To ArXiv or not to ArXiv: A Study Quantifying Pros and Cons of Posting Preprints Online

Anonymous authors
Paper under double-blind review

Abstract

Double-blind conferences have engaged in debates over whether to allow authors to post their papers online on arXiv or elsewhere during the review process. Independently, some authors of research papers face the dilemma of whether to put their papers on arXiv due to its pros and cons. We conduct a study to substantiate this debate and dilemma via quantitative measurements. Specifically, we conducted surveys of reviewers in two top-tier double-blind computer science conferences—ICML 2021 (5361 submissions and 4699 reviewers) and EC 2021 (498 submissions and 190 reviewers). Our two main findings are as follows. First, more than a third of the reviewers self-report searching online for a paper they are assigned to review. Second, outside the review process, we find that preprints from better-ranked affiliations see a weakly higher visibility, with a correlation of 0.06 in ICML and 0.05 in EC. In particular, papers associated with the top-10-ranked affiliations had a visibility of approximately 11% in ICML and 22% in EC, whereas the remaining papers had a visibility of 7% and 18% respectively.

1 Introduction

Across academic disciplines, peer review is used to decide on the outcome of manuscripts submitted for publication. Single-blind reviewing used to be the predominant method in peer review, where the authors' identities are revealed to the reviewer in the submitted paper. However, several studies have found various biases in single-blind reviewing. These include bias pertaining to affiliations or fame of authors (Blank, 1991; Sun et al., 2021; Manzoor & Shah, 2021), bias pertaining to gender of authors (Rossiter, 1993; Budden et al., 2008; Knobloch-Westerwick et al., 2013; Roberts & Verhoef, 2016), and others (Link, 1998; Snodgrass, 2006; Tomkins et al., 2017). These works span several fields of science and study the effect of revealing the authors' identities to the reviewer through both observational studies and randomized control trials. These biases are further exacerbated due to the widespread prevalence of the Matthew effect—rich get richer and poor get poorer—in academia (Merton, 1968; Squazzoni & Claudio, 2012; Thorngate & Chowdhury, 2013). Biases based on authors' identities in peer review, coupled with the Matthew effect, can have far-reaching consequences on researchers' career trajectories.

As a result, many peer-review processes have moved to *double-blind* reviewing, where authors' names and other identifiers are removed from the submitted papers. Ideally, in a double-blind review process, neither the authors nor the reviewers of any papers are aware of each others' identity. However, a challenge for ensuring that reviews are truly double-blind is the exponential growth in the trend of posting papers online before review (Xie et al., 2021). Increasingly, authors post their preprints on online publishing websites such as arXiv and SSRN and publicize their work on social media platforms such as Twitter. The conventional publication route via peer review is infamously long and time-consuming. On the other hand, online preprint-publishing venues provide a platform for sharing research with the community usually without delays. Not only does this help science move ahead faster, but it also helps researchers avoid being "scooped". However, the increase in popularity of making papers publicly available—with author identities—before or during the

review process, has led to the dilution of double-blinding in peer review. For instance, the American Economic Association, the flagship journal in economics, dropped double-blinding in their reviewing process citing its limited effectiveness in maintaining anonymity. The availability of preprints online presents a challenge in double-blind reviewing, which could lead to biased evaluations for papers based on their authors' identities, similar to single-blind reviewing.

This dilution has led several double-blind peer-review venues to debate whether authors should be allowed to post their submissions on the Internet, before or during the review process. For instance, top-tier machine learning conferences such as NeurIPS and ICML do not prohibit posting online. On the other hand, the Association of Computational Linguistics (ACL) recently introduced a policy for its conferences in which authors are prohibited from posting their papers on the Internet starting a month before the paper submission deadline till the end of the review process. The Conference on Computer Vision and Pattern Recognition (CVPR) has banned the advertisement of submissions on social media platforms for such a time period. Some venues are stricter, for example, the IEEE Communication Letters and IEEE International Conference on Computer Communications (INFOCOMM) disallows posting preprints to online publishing venues before acceptance.

Independently, authors who perceive they may be at a disadvantage in the review process if their identity is revealed face a dilemma regarding posting their work online. They stand to either hurt their paper's chances of acceptance by revealing their identity online or lose out on publicity for their paper by refraining from posting.

It is thus important to quantify the consequences of posting preprints online to (i) enable an evidence-based debate over conference policies, and (ii) help authors make informed decisions about posting preprints online. In our work, we conduct a large-scale survey-based study in conjunction with the review process of two top-tier publication venues in computer science that have double-blind reviewing: the 2021 International Conference on Machine Learning (ICML 2021) and the 2021 ACM Conference on Economics and Computation (EC 2021).[1] Specifically, we design and conduct experiments aimed at answering the following research questions:

(Q1) What fraction of reviewers, who had not seen the paper they were reviewing before the review process, deliberately search for the paper on the Internet during the review process?

(Q2) What is the relation between the rank of the authors' affiliations and the visibility of a preprint to its target audience?

By addressing these research questions, we aim to measure some of the effects of posting preprints online, and help quantify their associated risks and benefits for authors from different institutions.

2 Related work

Surveys of reviewers. Several studies survey reviewers to obtain insights into reviewer perceptions and practices. Nobarany et al. (2016) surveyed reviewers in the field of human-computer interaction to gain a better understanding of their motivations for reviewing. They found that encouraging high-quality research, giving back to the research community, and finding out about new research were the top general motivations for reviewing. Along similar lines, Tite & Schroter (2007) surveyed reviewers in biomedical journals to understand why peer reviewers decline to review. Among the respondents, they found the most important factor to be conflict with other workload.

Resnik et al. (2008) conducted an anonymous survey of researchers at a government research institution concerning their perceptions about ethical problems with journal peer review. They found that the most

[1] In Computer Science, conferences are typically the terminal publication venue and are typically ranked at par or higher than journals. Full papers are reviewed in CS conferences, and their publication has archival value.

common ethical problem experienced by the respondents was incompetent review. Additionally, 6.8% respondents mentioned that a reviewer breached the confidentiality of their article without permission. This survey focused on the respondents' perception, and not on the actual frequency of breach of confidentiality. In another survey, by Martinson et al. (2005), 4.7% authors self-reported publishing the same data or results in more than one publication. Fanelli (2009) provides a systematic review and meta analysis of surveys on scientific misconduct including falsification and fabrication of data and other questionable research practices.

Goues et al. (2018) surveyed reviewers in three double-blind conferences to investigate the effectiveness of anonymization of submitted papers. In their experiment, reviewers were asked to guess the authors of the papers assigned to them. Out of all reviews, 70%-86% of the reviews did not have any author guess. Here, absence of a guess could imply that the reviewer did not have a guess or they did not wish to answer the question. Among the reviews containing guesses, 72%-85% guessed at least one author correctly.

Analyzing papers posted versus not posted on arXiv. Bharadhwaj et al. (2020) aim to analyse the risk of selective de-anonymization through an observational study based on open review data from the International Conference on Learning Representations (ICLR). The analysis quantifies the risk of de-anonymization by computing the correlation between papers' acceptance rates and their authors' reputations separately for papers posted and not posted online during the review process. This approach however is hindered by the confounder that the outcomes of the analysis may not necessarily be due to de-anonymization of papers posted on arXiv, but could be a result of higher quality papers being selectively posted on arXiv by famous authors. Moreover, it is not clear how the paper draws conclusions based on the analysis presented therein. Our supporting analysis overlaps with the investigation of Bharadhwaj et al. (2020): we also investigate the correlation between papers' acceptance rates and their authors' associated ranking in order to support our main analysis and to account for confounding by selective posting by higher-ranked authors.

Aman (2014) also investigate possible benefits of publishing preprints on arXiv in *Quantitative Biology*, wherein they measure and compare the citations received by papers posted on arXiv and those received by papers not posted on arXiv. A similar confounder arises here that a positive result could be a false alarm due to higher quality papers being selectively posted on arXiv by authors.

In our work, we quantify the risk of de-anonymization by directly studying reviewer behaviour regarding searching online for their assigned papers. We quantify the effects of publishing preprints online by measuring their visibility using a survey-based experiment querying reviewers whether they had seen a paper before.

Studies on peer review in computer science. Our study is conducted in two top-tier computer science conferences and contributes to a growing list of studies on peer review in computer science. Lawrence & Cortes (2014); Beygelzimer et al. (2021) quantify the (in)consistencies of acceptance decisions on papers. Several studies (Madden & DeWitt, 2006; Tung, 2006; Tomkins et al., 2017; Manzoor & Shah, 2021) study biases due to single-blind reviewing. Shah et al. (2018) study several aspects of the NeurIPS 2016 peer-review process. Stelmakh et al. (2021c) study biases arising if reviewers know that a paper was previously rejected. Stelmakh et al. (2021b) study a pipeline for getting new reviewers into the review pool. Stelmakh et al. (2020) study herding in discussions. Stelmakh et al. (2022) study citation bias in peer review. A number of recent works (Charlin & Zemel, 2013; Stelmakh et al., 2021a; Kobren et al., 2019; Jecmen et al., 2020; Noothigattu et al., 2021) have designed algorithms that are used in the peer-review process of various computer science conferences. See Shah (2021) for an overview of such studies and computational tools to improve peer review.

3 Methods

We now outline the design of the experiment that we conducted to investigate the research questions in this work. First, in Section 3.1 we introduce the two computer science conferences ICML 2021 and EC 2021 that formed the venues for our investigation, and describe research questions Q1 and Q2 in the context of these two conferences. Second, in Section 3.2 we describe the experimental procedure. Finally, in Section 3.3 we provide the details of our analysis methods.

3.1 Preliminaries

Experiment setting. The study was conducted in the peer-review process of two conferences:

- **ICML 2021** International Conference on Machine Learning is a flagship machine learning conference. ICML is a large conference with 5361 submissions and 4699 reviewers in its 2021 edition.
- **EC 2021** ACM Conference on Economics and Computation is the top conference at the intersection of Computer Science and Economics. EC is a relatively smaller conference with 498 submissions and 190 reviewers in its 2021 edition.

Importantly, the peer-review process in both conferences, ICML and EC, is organized in a double-blind manner, defined as follows. In a **double-blind peer-review process**, the identity of all the authors is removed from the submitted papers. No part of the authors' identity, including their names, affiliations, and seniority, is available to the reviewers through the review process. At the same time, no part of the reviewers' identity is made available to the authors through the review process.

We now formally define some terminology used in the research questions Q1 and Q2. The first research question, Q1, focuses on the fraction of reviewers who deliberately search for their assigned paper on the Internet. The second research question, Q2, focuses on the correlation between the visibility to a target audience of papers available on the Internet before the review process, and the rank of the authors' affiliations. In what follows, we explicitly define the terms used in Q2 in the context of our experiments—target audience, visibility, preprint, and rank associated with a paper.

Paper's target audience. For any paper, we define its target audience as members of the research community that share similar research interests as that of the paper. In each conference, a 'similarity score' is computed between each paper-reviewer pair, which is then used to assign papers to reviewers. We used the same similarity score to determine the target audience of a paper (among the set of reviewers in the conference). We provide more details in Appendix A.

Paper's visibility. We define the visibility of a paper to a member of its target audience as a binary variable which is 1 if that person has seen this paper outside of reviewing contexts, and 0 otherwise. Visibility, as defined here, includes reviewers becoming aware of a paper through preprint servers or other platforms such as social media, research seminars and workshops. On the other hand, visibility does *not* include reviewers finding a paper during the review process (e.g., visibility does not include a reviewer discovering an assigned paper by deliberate search or accidentally while searching for references).

Preprint. To study the visibility of papers released on the Internet before publication, we checked whether each of the papers submitted to the conference was available online. Specifically, for EC, we manually searched for all submitted papers to establish their presence online. On the other hand, for ICML, owing to its large size, we checked whether a submitted paper was available on arXiv (`arxiv.org`). ArXiv is the predominant platform for pre-prints in machine learning; hence we used availability on arXiv as a proxy indicator of a paper's availability on the Internet.

Rank associated with a paper. In this paper, the rank of an author's affiliation is a measure of author's prestige that, in turn, is transferred to the author's paper. We determine the rank of affiliations in ICML and EC based on widely available rankings of institutions in the respective research communities. Specifically, in ICML, we rank (with ties) each institution based on the number of papers published in the ICML conference in the preceding year (2020) with at least one author from that institution (Ivanov, 2020). On the other hand, since EC is at the intersection of two fields, economics and computation, we merge three rankings—the QS ranking for computer science (QS, 2021a), the QS ranking for economics and econometrics (QS, 2021b), and the CS ranking for economics and computation (CSRankings, 2021)—by taking the best available rank for each institution to get our ranking of institutions submitting to EC. By convention, better ranks, representing

more renowned institutions, are represented by lower numbers; the top-ranked institution for each conference has rank 1. Finally, we define the rank of a paper as the rank of the best-ranked affiliation among the authors of that paper. Due to ties in rankings, we have 37 unique rank values across all the papers in ICML 2021, and 66 unique rank values across all the papers in EC 2021.

3.2 Experiment design

To address Q1 and Q2, we designed survey-based experiments for EC 2021 and ICML 2021, described next.

Design for Q1. To find the fraction of reviewers that deliberately search for their assigned paper on the Internet, we surveyed the reviewers. Importantly, as reviewers may not be comfortable answering questions about deliberately breaking the double-blindness of the review process, we designed the survey to be anonymous. We used the Condorcet Internet Voting Service (CIVS) (Myers, 2003), a widely used service to conduct secure and anonymous surveys. Further, we took some steps to prevent our survey from spurious responses (e.g., multiple responses from the same reviewer). For this, in EC, we generated a unique link for each reviewer that accepted only one response. In ICML we generated a link that allowed only one response per IP address and shared it with reviewers asking them to avoid sharing this link with anyone.[2] The survey form was sent out to the reviewers via CIVS after the initial reviews were submitted. In the e-mail, the reviewers were invited to participate in a one-question survey on the consequences of publishing preprints online. The survey form contained the following question:

> "During the review process, did you search for any of your assigned papers on the Internet?"

with two possible options: *Yes* and *No*. The respondents had to choose exactly one of the two options. To ensure that the survey focused on reviewers deliberately searching for their assigned papers, right after the question text, we provided additional text: "Accidental discovery of a paper on the Internet (e.g., through searching for related works) does not count as a positive case for this question. Answer *Yes* only if you tried to find an assigned paper itself on the Internet."

Following the conclusion of the survey, CIVS combined the individual responses, while maintaining anonymity, and provided the total number of *Yes* and *No* responses received.

Design for Q2. Recall that for Q2 we want to find the correlation between preprints' visibility to a target audience and its associated rank. Following the definitions provided in Section 3.1, we designed a survey-based experiment as follows. We conducted a survey to query reviewers about some papers for which they are considered a target audience. Specifically, we asked reviewers if they had seen these papers before outside of reviewing contexts. We provide more details about the survey, including the phrasing of the survey question, in Appendix A. We queried multiple reviewers about each paper, and depending on their response, we considered the corresponding visibility to be 1 if the reviewer said they had seen the paper before outside of reviewing contexts and 0 otherwise. We note that in ICML reviewers were queried about the papers they were assigned to review using the reviewer response form, in which a response to the question of visibility was required. Meanwhile, in EC, reviewers were queried about a set of papers that they were not assigned to review, using a separate optional survey form that was emailed to them by the program chairs after the rebuttal phase and before the announcement of paper decisions. The survey designed for Q2 had a response rate of 100% in ICML, while EC had a response rate of 55.78%.

3.3 Analysis

We now describe the analysis for the data collected to address Q1 and Q2. Importantly, our analysis is the same for the data collected from ICML 2021 and EC 2021. For Q1, we directly report the numbers obtained

[2] The difference in procedures between EC and ICML is due to a change in the CIVS policy that was implemented between the two surveys.

from CIVS regarding the fraction of reviewers who searched for their assigned papers online in the respective conference. In this section, we describe our analysis for Q2, where we want to analyse the effect of papers' ranking on visibility. Recall that for Q2, we collected survey responses and observational data about which papers submitted to ICML or EC were posted online before the corresponding review process. Since the latter data is observational, we describe two possible confounding factors in our setting.

3.3.1 Confounding factors

For a paper posted online, the amount of time for which it has been available on the Internet can affect the visibility of the paper. For instance, papers posted online well before the deadline may have higher visibility as compared to papers posted near the deadline. Moreover, the time of posting a paper online could vary across institutions ranked differently. Thus, time of posting can be a confounding factor. In order to control for this factor, for the papers that were posted online before the review process, we incorporate the time gap between posting of the papers and submission of reviews in our analysis, which is described next in Section 3.3.2.

Second, one should ideally study the visibility of all papers submitted to the conference. However, in our experiment, we are naturally limited to studying the visibility of the papers that were released on the internet before the conclusion of the review process. The choice of posting online before or after the review process could vary depending on the quality of the paper as well as on the rank of the authors' affiliations. This can form a confounding factor in our analysis. To understand whether papers posted online before the review process had significantly different quality and rank profile from papers not posted online, we provide supporting analysis in Section 3.3.3.

3.3.2 Analysis procedure

We now describe our analysis to compute the relation between a paper's visibility and associated rank. In the following analysis procedure, we consider each response obtained in the survey for Q2 as one unit. Each response corresponds to a paper-reviewer pair, wherein the reviewer was queried about seeing the considered paper. In case of no response from reviewer, we do not consider the corresponding paper-reviewer pairs in our data. We thus have two variables associated to each response: the visibility of the paper to the reviewer (in $\{0, 1\}$), and the rank associated with the paper. Recall that we define the rank of a paper as the rank of the best-ranked affiliation associated with that paper.

We first describe the approach to control for confounding due to time of posting. There is ample variation in the time of posting papers online within the papers submitted to ICML and EC: some papers were posted right before the review process began while some papers were posted two years prior. To account for the causal effect of time of posting on visibility, we divide the responses into bins based on the number of days between the paper being posted online and the deadline for submitting responses to the Q2 survey. Since similar conference deadlines arrive every three months roughly and the same conference appears every one year, we binned the responses accordingly into three bins. Specifically, if the number of days between the paper being posted online and the survey response is less than 90, it is assigned to the first bin, if the number of days is between 90 and 365, the response is assigned to the second bin, and otherwise, the response is assigned to the third bin. Following this binning, we assume that time of posting does not affect the visibility of papers within the same bin. Consequently, we analyse the correlation between papers' visibility and associated rank separately within each bin and then combine them to get the overall effect in two steps:

Step 1. We compute the correlation coefficient between papers' visibility and associated rank within each bin. For this we use Kendall's Tau-b statistic, which is closely related to the widely used Kendall's Tau rank correlation coefficient (Kendall, 1938). Kendall's Tau statistic provides a measure of the strength and direction of association between two variables measured on an ordinal scale. It is a non-parametric measure that does not make any assumptions about the data. However, it does not account for ties and our data has a considerable number of ties, since visibility is a binary variable and the rankings used contain ties. Therefore, we use a variant of the statistic, Kendall's Tau-b statistic, that accounts for ties in the data.

Within each bin we consider all the responses obtained and their corresponding visibility and rank value, and compute Kendall's Tau-b correlation coefficient between visibility and rank. The procedure for computing Kendall's Tau-b correlation coefficient between two real-valued vectors (of the same length) is described in Appendix B.1. We now make a brief remark of a notational convention we use in this paper, in order to address ambiguity between the terminology "high-rank institutions" as well as "rank 1, 2,... institutions", both of which colloquially refers to better-rank institutions. It is intuitive to interpret a positive correlation between visibility and rank as the visibility increasing with an *improvement* in the rank. Consequently, we flip the sign of all correlation coefficients computed with respect to the rank variable.

Step 2. With the correlation computed within each bin, we compute the overall correlation using a sample-weighted average (Corey et al., 1998). Formally, let N_1, N_2 and N_3 denote the number of responses obtained in the first, second and third bin respectively. Denote Kendall's Tau-b correlation coefficients within the three bins as τ_1, τ_2 and τ_3. Then the correlation T between papers' visibility and rank over all the time bins is computed as

$$T = \frac{N_1\,\tau_1 + N_2\,\tau_2 + N_3\,\tau_3}{N_1 + N_2 + N_3}. \tag{1}$$

The statistic T gives us the effect size for our research question Q2. Finally, to analyse the statistical significance of the effect, we conduct a permutation test, wherein we permute our data within each bin and recompute the test statistic T to obtain a p-value for our test. We provide the complete algorithm for the permutation test in Appendix B.2.

3.3.3 Supporting analysis

As mentioned earlier in Section 3.3.1, we have to account for the papers not posted online before the survey for Q2 was conducted, to have a complete understanding of the effect of rank on paper visibility. We are unable to measure the visibility of these papers, as they were not posted online yet. Thus, we investigate whether the pool of papers posted online before the review process is significantly different, in terms of their rank profile, from the rest of the papers submitted to the conference. In the following analysis, we consider all papers submitted to the conference, and we consider each submitted paper as one unit of analysis.

First, we analyse the relationship between a binary value indicating whether a submitted paper was posted online before the Q2 survey, and the paper's associated rank. For this, we compute Kendall's Tau-b statistic between the two values for all papers submitted to the conference, and flip the sign of the statistic with respect to the rank variable.

Second, we investigate whether there is a significant difference between the papers posted and not posted online before the review process, in terms of their quality and rank profile. Here we measure the quality of a paper as a binary variable based on its final decision in the conference (`accept` or `reject`). We give an example to understand the motivation for this supporting analysis. Suppose the double-blind process works perfectly, and assume that among the papers with better-ranked affiliations, only the high-quality papers are posted online, while there is no quality-based selection among other papers. Then, for better-ranked papers, the difference in acceptance rates for papers posted and those not posted online would be higher than the difference for the other papers. Suppose there is no causal effect of ranking on papers' visibility. Assuming higher-quality papers would enjoy higher visibility, such self-selection could lead to false discovery of effect.

We conduct this analysis by computing three statistics. First, for all papers posted online before the Q2 survey, we compute Kendall's Tau-b statistic between their rank and their final decision. Second, for all papers *not* posted online, we compute Kendall's Tau-b statistic between their rank and their final decision. Third, for each unique rank value, for the corresponding papers with that rank, we compute the difference between the average acceptance rate for papers posted online and those not posted online. Then, we compute Kendall's Tau-b statistic between the rankings and the difference in acceptance rate. Finally, we flip the sign of all correlation coefficients computed with respect to the rank variable. Hence, a positive correlation would imply that the (difference in) acceptance rate increases as the rank improves.

		EC 2021	ICML 2021
1	# Reviewers	190	4699
2	# Survey Respondents	97	753
3	# Survey respondents who said they searched for their assigned paper online	41	269
4	% Survey respondents who said they searched for their assigned paper online	42%	36%

Table 1: Outcome of survey for research question Q1.

4 Main results

We now discuss the results from the experiments conducted in ICML 2021 and EC 2021.

4.1 Q1 results

Table 1 provides the results of the survey for research question Q1. The percentage of reviewers that responded to the anonymous survey for Q1 is 16% (753 out of 4699) in ICML and 51% (97 out of 190) in EC. While the coverage of the pool of reviewers is small in ICML (16%), the number of responses obtained is large (753). As shown in Table 1, the main observation is that, in both conferences, at least a third of the Q1 survey respondents self-report deliberately searching for their assigned paper on the Internet. There is substantial difference between ICML and EC in terms of the response rate as well as the fraction of *Yes* responses received, however, the current data cannot provide explanations for these differences.

4.2 Q2 results

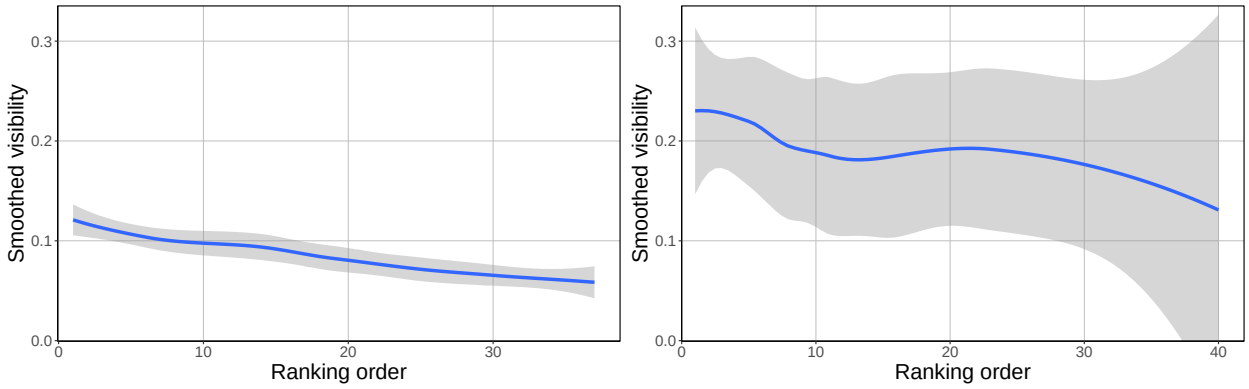
We discuss the results of the survey conducted for Q2 in ICML 2021 and EC 2021. First we discuss the results of the main analysis described in Section 3.3.2. Then we discuss the results of the supporting analysis described in Section 3.3.3.

Main analysis. Table 2 depicts the results of the survey for research question Q2. We received 7594 responses and 449 responses for the survey for Q2 in ICML and EC respectively (Row 1). Based on our binning rule based on time of posting described in Section 3.3.2, we see more papers in bin 1 and bin 2, compared to bin 3. This suggests that majority of preprints were posted online within one year of the review process (Row 2).

As shown in Table 2, for papers submitted to the respective conference and posted online before the review process, we observe that there is a weak positive correlation between the papers' visibility and its associated rank. The weak positive correlation implies that the visibility increases slightly as the rank improves.

		EC 2021	ICML 2021
1	# Responses overall	449	7594
2	# papers in bins 1, 2, 3	$63, 82, 38$	$968, 820, 146$
3	# Responses in bins 1, 2, 3	$159, 233, 57$	$3799, 3228, 567$
4	Correlation between rank and visibility $[-1, 1]$	0.05 ($p = 0.11$)	0.06 ($p < 10^{-5}$)
5	Correlation between rank and visibility in bins 1, 2, 3	$0.06, 0.04, 0.04$	$0.04, 0.10, 0.03$
6	p-value associated with correlations in Row 5	$0.36, 0.46, 0.66$	$0.004, < 10^{-5}, 0.19$
7	% Visibility overall $[0 - 100]$	20.5% (92 out of 449)	8.36% (635 out of 7594)
8	% Visibility for papers with top 10 ranks $[0 - 100]$	21.93% (59 out of 269)	10.91% (253 out of 2319)
9	% Visibility for papers below top 10 ranks $[0 - 100]$	18.33% (33 out of 180)	7.24% (382 out of 5275)

Table 2: Outcome of main analysis for research question Q2. A positive correlation in Row 4 and Row 5 implies that the visibility increases as the rank of the paper improves. Recall that for ICML, we consider the set of responses obtained for submissions that were available as preprints on arXiv. There were 1934 such submissions.



(a) ICML 2021

(b) EC 2021

Figure 1: Using responses obtained in Q2 survey, we plot the papers' visibility against papers' associated rank with smoothing. On the x-axis, we order papers by their ranks (i.e., paper with the best rank gets order 1, paper with the second best rank gets order 2, and so on). The range of x-axis is given by the number of unique ranks in the visibility analysis, which may be smaller than the total number of unique ranks associated with the papers in the respective conferences. The x-axis range is 37 in Figure 1a and 40 in Figure 1b due to ties in rankings used. On the y-axis, smoothed visibility lies in $[0, 1]$. We use local linear regression for smoothing (Cleveland & Loader, 1996). The solid line gives the smoothed visibility, and the grey region around the line gives the 95% confidence interval.

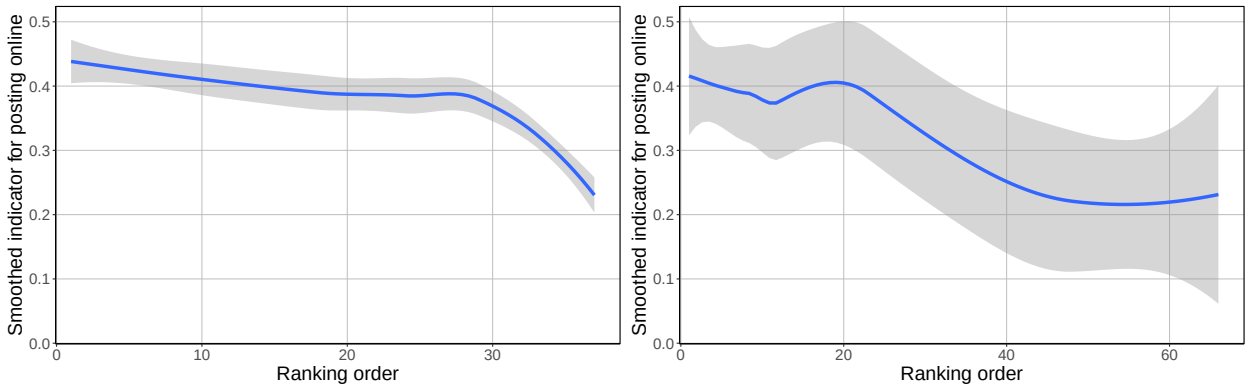
To provide some interpretation of the correlation coefficient values in Row 4, we compare the mean visibility within and without responses obtained for papers with at least one affiliation ranked 10 or better (Row 8 and 9). There are 10 and 23 institutions among the top-10 ranks in ICML and EC respectively. We see that there is more than 3 percentage points decrease in mean visibility across these two sets of responses in both ICML and EC. Figure 1 displays additional visualization that helps to interpret the strength of the effect of the papers' rank on visibility. The data suggests that top-ranked institutions enjoy higher visibility than lower-ranked institutions in both venues ICML 2021 and EC 2021.

In summary, in ICML the analysis supports a small but statistically significant effect of paper ranking on its visibility. In EC the effect size is comparable, but the effect does not reach statistical significance. Without further data, for EC the results are only suggestive.

As an aside, we note that the mean visibility in ICML 2021 (8.36%) is much lower than that in EC 2021 (20.5%). This may be attributed to the following reason: The research community in EC is smaller and more tight-knit, meaning that there is higher overlap in research interests within the members of the community (reviewers). On the other hand, ICML is a large publication venue with a more diverse and spread-out research community.

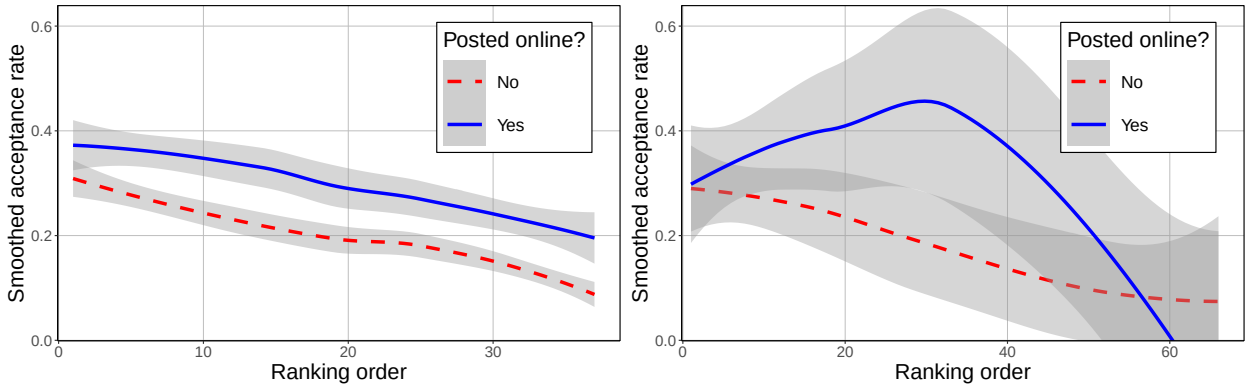
Supporting analysis. We provide the results for the supporting analysis described in Section 3.3.3 in Table 3. There were a total of 5361 and 498 papers submitted to ICML 2021 and EC 2021 respectively, out of which 1934 and 183 were posted online before the end of the review process respectively (Row 1 and 2). Thus, we see that more than a third of the papers submitted were available online. Among all the papers submitted, we observe that there is a positive correlation (Kendall's Tau-b) between paper's rank and whether it was posted online before the review process in both ICML and EC of 0.12 ($p < 10^{-5}$) and 0.09 ($p = 0.01$) respectively (Row 3). This implies that the authors from higher-ranked institutions are more likely to post their papers online before the review process. In Figure 2 we provide visualization to interpret the correlation between ranking and uploading behaviour.

Next, to understand if there is significant difference in the quality of papers uploaded online by authors from institutions with different ranks, we compare the final decision of the pool of papers posted online before



(a) ICML 2021 (b) EC 2021

Figure 2: For papers submitted to the respective conferences, we plot the indicator for paper being posted online before the end of the review process against papers' associated rank, with smoothing. On the x-axis, we have the ranking order as described in Figure 1. On the y-axis, smoothed indicator for posting online lies in $[0, 1]$. We use locally estimated smoothing to get the smoothed indicator for posting online across ranks, shown by the solid line, and a 95% confidence interval, shown by the grey region.



(a) ICML 2021 (b) EC 2021

Figure 3: For papers submitted to the respective conferences and (not) posted online before the review process, we plot the papers' final decision against papers' associated rank, with smoothing. On the x-axis, we have the ranking order as described in Figure 1. On the y-axis, smoothed acceptance rate lies in $[0, 1]$. We use locally estimated smoothing to get the smoothed acceptance rate across ranks, shown by the lines, and a 95% confidence interval, shown by the grey region. Note that in Figure 3b, the number of papers corresponding to the ranks on the right of the plot is very small.

the review process and the pool of papers that was not, across ranks. Here, we use the final decision as a proxy indicator for the quality of paper. Note that this proxy is not a perfect indicator, it could be affected by the ranking of the paper in case of de-anonymization of the paper, as we discussed in Section 1. Now, for the pool of papers posted online, we see that Kendall's Tau-b correlation between papers' rank and final decision is 0.11 ($p < 10^{-5}$) in ICML and 0.03 ($p = 0.58$) in EC (Row 4). Recall that a positive correlation implies that the acceptance rate increases as the rank improves. For the pool of papers *not* posted online, we see that Kendall's Tau-b correlation between papers' rank and final decision, 0.16 ($p < 10^{-5}$) in ICML and 0.13 ($p = 0.006$) in EC (Row 5). Further, the correlation between the rank values and the corresponding difference (between papers posted and not posted online) in mean acceptance rates is 0.01 ($p = 0.92$) in ICML and 0.12 ($p = 0.18$) in EC (Row 6).

To interpret these values, we provide visualization of the variation of mean acceptance rate as rank varies for the two pools of papers in Figure 3. In ICML, we see that the difference in acceptance rates between the two pools of papers roughly remains the same as the rank changes. Meanwhile, in EC this difference in acceptance rates does not suggest a clear trend across ranks. Now, recall the earlier discussion regarding self-selection of papers being posted online, that authors from high-rank institutions may upload only high quality papers, while authors from low-rank institutions may not select the papers to be uploaded online based on quality. Based on the results displayed in Table 3 and Figure 3, we do not see strong evidence for self-selection of papers being posted online by authors from high-rank institutions compared to those from low-rank institutions.

5 Discussion

To improve peer review and scientific publishing in a principled manner, it is important to understand the quantitative effects of the policies in place, and design policies in turn based on these quantitative measurements.

We find that more than a third of survey respondents self-report deliberately searching for their assigned papers online, thereby weakening the effectiveness of author anonymization in double-blind peer review. Further, the observed value of fraction of reviewers that searched for their assigned paper online in Table 1 might be an underestimate due of two reasons: (i) Reviewers who deliberately broke the double-blindedness of the review process may be more reluctant to respond to our survey for Q1. (ii) As we saw in Section 4.2, roughly 8% of reviewers in ICML 2021 had already seen their assigned paper before the review process began (Table 2 row 5). If these reviewers were not already familiar with their assigned paper, they may have searched for them online during the review process.

For Q2, the effect size is statistically significant in ICML, but not in EC. A possible explanation for the difference is in the method of assigning rankings to institutions, described in Section 3.1. For ICML, the rankings used are directly related to past representation of the institutions at ICML (Ivanov, 2020). In EC,

		EC 2021	ICML 2021
1	# Papers	498	5361
2	# Papers posted online before the end of review process	183	1934
3	Correlation between papers' rank and whether they were posted online $[-1, 1]$	0.09	0.12
4	Correlation for papers posted online between their rank and decision $[-1, 1]$	0.03	0.11
5	Correlation for papers not posted online between their rank and decision $[-1, 1]$	0.13	0.16
6	Correlation between ranking and corresponding difference, between papers posted and not posted online, in mean acceptance rate $[-1, 1]$	0.12	0.01

Table 3: Outcome of supporting analysis for research question Q2. A positive correlation in row 3, 4 and 5 implies that the value of the variable considered increases as the rank of the paper improves. For instance, in row 3, the rate of posting online increases as the rank improves.

we used popular rankings of institutions such as QS rankings and CS rankings. In this regard, we observe that there is no clear single objective measure for ranking institutions in a research area. This leads to many ranking lists that may not agree with each other. Our analysis also suffers from this limitation.

Next, while we try to carefully account for confounding factors based on time of posting in our analysis for Q2, our study remains dependent on observational data. Thus, the usual caveat of unaccounted for confounding factors applies to our work. For instance, the topic of research may be a confounding factor in the effect of papers' rank on visibility: If authors from better-ranked affiliations work more on cutting-edge topics compared to others, then their papers would be read more widely. This could potentially increase the observed effect.

Policy implications. Double-blind venues now adopt various policies for authors regarding posting or advertising their work online before and during the review process. A notable example is a recent policy change by the Association for Computational Linguistics in their conference review process, which includes multiple conferences: ACL, NAACL (North American Chapter of the ACL) and EMNLP (Empirical Methods in Natural Language Processing). ACL introduced an anonymity period for authors, starting a month before the paper submission deadline and extending till the end of the review process. According to their policy, within the anonymity period authors are not allowed to post or discuss their submitted work anywhere on the Internet (or make updates to existing preprints online). In this manner, the conference aims to limit the de-anonymization of papers from posting preprints online. A similar policy change has been instituted by the CVPR computer vision conference. We provide some quantitative insights on this front using the data we collected from the Q2 survey in ICML 2021 and EC 2021. There were 918 (out of 5361 submitted) and 74 (out of 498 submitted) papers posted online *during* the one month period right before the submission deadline in ICML and EC respectively. These papers enjoyed a visibility of 8.11% (292 out of 3600) and 23.81% (45 out of 189) respectively. Meanwhile, there were 1016 (out of 5361) and 109 (out of 498) papers posted online *prior to* the one month period right before the submission deadline in ICML and EC, and these papers enjoyed a visibility of 8.59% (343 out of 3994) and 18.08% (47 out of 260) respectively. These measurements may help inform subsequent policy decisions.

While our work finds dilution of anonymization in double-blind reviewing, any prohibition on posting preprints online comes with its own downsides. For instance, consider fields such as Economics where journal publication is the norm, which can often imply several years of lag between paper submission and publication. Double-blind venues must grapple with the associated trade-offs, and we conclude with a couple of suggestions for a better trade-off. First, many conferences, including but not limited to EC 2021 and ICML 2021, do not have clearly stated policies for reviewers regarding searching for papers online, and can clearly state as well as communicate these policies to the reviewers. Second, venues may consider policies requiring authors to use a different title and reword the abstract during the review process as compared to the versions available online, which may reduce the chances of reviewers discovering the paper or at least introduce some ambiguity if a reviewer discovers (a different version of) the paper online.

References

Valeria Aman. Is there any measurable benefit in publishing preprints in the arXiv section quantitative biology? *arXiv preprint arXiv:1411.1955*, 2014.

Alina Beygelzimer, Yann Dauphin, Percy Liang, and Jennifer Wortman Vaughan. The NeurIPS 2021 consistency experiment. `https://blog.neurips.cc/2021/12/08/the-neurips-2021-consistency-experiment/`, 2021.

Homanga Bharadhwaj, Dylan Turpin, Animesh Garg, and Ashton Anderson. De-anonymization of authors through arXiv submissions during double-blind review. *arXiv preprint arXiv:2007.00177*, 2020.

Rebecca M Blank. The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The American Economic Review. *American Economic Review*, 81(5):1041–1067, December 1991.

Amber E. Budden, Tom Tregenza, Lonnie W. Aarssen, Julia Koricheva, Roosa Leimu, and Christopher J. Lortie. Double-blind review favours increased representation of female authors. *Trends in ecology & evolution*, 23 1:4–6, 2008.

L. Charlin and R. S. Zemel. The Toronto Paper Matching System: An automated paper-reviewer assignment system. In *ICML Workshop on Peer Reviewing and Publishing Models*, 2013.

William S. Cleveland and Clive Loader. Smoothing by local regression: Principles and methods. In Wolfgang Härdle and Michael G. Schimek (eds.), *Statistical Theory and Computational Aspects of Smoothing*, pp. 10–49, Heidelberg, 1996. Physica-Verlag HD. ISBN 978-3-642-48425-4.

David Corey, William Dunlap, and Michael Burke. Averaging correlations: Expected values and bias in combined pearson rs and fisher's z transformations. *Journal of General Psychology - J GEN PSYCHOL*, 125:245–261, 07 1998. doi: 10.1080/00221309809595548.

CSRankings. Computer Science Rankings: Economics and Computation, 2021. `http://csrankings.org/#/fromyear/2011/toyear/2021/index?ecom&world` [Last Accessed: 10/15/2021].

Daniele Fanelli. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS one*, 4(5):e5738, 2009.

Claire Le Goues, Yuriy Brun, Sven Apel, E. Berger, Sarfraz Khurshid, and Yannis Smaragdakis. Effectiveness of anonymization in double-blind review. *Communications of the ACM*, 61:30 – 33, 2018.

Sergei Ivanov. ICML 2020. Comprehensive analysis of authors, organizations, and countries., 2020. `https://medium.com/criteo-engineering/icml-2020-comprehensive-analysis-of-authors-organizations-and-countries-c4d1bb847fde` [Last Accessed: 3/15/2022].

Steven Jecmen, Hanrui Zhang, Ryan Liu, Nihar B. Shah, Vincent Conitzer, and Fei Fang. Mitigating manipulation in peer review via randomized reviewer assignments. In *NeurIPS*, 2020.

Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

Silvia Knobloch-Westerwick, Carroll J. Glynn, and Michael Huge. The Matilda effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*, 35(5):603–625, 2013. doi: 10.1177/1075547012472684. URL `https://doi.org/10.1177/1075547012472684`.

Ari Kobren, Barna Saha, and Andrew McCallum. Paper matching with local fairness constraints. In *ACM KDD*, 2019.

N. Lawrence and C. Cortes. The NIPS Experiment. `http://inverseprobability.com/2014/12/16/the-nips-experiment`, 2014. [Online; accessed 11-June-2018].

Ann M. Link. US and Non-US Submissions: An Analysis of Reviewer Bias. *JAMA*, 280(3):246–247, 07 1998. ISSN 0098-7484. doi: 10.1001/jama.280.3.246. URL `https://doi.org/10.1001/jama.280.3.246`.

Samuel Madden and David DeWitt. Impact of double-blind reviewing on SIGMOD publication rates. *ACM SIGMOD Record*, 35(2):29–32, 2006.

Emaad Manzoor and Nihar B Shah. Uncovering latent biases in text: Method and application to peer review. In *AAAI*, 2021.

Brian C Martinson, Melissa S Anderson, and Raymond De Vries. Scientists behaving badly. *Nature*, 435 (7043):737–738, 2005.

Robert K. Merton. The Matthew effect in science. *Science*, 159(3810):56–63, 1968. doi: 10.1126/science.159.3810.56. URL https://www.science.org/doi/abs/10.1126/science.159.3810.56.

Andrew Myers. Condorcet internet voting service. https://civs1.civs.us/, 2003. [Online; accessed March-2021].

Syavash Nobarany, Kellogg S Booth, and Gary Hsieh. What motivates people to review articles? The case of the human-computer interaction community. *Journal of the Association for Information Science and Technology*, 67(6):1358–1371, 2016.

Ritesh Noothigattu, Nihar Shah, and Ariel Procaccia. Loss functions, axioms, and peer review. *Journal of Artificial Intelligence Research*, 2021.

QS. QS world university rankings by subject 2021: Computer Science and Information Systems, 2021a. https://www.topuniversities.com/university-rankings/university-subject-rankings/2021/computer-science-information-systems [Last Accessed: 10/15/2021].

QS. QS world university rankings by subject 2021: Economics & Econometrics, 2021b. https://www.topuniversities.com/university-rankings/university-subject-rankings/2021/economics-econometrics [Last Accessed: 10/15/2021].

David B Resnik, Christina Gutierrez-Ford, and Shyamal Peddada. Perceptions of ethical problems with scientific journal peer review: An exploratory study. *Science and engineering ethics*, 14(3):305–310, 2008.

Seán Roberts and Tessa Verhoef. Double-blind reviewing at EvoLang 11 reveals gender bias. *Journal of Language Evolution*, 1:163–167, 07 2016. doi: 10.1093/jole/lzw009.

Margaret W. Rossiter. The Matilda effect in science. *Social Studies of Science*, 23(2):325–341, 1993. URL http://www.jstor.org/stable/285482.

Nihar Shah, Behzad Tabibian, Krikamol Muandet, Isabelle Guyon, and Ulrike Von Luxburg. Design and analysis of the NIPS 2016 review process. *JMLR*, 19(1):1913–1946, 2018.

Nihar B Shah. An overview of challenges, experiments, and computational solutions in peer review. Communications of the ACM (to appear). Preprint available at http://bit.ly/PeerReviewOverview, July 2021.

Richard Snodgrass. Single- versus Double-blind reviewing: An analysis of the literature. *SIGMOD Record*, 35:8–21, 09 2006. doi: 10.1145/1168092.1168094.

Flaminio Squazzoni and Gandelli Claudio. Saint Matthew strikes again. An agent-based model of peer review and the scientific community structure. *Journal of Informetrics*, 6:265–27, 04 2012. doi: 10.1016/j.joi.2011.12.005,.

Ivan Stelmakh, Charvi Rastogi, Nihar B Shah, Aarti Singh, and Hal Daumé III. A large scale randomized controlled trial on herding in peer-review discussions. *arXiv preprint arXiv:2011.15083*, 2020.

Ivan Stelmakh, Nihar Shah, and Aarti Singh. PeerReview4All: Fair and accurate reviewer assignment in peer review. *JMLR*, 2021a.

Ivan Stelmakh, Nihar Shah, Aarti Singh, and Hal Daumé III. A novice-reviewer experiment to address scarcity of qualified reviewers in large conferences. In *AAAI*, 2021b.

Ivan Stelmakh, Nihar Shah, Aarti Singh, and Hal Daumé III. Prior and prejudice: The novice reviewers' bias against resubmissions in conference peer review. In *CSCW*, 2021c.

Ivan Stelmakh, Charvi Rastogi, Ryan Liu, Shuchi Chawla, Federico Echenique, and Nihar B Shah. Cite-seeing and reviewing: A study on citation bias in peer review. *arXiv preprint arXiv:2203.17239*, 2022.

Mengyi Sun, Jainabou Barry Danfa, and Misha Teplitskiy. Does double-blind peer review reduce bias? Evidence from a top computer science conference. *Journal of the Association for Information Science and Technology*, 2021.

Warren Thorngate and Wahida Chowdhury. By the numbers: Track record, flawed reviews, journal space, and the fate of talented authors. *Advances in Intelligent Systems and Computing*, 229, 01 2013. doi: 10.1007/978-3-642-39829-2_16.

Leanne Tite and Sara Schroter. Why do peer reviews decline to review? A survey. *Journal of epidemiology and community health*, 61:9–12, 02 2007. doi: 10.1136/jech.2006.049817.

Andrew Tomkins, Min Zhang, and William Heavlin. Reviewer bias in single- versus double-blind peer review. *Proceedings of the National Academy of Sciences*, 114:201707323, 11 2017. doi: 10.1073/pnas.1707323114.

Anthony KH Tung. Impact of double blind reviewing on SIGMOD publication: A more detail analysis. *ACM SIGMOD Record*, 35(3):6–7, 2006.

Boya Xie, Zhihong Shen, and Kuansan Wang. Is preprint the future of science? A thirty year journey of online preprint services. *ArXiv*, abs/2102.09066, 2021.

A Survey details for Q2.

Target audience selection Recall that our objective in target audience selection is to find reviewers for each paper whose research interests intersect with the paper, so that we can survey these reviewers about having seen the corresponding papers outside of reviewing contexts. We describe the exact process for target audience selection in EC and ICML.

In EC, the number of papers posted online before the end of the review process was small. To increase the total number of paper-reviewer pairs where the paper was posted online and the reviewer shared similar research interests with the paper, we created a new paper-reviewer assignment. For the new paper-reviewer assignment, for each paper we considered at most 8 members of the reviewing committee that satisfied the following constraints as its target audience—(1) they submitted a positive bid for the paper indicating shared interest, (2) they are not reviewing the given paper.

In ICML, a large number of papers were posted online before the end of the review process. So, we did not create a separate paper-reviewer assignment for surveying reviewers. Instead, in ICML, we consider a paper's reviewers as its target audience and queried the reviewers about having seen it, directly through the reviewer response form.

Survey question. For research question Q2, we conducted a survey to measure the visibility of papers submitted to the conference and posted online before or during the review process. We describe the details of the survey for EC 2021 and ICML 2021 separately. In EC 2021, we created a specialised reviewer-specific survey form shared with all the reviewers. Each reviewer was shown the title of five papers and asked to answer the following question for each paper:

"Have you come across this paper earlier, outside of reviewing contexts?"

In the survey form, we provided examples of reviewing contexts as "reviewing the paper in any venue, or seeing it in the bidding phase, or finding it during a literature search regarding another paper you were reviewing." The question had multiple choices as enumerated in Table 4, and the reviewer could select more than one choice. If they selected one or more options from (b), (c), (d) and (e), we set the visibility to 1, and if they selected option (a), we set the visibility to 0. We did not use the response in our analysis, if

the reviewer did not respond or only chose option (f). In Table 4, we also provide the number of times each choice was selected in the set of responses obtained.

In ICML 2021, we added a two-part question corresponding to the research question Q2 in the reviewer response. Each reviewer was asked the following question for the paper they were reviewing:

> "Do you believe you know the identities of the paper authors? If yes, please tell us how."

Each reviewer responded either *Yes* or *No* to the first part of the question. For the second part of the question, table 5 lists the set of choices provided for the question, and a reviewer could select more than one choice. If they responded *Yes* to the first part, and selected one or more options from (a), (d), (e) and (f) for the second part, then we set the visibility to 1, otherwise to 0. In Table 5, we also provide the number of times each choice was selected in the set of responses that indicated a visibility of 1.

B Analysis procedure details

In this section we provide some more details of the analysis procedure.

B.1 Kendall's Tau-b statistic

We describe the procedure for computing Kendall's Tau-b statistic between two vectors. Let n denote the length of each vector. Let us denote the two vectors as $[x_1, x_2, \ldots, x_n] \in \mathbb{R}^n$ and $[y_1, y_2, \ldots, y_n] \in \mathbb{R}^n$. Let P denote the number of concordant pairs in the two vectors, defined formally as

$$P = \sum_{\substack{(i,k)\in[n]^2 \\ i<k}} \left(\mathbb{I}\left(x_i > x_k\right)\mathbb{I}\left(y_i > y_k\right) + \mathbb{I}\left(x_i < x_k\right)\mathbb{I}\left(y_i < y_k\right)\right).$$

Following this, we let the number of discordant pairs in the two vectors be denoted by Q, defined as

$$Q = \sum_{\substack{(i,k)\in[n]^2 \\ i<k}} \left(\mathbb{I}\left(x_i > x_k\right)\mathbb{I}\left(y_i < y_k\right) + \mathbb{I}\left(x_i < x_k\right)\mathbb{I}\left(y_i > y_k\right)\right).$$

Observe that the concordant and discordant pairs do not consider pairs with ties in either of the two vectors. In our data, we have a considerable number of ties. To account for ties, we additionally compute the following statistics. Let A_x and A_y denote the number of pairs in the two vectors tied in exactly one of the two vectors as

$$A_x = \sum_{\substack{(i,k)\in[n]^2 \\ i<k}} \mathbb{I}\left(x_i = x_k\right)\mathbb{I}\left(y_i \neq y_k\right) \qquad \text{and} \qquad A_y = \sum_{\substack{(i,k)\in[n]^2 \\ i<k}} \mathbb{I}\left(x_i \neq x_k\right)\mathbb{I}\left(y_i = y_k\right).$$

List of choices for question in Q2 survey	Count
(a) I have NOT seen this paper before / I have only seen the paper in reviewing contexts	359
(b) I saw it on a preprint server like arXiv or SSRN	51
(c) I saw a talk/poster announcement or attended a talk/poster on it	22
(d) I saw it on social media (e.g., Twitter)	4
(e) I have seen it previously outside of reviewing contexts (but somewhere else or don't remember where)	29
(f) I'm not sure	24

Table 4: Set of choices provided to reviewers in EC in Q2 survey and the number of times each choice was selected in the responses obtained. There were 449 responses in total, out of which 92 responses indicated a visibility of 1.

List of choices for question in Q2 survey	Count
(a) I was aware of this work before I was assigned to review it.	373
(b) I discovered the authors unintentionally while searching web for related work during reviewing of this paper	47
(c) I guessed rather than discovered whose submission it is because I am very familiar with ongoing work in this area.	28
(d) I first became aware of this work from a seminar announcement, Archiv announcement or another institutional source	259
(e) I first became aware of this work from a social media or press posting by the authors	61
(f) I first became aware of this work from a social media or press posting by other researchers or groups (e.g. a ML blog or twitter stream)	52

Table 5: Set of choices provided to reviewers in ICML in Q2 survey question and the number of times each choice was selected in the set of responses considered that self-reported knowing the identities of the paper authors outside of reviewing contexts. There were a total of 635 such responses that indicated a visibility of 1. Recall that for ICML, we consider the set of responses obtained for submissions that were available as preprints on arXiv. There were 1934 such submissions.

Finally, let A_{xy} denote the number of pairs in the two vectors tied in both vectors, as

$$A_{xy} = \sum_{\substack{(i,k)\in[n]^2 \\ i<k}} \mathbb{I}\left(x_i = x_k\right) \mathbb{I}\left(y_i = y_k\right).$$

Observe that the five statistics mentioned above give a mutually exclusive and exhaustive count of pairs of indices, with $P + Q + A_x + A_y + A_{xy} = 0.5n(n-1)$. With this setup in place, we have the Kendall's Tau-b statistic between $[x_1, x_2, \ldots, x_n] \in \mathbb{R}^n$ and $[y_1, y_2, \ldots, y_n] \in \mathbb{R}^n$ denoted by τ as

$$\tau = \frac{P - Q}{\sqrt{\left(P + Q + A_x\right)\left(P + Q + A_y\right)}}. \tag{2}$$

This statistic captures the correlation between the two vectors.

Input : Samples $v_i, \alpha_i, \widetilde{t}_i$ for $i \in [N]$, iteration count γ.
(1) Compute the test statistic T defined in equation 1.
(2) For $z \leftarrow 1$ to γ:
(i) For all $b \in \{1, 2, 3\}$: Let V_b denote the number of responses with $v_i = 1$ in bin b. Take all the responses in bin b and reassign each response's visibility to 0 or 1 uniformly at random such that the total number of responses with a visibility of 1 remains the same as V_b.
(ii) Using the new values of visibility in all bins, recompute the test statistic in equation 1. Denote the computed test statistic as T_z.
Output : P value $= \frac{1}{\gamma}\sum_{z=1}^{\gamma} \mathbb{I}(T_z - T > 0)$.

Algorithm 1: Permutation test for correlation between papers' visibility and rank.

B.2 Permutation test

The test statistic T in equation 1 gives us the effect size for our test. Recall from equation 1 that the test statistic T is defined as:

$$T = \frac{N_1\,\tau_1 + N_2\,\tau_2 + N_3\,\tau_3}{N_1 + N_2 + N_3},$$

where for each bin value $b \in \{1, 2, 3\}$, we have N_b as the number of responses obtained in that bin, and τ_b represents the Kendall Tau-b correlation between visibility and rank in the responses obtained in that bin. To analyse the statistical significance of the effect, we define some notation for our data. Let N denote the total number of responses. For each response $i \in [N]$ we denote the visibility of the paper to the reviewer as $v_i \in \{0, 1\}$ and the rank associated with response i as $\alpha_i \in \mathbb{N}_{<0}$. Finally, we denote the bin associated with response i as $\widetilde{t_i} \in \{1, 2, 3\}$. With this, we provide the algorithm for permutation testing in Algorithm 1.