review by 2.4 percentage points (p<0.05, Fig. 3B). In contrast, the policy decreased the probability that middle- and high-prestige authors receive positive reviews by 1.8% (p = 0.048) and 1% (p = 0.253). The point-estimates for all three status groups are directionally consistent with expectations but imprecisely estimated given the recency of the policy.

We thus conclude that the key prediction of the anonymization nudge is strongly (but not conclusively) supported: if anonymized recommendations represent a relatively unbiased signal, then the policy helped reduce the bias in non-anonymized evaluations, which more highly cited authors benefited from before the policy. This effect is observed despite more cited authors anonymizing their papers less.

Final review. Figure 4 displays estimates from a regression where the outcome is an indicator for whether the paper was ultimately accepted for publication (Table S2 Models 5-6). The models were estimated on the subset of submissions that were sent out for peer review.

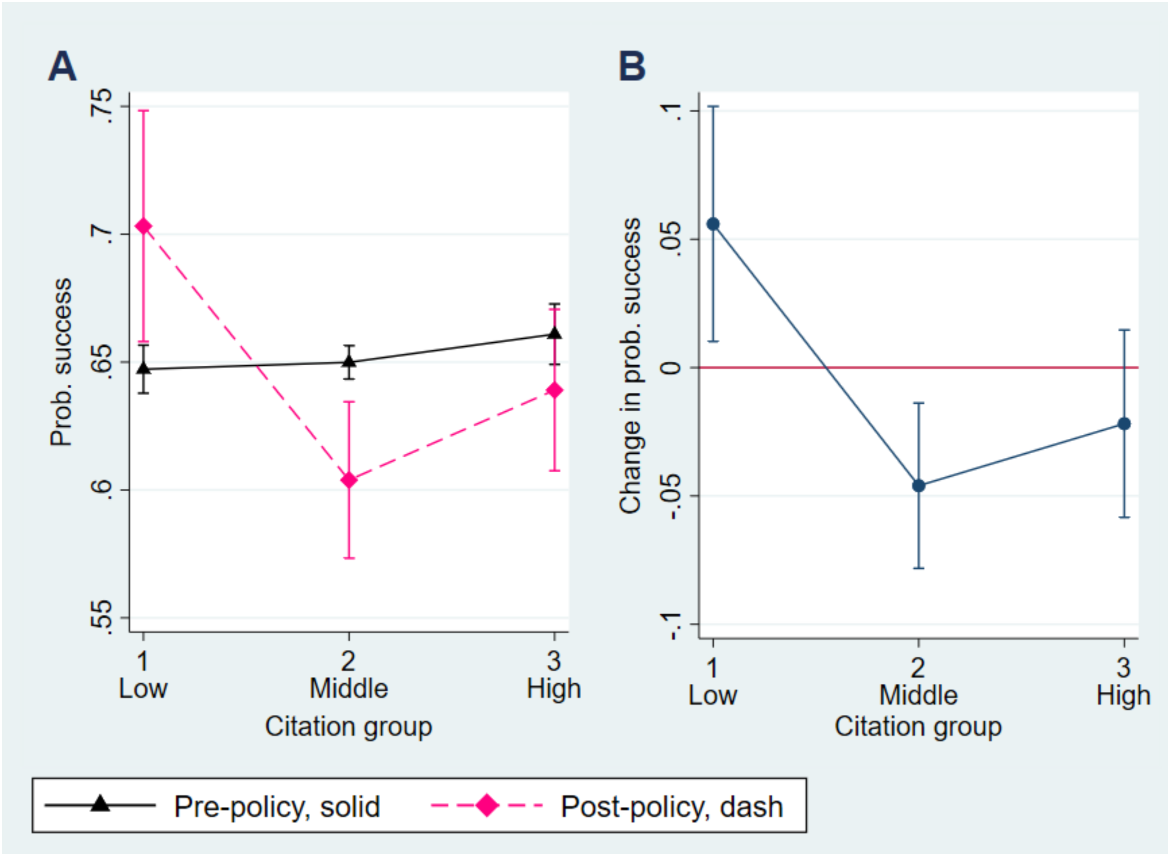

Figure 4. *Predicted effects of the new policy on final review outcome for various lead author prestige groups.* ***A:*** *Marginal effects.* ***B:*** *Difference in evaluation post- vs. pre-policy. Error bars represent 95% CIs.*

The differences in acceptances are even more pronounced than for reviewer recommendations. The estimated marginal effect of the policy on low-citations authors is 5.6% (Fig. 4B, p<0.05). In contrast, point estimates show that the policy caused the acceptance probability for both middle- and high-prestige authors to decrease by 4.6% (p = 0.005) and 2.2% (p = 0.241). The point-estimates for all three status groups are in line with our theoretical expectations but less robust for the highest prestige group.

Two mechanisms may account for why policy effects at this final stage are larger than at the peer review stage. First, in our analyses for the peer-review stage we focus on the first review rounds only whereas the final decision is based on potentially additional rounds of review. To test this mechanism, we ran an additional set of models with the last round of reviewer recommendations as the dependent variable in Models 7-8 of Table S2. The results are almost identical across first and last review rounds, so the differences are unlikely to be explained by additional review rounds.

Second, the editor's final decision may be driven by the most negative reviewer recommendation received, rather than their average. If so, a substantial change in just one review could lead to a small change in average reviews, but a substantial change in the final decision. To test this possibility, we ran models taking the lowest recommendation in the last round of reviews as the dependent variable in Models 9-10 of Table S2. We found that the results are very close to the ones in the final review stage, indicating that this explanation is likely driving the differences we observed between the estimates reported in Figures 3 and 4.

We also ran robustness tests, using models with alternative binning of status (Table S3) and focusing on corresponding rather than lead authors (Table S4). The results of these regressions are qualitatively similar to our main findings, although statistically weaker for certain coefficients.

Having established the causal effects of the policy on review decisions, we turn to identifying the exact mechanisms driving those effects.

Mechanisms

Anonymization

Our main hypothesis is that low-cited authors benefit under the new policy because they can anonymize their manuscripts and avoid bias against them, while highly-cited authors are harmed by it because some of them choose to be "good citizens," anonymize their manuscripts, and forego the prestige advantage. However, our estimates above are of the effects of the nudging policy, not anonymization directly. To estimate the latter, we used an instrumental variable regression approach. This approach assumes that the nudging policy operates through anonymization only, i.e., not by changing selection into reviewing or submitting, and estimates the effect size of *anonymization* (not the nudge to anonymize). We instrumented the anonymization indicator by the availability of the new policy and computed the local average treatment effect, i.e., the effect of anonymization on those lead authors who anonymized their papers. Table S5 reports the estimation results with the two-stage least squares for our three evaluation outcomes while Figure 5 presents the results in a graphical form.

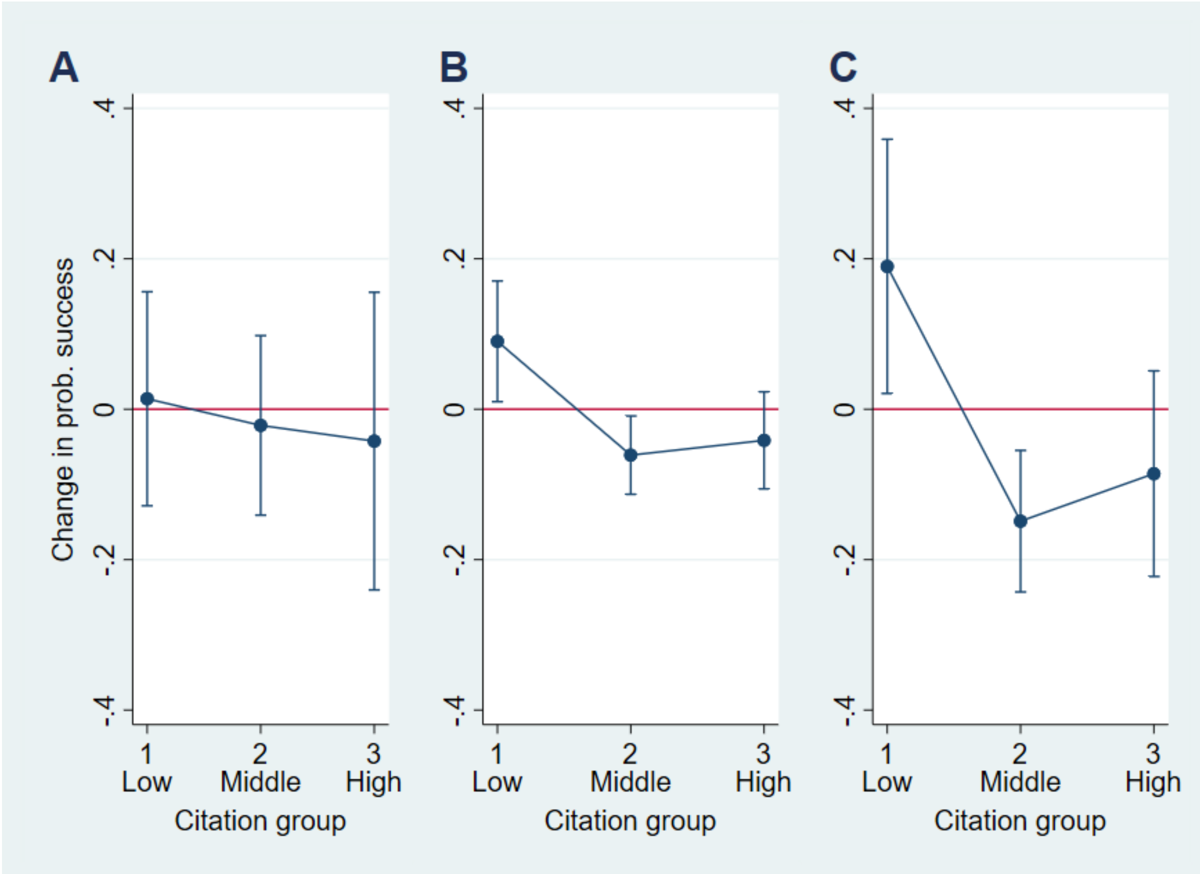

Figure 5. *Predicted evaluation differences between non-anonymized and anonymized manuscripts for various lead author prestige groups.* ***A:*** *Desk review.* ***B:*** *Peer review.* ***C:*** *Final review. Error bars represent 95% CIs.*

Figure 5A shows that for the desk-review outcome there are no statistically significant effects of anonymization on desk decisions for all three status groups, which is in line with our main results. The results for the peer-review and final-review outcomes, illustrated in Fig. 5B and 5C, corroborate our main findings. Low-prestige authors benefit from anonymizing their papers—their chances of success for a positive reviewer's recommendation and final decision are increased by 9 (p<0.05 in Fig. 5B) and 19 percentage points (p<0.05 in Fig. 5C), respectively. Middle-prestige authors, on the other hand, receive lower evaluation outcomes when their identities are hidden from reviewers—6.1% (p<0.05 in Fig. 5B) lower success chances for a positive reviewer's recommendation and 14.9% (p<0.01 in Fig. 5C) for the final paper acceptance, respectively. The results for high-prestige authors are qualitatively similar to the ones for middle-status authors (4.1% and 8.6% decrease in probability to receive positive reviewer and final recommendation, respectively), but not statistically significant (p>0.2 in both Fig. 5B and 5C). The effect sizes in these models are expectedly larger than the ones previously reported in Table S2 as these are the actual treatments rather than the so-called intention-to-treat effects.

The policy may have had effects through additional, non-anonymization mechanisms, which we consider below.

Author recruitment

One possible effect of the policy is to systematically affect the types of authors who submit to IOP journals. To account for this possibility, we estimated two sets of models with author publication and citation data as our two dependent variables and the availability of the new policy as a regressor in both sets. Keeping covariates constant, the results in Table S6 show no systematic author differences pre- vs. post-policy. We thus reject this hypothesis.

Reviewer recruitment

The policy may also have affected the types and quality of reviewers recruited before and after the new policy. To test this mechanism, we estimated two sets of models[2]: one with a binary dependent variable for each invited reviewer defined as 1 if the reviewer accepted an editor's invitation to review a given manuscript, and 0 otherwise (Table S7 Models 1-2); and the second set with reviewer publication and citation data as our two response variables (Table S7 Models 3-6). We used the availability of the new policy and lead authors' characteristics as our main predictors in both sets. The data suggest no visible patterns of review selection bias (as measured by average characteristics) in the pre- vs. post-policy observational period—the estimated relationship is not statistically significant in all regression models. We also do not find any association between the types of reviewers that different author groups attract before and after the policy. Together, we thus conclude that it is unlikely that the reviewer selection mechanism is driving our results and that the new policy solicited a substantially different reviewer crowd.

Finally, for robustness, in Table S8 we present the results for the effects of the new policy on first reviewer recommendation controlling for reviewer fixed-effects. This specification tests for behavior change among the same set of reviewers before and after the policy. Our main results partially hold, although the sample size is much smaller and the standard errors are larger. We find that, after the new policy, keeping time-invariant reviewer traits intact, the results match our previous findings for low-prestige authors but show weaker significance for the middle and top author groups.

---

[2] Given raw data's limitation to provide us with unique reviewer IDs, in these models we only focus on the reviewer sample who were successfully identified in the *OpenAlex* database search.

Discussion

In this paper, we analyzed paper submissions to 57 peer-reviewed physics journals in the IOP's portfolio to examine the bias-reducing effect of voluntary anonymization. The sample is unusually large-scale and naturalistic, with high stakes, involving over 390,000 submitters and 168,000 reviewers from around the globe. Our inquiry makes several contributions to understanding the effects of voluntary anonymization in evaluations. First, we examined which authors comply with anonymization when nudged (RQ1). We found evidence of self-selection: in line with our prediction about strategic behavior, high-prestige authors anonymize less. The decisions to anonymize are not fully strategic, however, and some high-prestige authors do indeed choose to hide their identities from reviewers. Take-up of anonymization by country revealed that information frictions can also account for authors' decisions to anonymize, thus improving our understanding of anticipated discrimination. We found that authors from core countries like USA and Germany, who may benefit most from visible identities, anonymize significantly more vs. authors from China and India, who may experience discrimination but do not seem to anticipate it.

Second, we estimated the causal effect of the anonymization-via-nudging policy in helping reduce prestige bias in scientific evaluations (RQ2). We examined paper evaluation outcomes at the three main review stages: desk, external peer review, and final editor review. First, in a placebo test, the new policy did not have an effect on desk-review outcomes, which is expected given that the policy did not affect whether editors see authors' identities. Second, the policy increased the likelihood of positive reviewer recommendations for low-prestige authors by 2.4% and lowered it for middle- and high-citations authors by 1.8% and 1%, respectively. Lastly, the policy had the biggest effects on reducing prestige bias on final paper decisions, increasing acceptance of low-prestige authors by 5.6% while lowering it by 4.6% and 2.2% for middle- and high-citations authors, respectively. Because authors' decisions to anonymize are not fully strategic (per our RQ1), this prevents anonymization from being a reliable signal of low status, making bias reduction of the new policy possible. It is important to note that while our bias-reducing results for low- and middle-status authors are statistically significant, the finding for the top author group is not. This may be due to substantive and statistical reasons. Given that high-status authors are least likely to anonymize, the nudging policy affected their behavior the least, and reductions in bias are correspondingly the smallest. Additionally, the policy is relatively recent, so the amount of post-policy observations is limited. More post-policy data should make the estimates more robust.

These unique data enabled us to directly test some of the unintended effects (unrelated to bias reduction) that the policy may have produced: the policy could change the quantity or characteristics of authors (RQ3) or reviewers (RQ4) who submit or review papers. We found no evidence that the policy affected the characteristics of the submitting authors. Similarly, we found no evidence that the policy hurt reviewer recruitment, i.e., characteristics of those accepting invitations to review, or the characteristics of those submitting reviews. Furthermore, our sizable treatment effects make it unlikely that reviewers interpreted authors' choosing to anonymize as a signal of low prestige and discriminated based on this signal. Evidence against these additional effects of the policy buttresses the focal mechanism—prestige bias reduction among reviewers—as the primary effect of the nudge policy.

Comparing the magnitude of our bias reduction via nudging to studies of peer review with enforced anonymization (Blank 1991; Okike et al. 2016; Tomkins, Zhang, and Heavlin 2017; Sun, Barry Danfa, and Teplitskiy 2022) shows that our estimates are generally smaller, suggesting that enforcement is necessary to achieve the full benefits of anonymization. However, the policy costs should also be considered given that "nudging" delivers some of the benefits of harder-to-implement policies at essentially no cost. With a proper cost-benefit analysis, the voluntary anonymization option might prove the superior solution for many organizations seeking to make their expert-based evaluations fairer.

Although we examined a rich number of evaluations, the setting is only in one domain (physical sciences) of one industry (academic science). This raises the question of external validity. There are strong reasons to expect the results to generalize to other industries relying on qualitative evaluations by experts. Goldin and Rouse (2000), for example, examined the setting of orchestral auditions and found that, similarly to scientific peer review (Fox and Paine 2019), blinding the musicians' identities increased the chances of female musicians to get selected to the advanced audition stages (by 50%) and eventually get hired (explaining 30% increase in the proportion of females among new hires). The similarities in procedures and findings highlight the notion that biases in evaluations arise from fundamental human features rather than specific settings where a particular study is conducted. Similarly, the field of Physics is a very conservative setting to estimate our effects due to the high popularity and adoption rates of preprints. In the "age of Google", the problem of simple searches may be exacerbated in the physical sciences, thus limiting the effectiveness of anonymization. Our results can thus be in fact smaller than in other industries dealing with complex qualitative judgments and that are of sufficient size that evaluators do not know all projects before entering the formal evaluation process.

Our study is subject to several scope conditions. The generalizability of our findings may not extend to fields with substantially different evaluation cultures or sizes. For example, anonymization, whether via nudging or enforcement, may not work well in very small fields because reviewers can accurately identify authors of anonymized projects (O'Connor et al. 2017). Additionally, our study focuses on prestige bias only at the last stages of project pipelines. Prestige bias may also affect various upstream parts of the pipelines, including problem choice (Hoppe et al. 2019) and the sorting of individuals into institutions (Clauset, Arbesman, and Larremore 2015). Lastly, the analyses used only individuals whom we could successfully link to publications data, leaving open the possibility that effect sizes might change had it been possible to analyze the entire sample of submitters and reviewers.

Our study raises a number of questions for future research. First, due to the large variation in anonymization by subfield (i.e., theoretical vs. environmental physics projects), our results are driven primarily by journals with the higher rates of policy take-up. Future investigations may assess how prestige bias varies across field topics, sizes, and journal/venue prestige. Second, given that the timeframe of our data is rather recent, future research may explore the longer-term outcomes of the nudge policy, such as whether it can improve the quality of the selections, breadth and novelty of the topics selected, or project submitters' and evaluators' careers. Third, an alternative explanation not ruled out by our design is that evaluators favor submissions that comply with anonymization simply due to their compliance, creating a new bias.

The attractiveness of the anonymization nudge may also turn it into a double-edged sword. On the one hand, performing policy experiments like IOP's and publicizing them makes it easier for other organizations to follow suit, possibly leading to wider adoption of costless nudges. On the other hand, our results suggest that, from a purely strategic perspective, prestigious individuals should not anonymize their projects if they would like to keep their reputational premium. In the presence of information frictions, however, authors should accurately assess their prestige in several hierarchies before making an anonymization decision. Consequently, increased attention to this topic may reduce information frictions and lead famous individuals to increase strategizing when submitting their projects, thus potentially limiting the policy's benefits. Finally, nudging can potentially lead to greater gender disparities. We found that female submitters tend to anonymize their projects more often. If anonymization hurts higher-prestige scientists, nudging may disproportionately hurt higher-prestige women. Future work should examine how potentially intersectional effects of the policy can be effectively managed to reduce prestige bias without increasing other biases.

References

Agan, Amanda, and Sonja Starr. 2018. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment*." *The Quarterly Journal of Economics* 133 (1): 191–235. https://doi.org/10.1093/qje/qjx028.

Anderson, Rick. 2020. "IOP Moves to Universal Double-Blind Peer Review: An Interview with Kim Eggleton." *The Scholarly Kitchen* (blog). September 10, 2020. https://scholarlykitchen.sspnet.org/2020/09/10/iop-moves-to-universal-double-blind-peer-review-

an-interview-with-kim-eggleton/.
Bacharach, Samuel B., Peter Bamberger, and Bryan Mundell. 1993. “Status Inconsistency in Organizations: From Social Hierarchy to Stress.” *Journal of Organizational Behavior* 14 (1): 21–36. https://doi.org/10.1002/job.4030140104.
Beshears, John, and Harry Kosowsky. 2020. “Nudging: Progress to Date and Future Directions.” *Organizational Behavior and Human Decision Processes* 161 (November): 3–19. https://doi.org/10.1016/j.obhdp.2020.09.001.
Blank, Rebecca M. 1991. “The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The American Economic Review.” *The American Economic Review* 81 (5): 1041–67.
Ceci, Stephen J., and Wendy M. Williams. 2011. “Understanding Current Causes of Women’s Underrepresentation in Science.” *Proceedings of the National Academy of Sciences* 108 (8): 3157–62. https://doi.org/10.1073/pnas.1014871108.
Chen, Christine Yifeng, Sara S Kahanamoku, Aradhna Tripati, Rosanna A Alegado, Vernon R Morris, Karen Andrade, and Justin Hosbey. 2022. “Systemic Racial Disparities in Funding Rates at the National Science Foundation.” *ELife* 11 (November): e83071. https://doi.org/10.7554/eLife.83071.
Chong, Phillipa K. 2020. “Inside the Critics’ Circle: Book Reviewing in Uncertain Times.” *Social Forces* 99 (1): e17–e17. https://doi.org/10.1093/sf/soaa024.
Clauset, Aaron, Samuel Arbesman, and Daniel B. Larremore. 2015. “Systematic Inequality and Hierarchy in Faculty Hiring Networks.” *Science Advances* 1 (1): e1400005. https://doi.org/10.1126/sciadv.1400005.
Coffman, Katherine Baldiga. 2014. “Evidence on Self-Stereotyping and the Contribution of Ideas*.” *The Quarterly Journal of Economics* 129 (4): 1625–60. https://doi.org/10.1093/qje/qju023.
Criscuolo, Paola, Linus Dahlander, Thorsten Grohsjean, and Ammon Salter. 2017. “Evaluating Novelty: The Role of Panels in the Selection of R&D Projects.” *Academy of Management Journal* 60 (2): 433–60. https://doi.org/10.5465/amj.2014.0861.
Edmondson, Amy C., and Ranjay Gulati. 2021. “Agility Hacks How to Create Temporary Teams That Can Bypass Bureaucracy and Get Crucial Work Done Quickly.” *Harvard Business Review* 99 (6): 46–49.
Ernst, E., and T. Kienbacher. 1991. “Chauvinism.” *Nature* 352 (6336): 560–560. https://doi.org/10.1038/352560b0.
Fischman, Josh. 2011. “Leading Economics Journals Drop ‘Double Blind’ Peer Review.” *The Chronicle of Higher Education* (blog). May 31, 2011. https://www.chronicle.com/blogs/ticker/leading-economics-journals-drop-double-blind-peer-review.
Fox, Charles W., and C. E. Timothy Paine. 2019. “Gender Differences in Peer Review Outcomes and Manuscript Impact at Six Journals of Ecology and Evolution.” *Ecology and Evolution* 9 (6): 3599–3619. https://doi.org/10.1002/ece3.4993.
Frank, Robert H. 2008. “To Disclose or Not? Ask the Frogs,” April 13, 2008. https://www.nytimes.com/2008/04/13/business/13view.html.
Goldin, Claudia, and Cecilia Rouse. 2000. “Orchestrating Impartiality: The Impact of ‘Blind’ Auditions on Female Musicians.” *American Economic Review* 90 (4): 715–41. https://doi.org/10.1257/aer.90.4.715.
Goues, C. Le, Y. Brun, S. Apel, E. Berger, S. Khurshid, and Y. Smaragdakis. 2018. “Effectiveness of Anonymization in Double-Blind Review.” *Communications of the ACM* 61 (6): 30–33. https://doi.org/10.1145/3208157.
Guo, Yue, Fu Xin, and Stuart J. Barnes. 2022. “The Fiction of Double-Blind Reviewing: Evidence From the Social Science Research Network.” *International Journal of Business Communication* 59 (1): 48–55. https://doi.org/10.1177/2329488418803655.
Han, Jung-Hoon, and Timothy G. Pollock. 2021. “The Two Towers (or Somewhere in Between): The

Behavioral Consequences of Positional Inconsistency across Status Hierarchies." *Academy of Management Journal* 64 (1): 86–113. https://doi.org/10.5465/amj.2018.1091.
Harris, Matthew, Joachim Marti, Hillary Watt, Yasser Bhatti, James Macinko, and Ara W. Darzi. 2017. "Explicit Bias Toward High-Income-Country Research: A Randomized, Blinded, Crossover Experiment Of English Clinicians." *Health Affairs* 36 (11): 1997–2004. https://doi.org/10.1377/hlthaff.2017.0773.
Hertwig, Ralph, and Till Grüne-Yanoff. 2017. "Nudging and Boosting: Steering or Empowering Good Decisions." *Perspectives on Psychological Science* 12 (6): 973–86. https://doi.org/10.1177/1745691617702496.
Ho, Roger Chun-Man, Kwok-Kei Mak, Ren Tao, Yanxia Lu, Jeffrey R Day, and Fang Pan. 2013. "Views on the Peer Review System of Biomedical Journals: An Online Survey of Academics from High-Ranking Universities." *BMC Medical Research Methodology* 13 (1): 74. https://doi.org/10.1186/1471-2288-13-74.
Hoppe, Travis A., Aviva Litovitz, Kristine A. Willis, Rebecca A. Meseroll, Matthew J. Perkins, B. Ian Hutchins, Alison F. Davis, et al. 2019. "Topic Choice Contributes to the Lower Rate of NIH Awards to African-American/Black Scientists." *Science Advances* 5 (10): eaaw7238. https://doi.org/10.1126/sciadv.aaw7238.
Huber, Jürgen, Sabiou Inoua, Rudolf Kerschbamer, Christian König-Kersting, Stefan Palan, and Vernon L. Smith. 2022. "Nobel and Novice: Author Prominence Affects Peer Review." *Proceedings of the National Academy of Sciences* 119 (41): e2205779119. https://doi.org/10.1073/pnas.2205779119.
IOP Publishing. 2022. "Checklist for Anonymising Your Manuscript for Double-Anonymous Peer Review." IOP Publishing Website. May 2022. https://publishingsupport.iopscience.iop.org/questions/checklist-for-anonymising-your-manuscript/.
Jamali, Hamid R., David Nicholas, Anthony Watkinson, Abdullah Abrizah, Blanca Rodríguez‐Bravo, Cherifa Boukacem‐Zeghmouri, Jie Xu, Tatiana Polezhaeva, Eti Herman, and Marzena Świgon. 2020. "Early Career Researchers and Their Authorship and Peer Review Beliefs and Practices: An International Study." *Learned Publishing* 33 (2): 142–52. https://doi.org/10.1002/leap.1283.
Kang, Sonia K., Katherine A. DeCelles, András Tilcsik, and Sora Jun. 2016. "Whitened Résumés: Race and Self-Presentation in the Labor Market." *Administrative Science Quarterly* 61 (3): 469–502. https://doi.org/10.1177/0001839216639577.
Larivière, Vincent, Stefanie Haustein, and Philippe Mongeon. 2015. "The Oligopoly of Academic Publishers in the Digital Era." Edited by Wolfgang Glanzel. *PLOS ONE* 10 (6): e0127502. https://doi.org/10.1371/journal.pone.0127502.
Lenski, Gerhard E. 1967. "Status Inconsistency and the Vote: A Four Nation Test." *American Sociological Review* 32 (2): 298. https://doi.org/10.2307/2091819.
Li, Danielle. 2017. "Expertise versus Bias in Evaluation: Evidence from the NIH." *American Economic Journal: Applied Economics* 9 (2): 60–92. https://doi.org/10.1257/app.20150421.
McGillivray, Barbara, and Elisa De Ranieri. 2018. "Uptake and Outcome of Manuscripts in Nature Journals by Review Model and Author Characteristics." *Research Integrity and Peer Review* 3 (1): 5. https://doi.org/10.1186/s41073-018-0049-z.
Mertens, Stephanie, Mario Herberz, Ulf J. J. Hahnel, and Tobias Brosch. 2022. "The Effectiveness of Nudging: A Meta-Analysis of Choice Architecture Interventions across Behavioral Domains." *Proceedings of the National Academy of Sciences* 119 (1): e2107346118. https://doi.org/10.1073/pnas.2107346118.
Merton, Robert K. 1968. "The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered." *Science* 159 (3810): 56–63. https://doi.org/10.1126/science.159.3810.56.
Mulligan, Adrian, Louise Hall, and Ellen Raphael. 2013. "Peer Review in a Changing World: An International Study Measuring the Attitudes of Researchers." *Journal of the American Society for*

Information Science and Technology 64 (1): 132–61. https://doi.org/10.1002/asi.22798.
Murray, Dakota, Kyle Siler, Vincent Larivière, Wei Mun Chan, Andrew M. Collings, Jennifer Raymond, and Cassidy R. Sugimoto. 2019. "Author-Reviewer Homophily in Peer Review." Preprint. BioRxiv. https://doi.org/10.1101/400515.
Nakamura, Richard K, Lee S Mann, Mark D Lindner, Jeremy Braithwaite, Mei-Ching Chen, Adrian Vancea, Noni Byrnes, Valerie Durrant, and Bruce Reed. 2021. "An Experimental Test of the Effects of Redacting Grant Applicant Identifiers on Peer Review Outcomes." *ELife* 10 (October): e71368. https://doi.org/10.7554/eLife.71368.
O'Connor, E.E., M. Cousar, J.A. Lentini, M. Castillo, K. Halm, and T.A. Zeffiro. 2017. "Efficacy of Double-Blind Peer Review in an Imaging Subspecialty Journal." *American Journal of Neuroradiology* 38 (2): 230–35. https://doi.org/10.3174/ajnr.A5017.
Okike, Kanu, Kevin T. Hug, Mininder S. Kocher, and Seth S. Leopold. 2016. "Single-Blind vs Double-Blind Peer Review in the Setting of Author Prestige." *JAMA* 316 (12): 1315. https://doi.org/10.1001/jama.2016.11014.
Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108 (5): 937–75. https://doi.org/10.1086/374403.
Quillian, Lincoln, Anthony Heath, Devah Pager, Arnfinn Midtbøen, Fenella Fleischmann, and Ole Hexel. 2019. "Do Some Countries Discriminate More than Others? Evidence from 97 Field Experiments of Racial Discrimination in Hiring." *Sociological Science* 6: 467–96. https://doi.org/10.15195/v6.a18.
Reimers, Imke, and Joel Waldfogel. 2021. "Digitization and Pre-Purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings." *American Economic Review* 111 (6): 1944–71. https://doi.org/10.1257/aer.20200153.
Reinstein, David A., and Christopher M. Snyder. 2005. "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics*." *The Journal of Industrial Economics* 53 (1): 27–51. https://doi.org/10.1111/j.0022-1821.2005.00244.x.
Rivera, Lauren A., and András Tilcsik. 2019. "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84 (2): 248–74. https://doi.org/10.1177/0003122419833601.
Sauder, Michael, Freda Lynn, and Joel M. Podolny. 2012. "Status: Insights from Organizational Sociology." *Annual Review of Sociology* 38 (1): 267–83. https://doi.org/10.1146/annurev-soc-071811-145503.
Sharkey, Amanda, Balazs Kovacs, and Greta Hsu. 2022. "Expert Critics, Rankings, and Review Aggregators: The Changing Nature of Intermediation and the Rise of Markets with Multiple Intermediaries." *Academy of Management Annals*, April, annals.2021.0025. https://doi.org/10.5465/annals.2021.0025.
Shukla, Arun Kumar. 2021. "India — Stop Looking down on International Collaborations." *Nature* 600 (7889): 361–361. https://doi.org/10.1038/d41586-021-03700-2.
Simcoe, Timothy S., and Dave M. Waguespack. 2011. "Status, Quality, and Attention: What's in a (Missing) Name?" *Management Science* 57 (2): 274–90. https://doi.org/10.1287/mnsc.1100.1270.
Singh Chawla, Dalmeet. 2022. "Massive Open Index of Scholarly Papers Launches." *Nature*, January, d41586-022-00138-y. https://doi.org/10.1038/d41586-022-00138-y.
Snodgrass, Richard. 2006. "Single- versus Double-Blind Reviewing: An Analysis of the Literature." *ACM SIGMOD Record* 35 (3): 8–21. https://doi.org/10.1145/1168092.1168094.
Stryker, Sheldon, and Anne Statham Macke. 1978. "Status Inconsistency and Role Conflict." *Annual Review of Sociology* 4: 57–90.
Subbaraman, Nidhi. 2021. "Scientists' Fears of Racial Bias Surge amid US Crackdown on China Ties." *Nature*, October, d41586-021-02976–78. https://doi.org/10.1038/d41586-021-02976-8.
Sun, Mengyi, Jainabou Barry Danfa, and Misha Teplitskiy. 2022. "Does Double‐blind Peer Review Reduce Bias? Evidence from a Top Computer Science Conference." *Journal of the Association for Information Science and Technology* 73 (6): 811–19. https://doi.org/10.1002/asi.24582.

Teplitskiy, Misha, Eamon Duede, Michael Menietti, and Karim R. Lakhani. 2022. “How Status of Research Papers Affects the Way They Are Read and Cited.” *Research Policy* 51 (4): 104484. https://doi.org/10.1016/j.respol.2022.104484.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Penguin Group, New York.

Tomkins, Andrew, Min Zhang, and William D. Heavlin. 2017. “Reviewer Bias in Single- versus Double-Blind Peer Review.” *Proceedings of the National Academy of Sciences* 114 (48): 12708–13. https://doi.org/10.1073/pnas.1707323114.

Torvik, Vetle I., and Sneha Agarwal. 2016. “Ethnea -- an Instance-Based Ethnicity Classifier Based on Geo-Coded Author Names in a Large-Scale Bibliographic Database.” Presented at the International Symposium on Science of Science, Library of Congress, Washington DC, USA, March 22. https://www.ideals.illinois.edu/bitstream/handle/2142/88927/ethnea.pdf.

Ucci, Matteo Antonio, Francesco D’Antonio, and Vincenzo Berghella. 2022. “Double- vs Single-Blind Peer Review Effect on Acceptance Rates: A Systematic Review and Meta-Analysis of Randomized Trials.” *American Journal of Obstetrics & Gynecology MFM* 4 (4): 100645. https://doi.org/10.1016/j.ajogmf.2022.100645.

Vollaard, Ben, and Jan C. van Ours. 2022. “Bias in Expert Product Reviews.” *Journal of Economic Behavior & Organization* 202 (October): 105–18. https://doi.org/10.1016/j.jebo.2022.08.002.

Wiley. 2022. “Types of Peer Review.” Wiley Website. June 2022. https://authorservices.wiley.com/Reviewers/journal-reviewers/what-is-peer-review/types-of-peer-review.html.

Supplementary Appendix for "The bias-reducing effect of voluntary anonymization of authors' identities: Evidence from peer review"

Table S1. Anonymization compliance (logistic regressions—odds ratios)

VARIABLES	(1) DV: likelihood of anonymizing	(2) DV: likelihood of anonymizing	(3) DV: likelihood of anonymizing
Middle prestige	0.96	0.96	0.92
	(0.073)	(0.075)	(0.074)
High prestige	0.80**	0.79**	0.69***
	(0.060)	(0.058)	(0.050)
Female		1.19***	1.20***
		(0.057)	(0.058)
Country: China			0.74***
			(0.065)
Country: Germany			1.00
			(0.130)
Country: India			0.57***
			(0.051)
Country: Iran			0.68*
			(0.108)
Constant	0.25***	0.25***	0.30**
	(0.096)	(0.095)	(0.118)
Observations	11,150	10,637	10,329
All country dummies	NO	NO	YES
Journal dummies	YES	YES	YES
Year-Month dummies	YES	YES	YES

USA is a baseline for country-level comparisons, only coefficients from the top submitting countries are reported. Robust standard errors clustered at the journal level are reported in parentheses.
+p < .10; *p < .05; **p < .01; ***p < .001 (two-tailed tests).

Table S2. Effect of new policy on paper evaluation outcomes (linear probability models)

VARIABLES	(1) DV: passed the desk	(2) DV: passed the desk	(3) DV: first reviewer recommendation	(4) DV: first reviewer recommendation	(5) DV: final decision	(6) DV: final decision	(7) DV: last reviewer recommendation	(8) DV: last reviewer recommendation	(9) DV: min(last reviewer recommendation)	(10) DV: min(last reviewer recommendation)
New policy	-0.00443	0.00292	-0.00362	0.02425*	-0.01005	0.05596*	-0.00483	0.02537*	0.00165	0.06024**
	(0.018)	(0.021)	(0.008)	(0.010)	(0.016)	(0.023)	(0.009)	(0.010)	(0.014)	(0.018)
Middle prestige	0.02214***	0.02319***	-0.00138	0.00348	-0.00722	0.00268	-0.00522	0.00005	-0.00490	0.00339
	(0.005)	(0.005)	(0.005)	(0.004)	(0.007)	(0.007)	(0.005)	(0.005)	(0.007)	(0.007)
High prestige	0.02868**	0.03025**	0.01127*	0.01516**	0.00635	0.01370	0.00633	0.01046+	0.01020	0.01730*
	(0.010)	(0.010)	(0.005)	(0.006)	(0.009)	(0.009)	(0.006)	(0.006)	(0.008)	(0.009)
Female	0.01396+	0.01397+	0.00736	0.00736	0.01852**	0.01860**	0.01134*	0.01138*	0.02268***	0.02271***
	(0.008)	(0.008)	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)
New policy X Middle prestige		-0.00882		-0.04176***		-0.10198***		-0.04559***		-0.08596***
		(0.013)		(0.010)		(0.022)		(0.010)		(0.017)
New policy X High prestige		-0.01207		-0.03405***		-0.07782***		-0.03646***		-0.07452***
		(0.014)		(0.009)		(0.020)		(0.009)		(0.018)
Constant	0.81334***	0.81250***	0.81044***	0.80725***	0.76924***	0.76313***	0.81979***	0.81635***	0.68290***	0.67747***
	(0.027)	(0.026)	(0.011)	(0.011)	(0.016)	(0.016)	(0.011)	(0.011)	(0.016)	(0.016)
Observations	74,255	74,255	102,234	102,234	44,574	44,574	105,288	105,288	44,299	44,299
All country dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Journal FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year-Month dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.20	0.20	0.04	0.04	0.09	0.09	0.04	0.04	0.08	0.08

Robust standard errors clustered at the journal (Models 1-10) and manuscript (Models 3-4, 7-8) level are reported in parentheses.
+p < .10; *p < .05; **p < .01; ***p < .001 (two-tailed tests).

Table S3. Robustness test using alternative prestige binning with 0 citations defined as "low prestige" (n = 29,450), 1-100 citations as "lower-middle prestige" (n = 32,079), 101-1000 citations as "upper-middle prestige" (n = 16,639), and 1001+ citations as "high prestige" (n = 6,859): effect of new policy on paper evaluation outcomes (linear probability models)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	DV: passed the desk	DV: passed the desk	DV: first reviewer recommendation	DV: first reviewer recommendation	DV: final decision	DV: final decision	DV: last reviewer recommendation	DV: last reviewer recommendation	DV: min(last reviewer recommendation)	DV: min(last reviewer recommendation)
New policy	-0.00443	0.00305	-0.00373	0.02424*	-0.01020	0.05591*	-0.00496	0.02537*	0.00149	0.06024**
	(0.018)	(0.021)	(0.008)	(0.010)	(0.016)	(0.023)	(0.009)	(0.010)	(0.014)	(0.018)
Lower-middle prestige	0.02221***	0.02325***	-0.00130	0.00355	-0.00709	0.00281	-0.00514	0.00012	-0.00478	0.00349
	(0.005)	(0.005)	(0.005)	(0.004)	(0.007)	(0.007)	(0.005)	(0.005)	(0.007)	(0.007)
Upper-middle prestige	0.02425**	0.02460**	0.00560	0.00922	-0.00518	0.00225	0.00043	0.00367	0.00029	0.00650
	(0.009)	(0.008)	(0.006)	(0.006)	(0.010)	(0.009)	(0.006)	(0.006)	(0.009)	(0.009)
High prestige	0.04098*	0.04583**	0.02495***	0.02955***	0.03469**	0.04201***	0.02059**	0.02679***	0.03464**	0.04382***
	(0.016)	(0.016)	(0.007)	(0.006)	(0.010)	(0.009)	(0.006)	(0.006)	(0.011)	(0.011)
Female	0.01431+	0.01425+	0.00784	0.00783	0.01951**	0.01958**	0.01184**	0.01183**	0.02353***	0.02352***
	(0.008)	(0.008)	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)
New policy X Lower-middle prestige		-0.00881		-0.04171***		-0.10187***		-0.04556***		-0.08589***
		(0.013)		(0.010)		(0.022)		(0.010)		(0.017)
New policy X Upper-middle prestige		-0.00424		-0.03213**		-0.07861***		-0.02984***		-0.06641**
		(0.015)		(0.009)		(0.021)		(0.008)		(0.021)
New policy X High prestige		-0.03112+		-0.03893**		-0.07746**		-0.05096**		-0.09286***
		(0.018)		(0.013)		(0.024)		(0.015)		(0.023)
Constant	0.81267***	0.81192***	0.80998***	0.80681***	0.76835***	0.76223***	0.81934***	0.81596***	0.68222***	0.67687***
	(0.026)	(0.026)	(0.011)	(0.011)	(0.016)	(0.016)	(0.011)	(0.011)	(0.016)	(0.016)
Observations	74,255	74,255	102,234	102,234	44,574	44,574	105,288	105,288	44,299	44,299
All country dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Journal FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year-Month dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.20	0.20	0.04	0.04	0.09	0.09	0.04	0.04	0.08	0.08

Robust standard errors clustered at the journal (Models 1-10) and manuscript (Models 3-4, 7-8) level are reported in parentheses.
+p < .10; *p < .05; **p < .01; ***p < .001 (two-tailed tests).

Table S4. Robustness test using corresponding authors: effect of new policy on paper evaluation outcomes (linear probability models)

VARIABLES	(1) DV: passed the desk	(2) DV: passed the desk	(3) DV: first reviewer recommendation	(4) DV: first reviewer recommendation	(5) DV: final decision	(6) DV: final decision	(7) DV: last reviewer recommendation	(8) DV: last reviewer recommendation	(9) DV: min(last reviewer recommendation)	(10) DV: min(last reviewer recommendation)
New policy	0.00191	-0.00495	-0.00594	0.00991	-0.00217	0.05079*	-0.00538	0.01626	0.00478	0.04329*
	(0.019)	(0.019)	(0.006)	(0.009)	(0.014)	(0.020)	(0.007)	(0.010)	(0.011)	(0.017)
Middle prestige	0.03006***	0.03129***	0.00147	0.00515	0.00179	0.01023	-0.00222	0.00216	-0.00319	0.00322
	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.007)	(0.004)	(0.005)	(0.007)	(0.008)
High prestige	0.07832***	0.07562***	0.03031***	0.03159***	0.04114***	0.04573***	0.02602***	0.02815***	0.03509***	0.03821***
	(0.007)	(0.007)	(0.004)	(0.004)	(0.007)	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)
New policy X Middle prestige		-0.00670		-0.03090**		-0.08796***		-0.03752**		-0.06671***
		(0.008)		(0.010)		(0.023)		(0.012)		(0.019)
New policy X High prestige		0.01884		-0.01282		-0.05168***		-0.02016*		-0.03589*
		(0.012)		(0.008)		(0.015)		(0.009)		(0.015)
Female	-0.00112	-0.00108	0.00424	0.00429	0.00471	0.00488	0.00712*	0.00718*	0.01290*	0.01302*
	(0.007)	(0.007)	(0.003)	(0.003)	(0.006)	(0.006)	(0.003)	(0.003)	(0.006)	(0.006)
Constant	0.53155***	0.53276***	0.48763**	0.48618**	0.28615+	0.28193+	0.47105**	0.46884**	0.14896	0.14601
	(0.091)	(0.092)	(0.161)	(0.161)	(0.167)	(0.167)	(0.140)	(0.140)	(0.118)	(0.118)
Observations	105,693	105,693	143,613	143,613	62,403	62,403	147,925	147,925	62,045	62,045
All country dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Journal FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year-Month dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.19	0.19	0.04	0.04	0.09	0.09	0.04	0.04	0.07	0.07

Robust standard errors clustered at the journal (Models 1-10) and manuscript (Models 3-4, 7-8) level are reported in parentheses.
+p < .10; *p < .05; **p < .01; ***p < .001 (two-tailed tests).

Table S5. Explaining the effect of new policy on paper evaluation outcomes (instrumental variable regressions)

VARIABLES	(1) DV: passed the desk	(2) DV: passed the desk	(3) DV: first reviewer recommendation	(4) DV: first reviewer recommendation	(5) DV: final decision	(6) DV: final decision
Anonymized	-0.01635	0.01389	-0.01223	0.09026*	-0.03218	0.18988*
	(0.067)	(0.073)	(0.025)	(0.041)	(0.049)	(0.086)
Middle prestige	0.02211***	0.02317***	-0.00139	0.00353	-0.00727	0.00275
	(0.005)	(0.005)	(0.005)	(0.004)	(0.007)	(0.007)
High prestige	0.02855**	0.03025**	0.01118*	0.01522**	0.00611	0.01377
	(0.010)	(0.010)	(0.005)	(0.005)	(0.009)	(0.009)
Female	0.01404+	0.01412+	0.00741	0.00758	0.01856**	0.01896***
	(0.008)	(0.008)	(0.005)	(0.005)	(0.006)	(0.006)
Anonymized X Middle prestige		-0.03540		-0.15130***		-0.33874***
		(0.052)		(0.041)		(0.091)
Anonymized X High prestige		-0.05629		-0.13167**		-0.27561**
		(0.067)		(0.040)		(0.085)
Constant	0.51806***	0.51714***	0.77861***	0.77529***	0.64200***	0.63574***
	(0.025)	(0.025)	(0.011)	(0.011)	(0.016)	(0.016)
Observations	74,256	74,256	102,234	102,234	44,574	44,574
All country dummies	YES	YES	YES	YES	YES	YES
Journal FE	YES	YES	YES	YES	YES	YES
Year-Month dummies	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.20	0.20	0.04	0.04	0.09	0.09

Availability of the new policy is an instrument for the anonymization. Robust standard errors clustered at the journal (Models 1-6) and manuscript (Models 3-4) level are reported in parentheses.
+p < .10; *p < .05; **p < .01; ***p < .001 (two-tailed tests).

Table S6. Effect of new policy on author characteristics (OLS models)

	(1)	(2)	(3)	(4)
VARIABLES	DV: author # of works	DV: author # of works	DV: author # of cites	DV: author # of cites
New policy	0.84722	1.56246	14.86877	35.00669
	(1.902)	(1.731)	(64.728)	(63.321)
Female		-12.26233***		-218.93045***
		(1.208)		(37.192)
Constant	25.80345***	46.27434***	350.21483***	986.68752***
	(2.157)	(3.340)	(53.264)	(121.808)
Observations	85,026	75,965	85,026	75,965
All country dummies	NO	YES	NO	YES
Journal FE	YES	YES	YES	YES
Year-Month dummies	YES	YES	YES	YES
Adj. R-squared	0.03	0.08	0.02	0.04

Robust standard errors clustered at the journal level are reported in parentheses.
+p < .10; *p < .05; **p < .01; ***p < .001 (two-tailed tests).

Table S7. Effect of new policy and author characteristics on reviewer behavior and reviewer characteristics (OLS models)

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	DV: reviewer accepted invitation to review	DV: reviewer accepted invitation to review	DV: reviewer # of works	DV: reviewer # of works	DV: reviewer # of cites	DV: reviewer # of cites
New policy	-0.00239	-0.00418	-2.01338	-3.29871	-39.82226	-149.78753
	(0.008)	(0.011)	(2.360)	(2.574)	(94.305)	(110.109)
Middle prestige	0.00362	0.00285	-1.12280	-1.36774	-51.51275	-67.34661
	(0.004)	(0.004)	(0.951)	(1.053)	(44.309)	(46.747)
High prestige	0.00159	0.00178	2.02863+	1.79529	84.84934+	58.76461
	(0.004)	(0.005)	(1.174)	(1.197)	(47.366)	(44.025)
New policy X Middle prestige		0.00516		1.79368		121.30486+
		(0.011)		(2.034)		(70.186)
New policy X High prestige		-0.00075		1.70202		183.06720*
		(0.008)		(2.196)		(87.778)
Female	-0.00358	-0.00359	-1.89951*	-1.89824*	-27.91610	-27.72000
	(0.003)	(0.003)	(0.855)	(0.854)	(40.971)	(40.869)
Constant	0.26069***	0.26091***	101.16163***	101.33127***	2,234.68880***	2,249.41031***
	(0.012)	(0.012)	(3.164)	(3.139)	(140.059)	(137.413)
Observations	148,918	148,918	148,918	148,918	148,918	148,918
All country dummies	YES	YES	YES	YES	YES	YES
Journal FE	YES	YES	YES	YES	YES	YES
Year-Month dummies	YES	YES	YES	YES	YES	YES
Adj. R-squared	0.02	0.02	0.02	0.02	0.03	0.03

Models are run only on reviewers with *OpenAlex* ID. Robust standard errors clustered at the journal level are reported in parentheses. +p < .10; *p < .05; **p < .01; ***p < .001 (two-tailed tests).

Table S8. Effect of new policy on first reviewer recommendation (linear probability models with reviewer fixed-effects)

	(1)	(2)	(3)
VARIABLES	DV: first reviewer recommendation	DV: first reviewer recommendation	DV: first reviewer recommendation
New policy	0.00190	0.00325	0.03170*
	(0.007)	(0.010)	(0.013)
Middle prestige		0.00059	0.00591
		(0.006)	(0.007)
High prestige		0.01714*	0.02060*
		(0.008)	(0.008)
New policy X Middle prestige			-0.04524**
			(0.014)
New policy X High prestige			-0.03112*
			(0.015)
Female		0.00595	0.00592
		(0.008)	(0.008)
Constant	0.72076***	0.81067***	0.80735***
	(0.017)	(0.020)	(0.020)
Observations	96,904	47,706	47,706
Reviewer FE	YES	YES	YES
All country dummies	NO	YES	YES
Journal FE	YES	YES	YES
Year-Month dummies	YES	YES	YES
Adj. R-squared	0.18	0.20	0.20

Models are run only on reviewers with *OpenAlex* ID. Robust standard errors clustered at the journal, reviewer and manuscript levels are reported in parentheses.
+p < .10; *p < .05; **p < .01; ***p < .001 (two-tailed tests).