Exploring blinding in peer review system

Fu Zixuan

Research proposal

December 29, 2024

1 Introduction

An eye for an eye, makes the whole world blind. **?**

2 Related Literature

There is a large literature on the topic of peer review, upon which a new scientific field is built called ”Journalogy”.

Many empirical evidence focuses on detecting bias, evaluating fairness, measuring quality of the peer review under different blinding mechanism, notably single-blind and double-blind. Interestingly, as tabulated and summarized in the book chapter by Largent and Snodgrass (2016), experimental and statistical results are divided. And there is no congruence even til now with respect to the effect or non-effect of peer review practice Blank (1991); Tomkins et al. (2017).

Tan (2018) has provided a summary of the strength and weakness of different referee practices. For example, single-blind may result in reviewers favoring big names compared to double-blind where author's identifying information is removed. Yet in double-blind, there is the question of true anonymity and the cost of preparing paper for review. Snodgrass (2007) lists a set of 6 benefits against 21 potential costs of double-blind while mentioning that the complete masking of authorship is not feasible in practice. Some pushes for a more open process with varying degree of transparency ranging from open feedback to open referee identity. Inevitably, there are concerns from all sides (editor, reviewer, author) about the maximal openess in the process. Therefore, I especially appreciate the project by Soergel et al. (2013), which gives birth to the platform OpenReview.net that supports a myriad of configurations such that journal and conferences can experiment with different dimension of open scholarship. [1]

Finally, I intend to

[1] *”The word ”open” denotes access to information. To characterize a system, then, we must state who has Open Scholarship and Peer Review access to what information, and when. (Additionally there may be special conditions on that access).”*

3 Model

Consider a peer review system with one journal editor and one reviewer. The system is described as follows.

1. A paper is submitted to the journal editor. It has a true quality of either low or high, represented by $h_i \in \{0, 1\}$.
2. The author of the paper has a binary type as well denoted by $a_i \in \{o, e\}$. The notation here stands for *old and new* which we can think of as
   - *old*: more experienced, well-known in the research field
   - *new*: less experienced researchers
3. Having received the paper, the editor needs to make a decision on acceptance $\delta_i \in \{0, 1\}$. He wants to accept good papers ($\delta_i = 1$ when $h_i = 1$) and reject bad papers. However, the editor does not know the true quality of the paper.
4. Since she won't evaluate it herself, she delegates the referee work to a reviewer.
5. The reviewer reads the paper more or less carefully by exerting some effort e. He then gives recommendations about whether to accept or reject.
6. The reviewer can not fully discover the true quality of the paper. The probability that the reviewer recommends acceptance when the paper is good is denoted by p_1 and rejection when the paper is bad by p_0.

$$\mathbb{P}(\delta_i = 1 \mid h_i = 1) = p_1,$$

$$\mathbb{P}(\delta_i = 0 \mid h_i = 0) = p_0.$$

   - For the editor, the higher these probabilities are, the more accurate her decision is.
   - For the reviewer, he needs to spend effort e to discover the true quality h_i. Both p_1 and p_0 increases with effort level. i.e.,

$$\frac{\partial p_1(e)}{\partial e} > 0, \quad \frac{\partial p_0(e)}{\partial e} > 0.$$

7. The editor follows the reviewer's recommendation exactly.

3.1 Reviewer

Reviewer's effort The reviewer exerts effort to review papers. The cost of effort is measured in time, denoted by $c(e)$, with $\frac{\partial c(e)}{\partial e} > 0$.

The reviewer faces two reviewing systems $\{\text{Double blind}, \text{Single blind}\}$. Under the double-blind system, the author's name is removed from the paper under referee, while under single-blind, the anonymity comes from the reviewer's side. I first discuss the two extreme cases.

In a world where the paper can stay truly anonymous in double-blind [2] the reviewer assigns the same effort e_b to every paper. Given N papers, the total cost is

$$Nc(e_b).$$

On the other hand, if reviewer can see the author's name and other information in single-blind, I assume that in this way he knows exactly the author's type $a_i \in \{o, e\}$. Since it is of our natural tendency to trust more experienced people, reviewer would exert less effort checking the paper (proof, theorem, etc.) if he knows that the author is of type o. Therefore, I have

$$e_o < e_n.$$

If there are $k\%$ authors of type o, the total cost of effort is

$$N\left(kc(e_o) + (1-k)c(e_n)\right).$$

It is worth mentioning that in reality however, even under double-blind, full anonymity is often not achieved. The reviewer simply knows or can guess the author. Let us say that $\lambda\%$ of the time reviewer still knows the authors' type. Then the total cost is

$$N\left[\lambda\big(kc(e_o) + (1-k)c(e_n)\big) + (1-\lambda)c(e_b)\right].$$

The parameter λ measures the level of non-anonymity in a research field. A larger λ represents a case where it is easy for the reviewer to find out who the author is.

Reviewer's indifference Reviewer does not receive monetary award for referee work, this altruistic action arises out of professional ethics. Meanwhile, he also faces time constraints (e.g., own research, teaching, personal time) thus won't spend as much effort as possible on refereeing. I assume that they assign **a fixed amount of time** to reviewing since there is no incentive to increase or decrease the time spent. They are indifferent between double-blind and single-blind because the total cost of effort is the same under the two mechanisms, i.e.,

$$\lambda_{sq}\big(kc(e_o) + (1-k)c(e_n)\big) + (1-\lambda_{sq})c(e_b) = kc(e_o) + (1-k)c(e_n)$$

where λ_{sq} represents the level of non-anonymity at status quo under double-blind system.

[2] Non-existence of pre-print, no talk has been given etc.

Reviewer's bias As mentioned before, if the reviewer knows the identity of the author, he may exert more or less effort depending on the type. From the editor's perspective, it is always better to have the reviewer exerts more effort by regarding the author as type n. In reality, it may also introduce bias when he knows the author's type. To be more specific, let us define a new probability $p_1(e,a), p_0(e,a)$ which depends on the effort level of reviewer and the reviewer's knowledge of author's type. The bias that occurred with type o is $p_0(e,o) = p_0(e) - b_o$,which translates to that condition on the paper being low quality, the reviewer is somehow more tolerant if he knows the author is of type o. While for type e, the bias takes place in the form of $p_1(e,n) = p_1(e) - b_n$, which means that conditional on good paper, reviewer is more skeptical of accepting it. All in all, I have

$$p_1(e,a) = \begin{cases} p_1(e) & \text{if } a = o, \\ p_1(e) - b_n & \text{if } a = n, \end{cases}$$

$$p_0(e,a) = \begin{cases} p_0(e) - b_o & \text{if } a = o, \\ p_0(e) & \text{if } a = n. \end{cases}$$

3.2 Editor

Editor's objective The editor acts in the interest of the journal. She wants to accept good papers and reject bad ones. Recall that the decision is δ_i and quality h_i. There are 4 types of outcomes that can enter the editor's objective function.

	$h_i = 1$	$h_i = 0$
$\delta_i = 1$	$h_i\delta_i$	$(1-h_i)\delta_i$
$\delta_i = 0$	$h_i(1-\delta_i)$	$(1-h_i)(1-\delta_i)$

Table 1

Consider the case where the editor maximizes the number of correctly accepted and correctly rejected paper, that is, $\sum_i h_i\delta_i$ and $\sum_i(1-h_i)(1-\delta_i)$. The objective function is

$$\max \mathbb{E}[\sum_i \delta_i h_i] + \tau\mathbb{E}[\sum_i (1-\delta_i)(1-h_i)] \tag{1}$$

which is equivalent to

$$\max \{\mathbb{P}(\delta_i = 1|h_i = 1)\mathbb{P}(h_i = 1) + \tau\mathbb{P}(\delta_i = 0|h_i = 0)\mathbb{P}(h_i = 0)\}$$

Editor's payoff I first define some institutional setting before presenting editor's payoff under different scenarios.

- α is the proportion of high quality paper $\mathbb{P}(h_i = 1)$.

- β_o is the proportion of good paper from o author and β_n from n author, satisfying

$$k\beta_o + (1-k)\beta_n = \alpha$$

Now I write the editor's payoff under different scenarios.

- Base case: I assume that without reading the paper, or delegating anyone to read the paper, the editor accepts and rejects with equal probability ($\mathbb{P}(\delta_i = 1|h_i) = 1/2$). [3] Her payoff in this case is

$$1/2\alpha + \tau 1/2(1-\alpha)$$

- Double blind with $\lambda = 0$: If the full anonymity can be achieved under double-blind system ($\lambda = 0$), the payoff is

$$p_1(e_b)\alpha + \tau p_0(e_b)(1-\alpha)$$

- Single blind with $\lambda = 1$: Under single-blind, the payoff conditioning on the author's type a_i is

$$\begin{aligned}
&\mathbb{E}[\delta_i h_i + \tau(1-\delta_i)(1-h_i)|a_i] \\
=&\mathbb{P}(\delta_i = 1|h_i = 1, a_i)\mathbb{P}(h_i = 1|a_i) + \tau\mathbb{P}(\delta_i = 0|h_i = 0, a_i)\mathbb{P}(h_i = 0|a_i) \\
=&p_1(e_a, a)\beta_a + \tau p_0(e_a, a)(1-\beta_a)
\end{aligned}$$

  Therefore, the total payoff is

$$k\left[p_1(e_o)\beta_o + \tau(p_0(e_o) - b_o)(1-\beta_o)\right] + (1-k)\left[(p_1(e_n) - b_n)\beta_n + \tau p_0(e_n)(1-\beta_n)\right]$$

- Double blind with $0 < \lambda < 1$: the editor's payoff is a linear combination of the two cases above.

3.3 Problem Formulation

The institutional parameters are $\{k, \lambda_{sq}, \alpha, \beta_o, \beta_n, \tau\}$. The editor's preference parameter is τ while the reviewer's bias is $\{b_o, b_n\}$. The reviewer solves the following (indifference) problem to determine the effort level e_b^*, e_o^*, e_n^*.

$$\lambda_{sq}\big(kc(e_o) + (1-k)c(e_n)\big) + (1-\lambda_{sq})c(e_b) = kc(e_o) + (1-k)c(e_n) \tag{2}$$

Given the e_b^*, e_o^*, e_n^*, the editor solves the following (maximization) problem to determine the optimal mechanism $\in \{\text{double-blind}, \text{single-blind}\}$.

$$\begin{aligned}
\max_{\lambda \in \{\lambda_{sq}, 1\}} &\lambda\left(k\left[p_1(e_o)\beta_o + \tau(p_0(e_o) - b_o)(1-\beta_o)\right] + (1-k)\left[(p_1(e_n) - b_n)\beta_n + \tau p_0(e_n)(1-\beta_n)\right]\right) \\
&+ (1-\lambda)\left(p_1(e_b)\alpha + \tau p_0(e_b)(1-\alpha)\right)
\end{aligned} \tag{3}$$

[3] This is not saying that $\delta_i = 1/2$ is the optimal decision rule under ignorance.

Therefore, emulating the principal-agent problem, I write

$$\begin{aligned} \max_{\lambda\in\{\lambda_{sq},1\},e_b,e_n,e_o} \quad & \text{editor's payoff} \\ \text{s.t.} \quad & \text{reviewer's indifference} \end{aligned} \tag{4}$$

4 Analysis

The reviewer's indifference condition simplifies to

$$c(e_b) = kc(e_o) + (1-k)c(e_n).$$

Since the decision on λ^* is binary, the editor's optimisation boils down to comparing the first and second term in equation 3. Rearranging the terms, I only need tocompare

$$k\beta_o p_1(e_o)+(1-k)\beta_n p_1(e_n)+\tau(k(1-\beta_o)p_0(e_o)+(1-k)(1-\beta_n)p_0(e_n)) \underbrace{-k\tau b_o(1-\beta_o)-(1-k)\tau b_n\beta_n}_{\text{bias}}$$

with

$$\alpha p_1(e_b) + \tau(1-\alpha)p_0(e_b).$$

I first make comparison without taking into account the bias term.

Linear case Starting with the simplest, I assume that $c(e) = p_1(e) = p_0(e) = e$. Comparing $k\beta_o p_1(e_o) + (1-k)\beta_n p_1(e_n)$ and $\alpha p_1(e_b)$ boils down to

$$\begin{bmatrix} k\beta_o & (1-k)\beta_n \end{bmatrix} \left(\begin{bmatrix} e_0 \\ e_n \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} k & 1-k \end{bmatrix} \begin{bmatrix} e_0 \\ e_n \end{bmatrix} \right) \overset{?}{\leq} 0$$

$$k\beta_o(k-1) + (1-k)\beta_n k \Leftrightarrow \beta_n - \beta_o \overset{?}{\leq} 0.$$

By definition $\beta_n < \beta_o$, which implies that the second term is larger than the first term even without bias. Adding back the bias, editor prefers double-blind undoubtedly.

General case My intuition is that if $c''(e) > p''(e)\ \forall e$, then even without the bias, the editor prefers double-blind mechanism. To begin with, I take $p(e) = e$ and $c(e) = 1/2e^2$. I can write the condition as

$$\begin{bmatrix} k\beta_o & (1-k)\beta_n \end{bmatrix} \left(\begin{bmatrix} p(e_0) \\ p(e_n) \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} p(e_b) \right) \overset{?}{\leq} 0$$

while

$$e_b^2 = ke_o^2 + (1-k)e_n^2 \Leftrightarrow e_b = k^*e_o + (1-k^*)e_n \text{ for some } k^* < k$$

The condition is more negative in this case.

$$k\beta_o(k^*-1) + (1-k)\beta_n k^* < \beta_n - \beta_o < 0.$$

Now I assume that $c(x) = \log(x+1)$ and $p(x) = x$, then the fact that $e_b = k^* e_o + (1 - k^*)e_n$ for some $k^* > k$ implies that the condition $k\beta_o(k^* - 1) + (1-k)\beta_n k^*$ is ambiguous. If β_n is not too small compared to β_o (the proportion of good paper from n author is not that different from o author), and the bias b_o, b_n not too large, then it is possible that the editor prefers single-blind mechanism with $\lambda = 1$.

Proposition 4.1. If $c''(e) \geq p''(e)$ for all e, the editor always prefers double-blind mechanism subject to a status quo level of anonymity λ_{sq}. When it is not the case and that the author's productivity difference, reviewer's bias are not too large, editor may prefer single-blind mechanism.

While I was expecting that the status quo anonymity λ_{sq}[4] would play a role in the editor's decision $\lambda \in \{\lambda_{sq}, 1\}$, it turns out that it doesn't matter at all under the model I set up. This is a question to be investigated in the next section.

5 Discussion

In this section, I discuss several aspects of the model that I have thought about but ignored in the previous section.

Non-anonymity level λ_{sq} Several papers have provided evidence that for some research field, there's no point bothering with double-blind. Because reviewers are assigned papers in their respective fields, chances are that they have already seen the work as preprint in archive, working paper or attended the author's talk themselves. This is especially true if the field is niche. This evidence suggests that intuitively speaking, a higher level of λ_{sq} should favor single-blind.

As discussed in section 4, the editor chooses double-blind for sure if full anonymity gives a higher payoff than full-disclosure. However, even under double-blind, full anonymity is not achieved because there exists an inherent non-anonymity level λ_{sq} for each research field. All other things being equal, a research field where it is easier for papers to stay truly blind (low λ_{sq}) should more likely end up in a double-blind system than a field where everyone knows everyone (high λ_{sq}). To incorporate the idea into a model, I should introduce a fixed cost c for the editor if she implements a double-blind mechanism. The fixed cost spent in anonymizing papers[5] is the same for every field. Therefore, even if

$$\begin{aligned}\alpha p_1(e_b) + \tau(1-\alpha)p_0(e_b) > & k\beta_o p_1(e_o) + (1-k)\beta_n p_1(e_n) \\ & + \tau(k(1-\beta_o)p_0(e_o) + (1-k)(1-\beta_n)p_0(e_n)) \\ & \underbrace{-k\tau b_o(1-\beta_o) - (1-k)\tau b_n \beta_n}_{\text{bias}}\end{aligned}$$

[4] i.e., how difficult it is for the paper to be truly blind

[5] This is more than just removing names but also verifying if there's any revealing information in the paper

the anonymization cost may be too large to make double-blinding beneficial if the weight $(1-\lambda_{sq})$ on the larger term $\alpha p_1(e_b) + \tau(1-\alpha)p_0(e_b) >$ is small. By incorporating a fixed anonymization cost, the editor's choice of mechanism will depend on the status quo non-anonymity level λ_{sq} (as I wish).

Reviewer's indifference condition I have assumed reviewer is indifferent between the two mechanism since they do not get compensation or reward for their time spent, as in reality. This indifference condition serves as a constraint in the combined problem 4. However, I argue that this condition is not enough to pin down the effort level (e_b, e_o, e_n) given all the parameters specified before. Because for either $\lambda \in \{\lambda_{sq}, 1\}$ the editor's payoff is maximized when effort is as high as possible. Yet there's no constraint on the upper bound of $c(e)$. It should be imposed e.g., $kc(e_o) + (1-k)(e_n) < \bar{c}$ so that all decision variable λ, e_b, e_o, e_n can be pinned down.

Another extension is to introduce reviewer's incentive in their effort decision. Currently, since both mechanism *blind* reviewer's identity such that they only get "We thank the anonymized reviewers for their helpful feedback." as reward. It would be interesting to think about rewarding reviewer in a non-pecuniary way. Propositions include credit system, promotion scheme etc. Or we can imagine a world where there's openness to the reviewer's feedback or even revelation of his identity. If reviewer is incentivized in some ways, the original question would take on the flavor of the classical principal-agent problem. In this case, the editor (principal) will juggle between many things in her choice of mechanism.

Editor's objective function I have briefly alluded to the 4 types of editor's concern as in table 1. Intuitively, the editor wants to maximize diagonal terms and minimized the off-diagonal ones. I argue that it is enough to take 2 out of 4 such that the there exists a countervailing pair $h_i\delta_i$ and $-\delta_i$.[6]

[6] By labeling the element in the table horizontally, I can take $(1,2), (1,4), (2,4), (2,3), (4,3)$. Taking only 4 also works if I treat the two forces (accept good and reject bad) equally.

References

Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the american economic review. *The American Economic Review*, pages 1041–1067.

Largent, E. A. and Snodgrass, R. T. (2016). Blind peer review by academic journals. *Blinding as a solution to bias: strengthening biomedical science, forensic science, and law*, pages 75–95.

Snodgrass, R. T. (2007). Single-versus double-blind reviewing.

Soergel, D., Saunders, A., and McCallum, A. (2013). Open scholarship and peer review: a time for experimentation .

Tan, M. H. (2018). Peer review—past, present, and future. In *Medical and Scientific Publishing*, pages 55–68. Elsevier.

Tomkins, A., Zhang, M., and Heavlin, W. D. (2017). Reviewer bias in single-versus double-blind peer review. *Proceedings of the National Academy of Sciences*, 114(48):12708–12713.