When do referees shirk in a peer review process?

Sahana Roy Chowdhury[*]

International Management Institute - Kolkata, India

Abstract

This note obtains conditions for existence of shirking referees in peer review process. When referees are heterogeneous say, bad (b) and good (g), only for a medium range of probability of getting a good paper p, both referees prefer reading over shirking. It never happens that b reads while g shirks. Both prefer 'shirking and rejecting (accepting)' if p is low (high) enough. The paper shows that a two-referee cross-examination review reduces the error of accepting a bad paper, though for a small range of probability.

Keywords: peer review system; refereeing; shirking
JEL Classification Codes: A20, D01

1. Introduction

There has been rising concerns over the efficiency of peer review process, as illuminated by Blank (1991), Hamermesh (1994), Ellison (2002). Inefficiencies crop up due to various reasons: 'market structure' as Atal (2010) adds to the literature, more often because there is hardly any easy means to monitor a referee. Under such circumstances, the possibility that a paper sent to the referee may not be attended at all and desk rejected - can't be negated. Ineffi- ciencies may be due to poor completion rate or long review time: Mrowinski et al. (2016) shows in an empirical exercise that the completion rate of reviewers known personally by the editor is very high, which means that they are much more likely to finish the review (hence become more efficient in the peer review process) than other reviewers.

   The present paper sheds light on the referee's credibility, rather than his efficiency, while accepting the facts that refereeing is not costless as it involves precious time and effort while there is no credible threat to defy him shirking. To the best of our knowledge, there is no

---

[*] E-mail: sahana.isi@gmail.com.

discussion on a shirking referee and the no-shirking conditions. The theory, even if it exists, is difficult to be supported empirically!

In Section2 the paper constructs the model first, with a single homogeneous referee case, then considering heterogeneous types, lastly with a two-type matching cross-examination review. Section 3 provides concluding remarks.

2. The model

Let $v_H > 0$ ($v_L < 0$) be, in the referee's perception, the value addition to the literature by a good (bad) paper. This may be attributed to the perception that the quality or rank of a journal might go down (up) by publishing a bad (good) paper. Referees derive utility from any potential positive contribution to the literature to which he is also a contributor. Let us suppose that referees are homogeneous (which will be relaxed later).

Let p be the probability of getting matched with a good paper in the peer review process. Then the (expected) utility derived from various actions are:
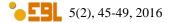
$$U = \begin{cases} pv_H - c & \text{if he reads (R)} & \text{(i)} \\ pv_H + (1-p)v_L & \text{if shirks and accepts (NR-A)} & \text{(ii)} \\ 0 & \text{if shirks and rejects (NR-R)} & \text{(iii)} \end{cases}$$

Only on reading a paper the referee understands whether it has any positive value to add[1]. When matched with a good paper with probability p, if he reads, he accepts it[2] and he derives utility v_H as it gets published adding value to the literature. When he reads a bad paper, matched with probability $(1 - p)$, he chooses to reject it. Thus 'negative contribution' is avoided by reading a paper while an effort cost $c > 0$ is forgone. Also, we assume $c < v_H$; otherwise, no referee will ever read.

In case he shirks, he invites the possibility of a negative contribution by accepting a bad paper to which he is matched with probability $(1 - p)$, on the other hand, rejecting a paper while shirking risks the rejection of a good one[3].

Proposition 1. *(i) If c is not high i.e.* $\frac{c}{|v_L|} < \left(1 - \frac{c}{v_H}\right)$, *then for all* $p < p_1 = \frac{c}{v_H}$ *he never reads and simply rejects. For very high p,* $p > p_2 = \left(1 - \frac{c}{|v_L|}\right)$ *he always shirks and accepts. Only for a medium range of probability* $p \in (p1, p2)$ *he prefers reading over shirking. (ii) If* $\frac{c}{|v_L|} > \left(1 - \frac{c}{v_H}\right)$ *he never reads, always shirks and rejects (accepts) for all* $p < (>)\hat{p} = \frac{1}{1 + \frac{v_H}{|v_L|}}$.

---

[1] By this we rule out possibility of Type I and Type II errors caused solely by his 'in- efficiency'. However, this is a simplifying assumption. In real life it may so happen that either 'shirking/not reading' or 'inefficiency' might have actually caused the Type I or Type II errors, both errors being costly to the society. George Akerlof first submitted his "The Market for Lemons" in 1967 which got rejected three times before publication. He mentioned that, "I do think its early rocky reception did have an eect on my own work. I believe I would have done follow-up work sooner, if I had not been made to feel lucky just to have it published at all" (Gans and Shepherd, 1994).
[2] We assume, for mathematical simplicity, that 'revise and resubmit' for re-consideration is akin to acceptance by the referee.
[3] Rejecting a good paper may often become costly to the referee so that NR-R gives him a payoff $-pv_H$. In that case it can be checked that all the propositions still hold qualitatively. Here for mathematical simplicity we eliminate any such cost.

Intuitively, too low a probability for getting a good paper with positive effort cost puts lesser incentive to read, also under such circumstances, rejecting while shirking gives zero utility whereas accepting provides a negative utility.

Presumably, a good journal has a high probability of getting good paper. Simultaneously, a bad journal has a low probability. But such a set up (even under $c/|v_L| < (1 - (c/v_H))$) can not negate possibilities of shirking and 'all-time rejection' for a bad journal that has a bleak chance of attracting good submission! This questions the existence of such a journal. To bridge the gap let us bring heterogeneity of referees. Suppose, there are two types of referees: good (*g*) and bad (*b*), who differ in value perceptions only. While a good paper is unanimously valued as good ($v_H{}^g = v_H{}^b$), a bad paper gives more disutility to *g* than it gives to *b* ($v_L{}^g < v_L{}^b$).

Let us see how adding good referee might increase the range of probability for reading (p_1, p_2), which we find only when cost of reading is not that high to eliminate all referees from reading. Note that p_1 will be same for both but $p_2{}^b < p_2{}^g$. From this it follows:

Proposition 2. *(i) Under $\left\{\frac{c}{v_L^g}, \frac{c}{v_L^b}\right\} < \left(1 - \frac{c}{v_H}\right)$ if $p < p_1 < p_2^b$ holds, both referees prefer reading over shirking. Both shirk and reject (accept) if p is low (high) enough, $p < p_1 \left(> p_2^g\right)$. (ii) When $\frac{c}{|v_L^g|} < \left(1 - \frac{c}{v_H}\right) < \frac{c}{|v_L^b|}$ only g reads for $p \in \{p_1, p_2^g\}$. (iii) It never happens that b reads while g shirks; simultaneously, it never happens that b rejects a paper while g accepts it.*

Intuitively, once decided to shirk, the good referee having a higher threshold probability for accepting a paper over rejecting it (since he derives more disutility from a bad paper getting published), never chooses to accept it when the bad type rejects.

2.1. Two-referee cross-examination

It may be of interest to find if cross-examination by two referees does better. Here each referee takes account of the other referee's action to fix his own decision. Here we assume: { $c/v_L{}^g$, $c/v_L{}^b$ } < (1 − (c/v_H)). Now, let us fix the rule of the game as follows:
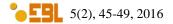
Reject: if any one of the referees rejects; Accept: if both accept

Note that shirking and accepting here may have higher pay-off (in case the other referee reads) compared to the single referee case, since the pay-off for the former is pv_H[4] while that for the latter is $pv_H + (1 - p)v_L$. If the type of referees are same (*b*, *b*) or (*g*, *g*) the outcome is same as a single referee case with *b* and *g* respectively. Two-type-referee case is thus more interesting to discuss.

For $p < p_1$ both *b* and *g* have no incentive to defect from *status quo* (say, SQ - single referee case where he chooses NR-R under $p < p_1$). This is because, if he rejects and the other rejects as well, he is no well-off; if he accepts and other rejects he is no well-off, but if other accepts as well then the paper will get accepted and he will be worse-off. Thus the outcome is same as in single referee case,

For $p_1 < p < p_2{}^b$ the SQ for both is R. Here both know that the other will never choose NR-R since it would make him worse-off. So the choice is between NR-A and R. If any of the referees chooses NR-A while the other reads he can save his cost with the same outcome.

---

[4] This is because, the other player might accept it on reading, which occurs with proba- bility p, or reject with probability $(1 - p)$ that give the expected value as pv_H. In the single referee case he might accept a wrong paper adding the possibility of negative contribution and thus an expected pay-off of $pv_H + (1 - p)v_L$.

However, if both choose NR-A then there is a possibility of adding negative value. Thus the pay-off functions are as follows:

Table 1. Pay-off functions.

		g R	NR-A
b	R	$pv_H - c$, $pv_H - c$	$pv_H - c$, pv_H
	NR-A	pv_H, $pv_H - c$	$pv_H + (1-p)v_L{}^b$, $pv_H + (1-p)v_L{}^g$

Clearly, we get 2 NEs here: (R,NR-A) and (NR-A,R) and outcome: accept with probability p, as in single referee case. Note that here both types do not read simultaneously for this range of probability, unlike the single referee case.

Refer to the table above. For $p_2{}^b < p < p_2{}^g$ we get (NR-A,R) as the NE and outcome: accept with probability *p*. Here the problem that a paper may get accepted on not reading (what we find under single referee scenario with a *b* type referee) becomes less intense, in other words the Type-II error (accepting a bad paper) gets reduced.

For $p > p_2{}^g$ we find (NR-A,NR-A) as the unique NE with the outcome accept, as in single referee case and cross-examination doesn't help. Thus, if we assume that Type-II error gets reduced if at least one of the referees reads and that makes the review process more efficient then we reach the following proposition:

Proposition 3. *When the two referees are of two different types cross-examination or double referee examination helps in reducing errors in review only under $p_2^b < p < p_2^g$.*

3. Conclusion

This note tries to add in the literature that analyses inefficiencies of peer re- view process. It searches for the possibility that under what condition the referee chooses to shirk over read. Introducing heterogeneity of referees the paper then obtains the no-shirking conditions for the good and the bad type. Interestingly, only for a medium range of probability of getting a good paper *p* both referees prefer reading over shirking. It never happens that *b* reads while *g* shirks, and also it never happens that *b* rejects a paper while *g* accepts it. Both shirk and reject (accept) if *p* is low (high) enough, however it depends upon the effort cost and the value perceptions. Good journals thus have lower critical probability for blind rejection (NR−R), because the effort in reviewing a paper submitted to a good journal is lower as it emphasizes certain standard submission practices (e.g, brevity, lucidity in expression, explanation and representation). On introducing a two-referee cross-examination for reviewing, the paper obtains that similar-type referee matching gives same outcome as under single review case. Type-II error can be reduced by cross-examination, although under a small range of *p*.

The model introduced here is very simplistic but can be extended to capture more realistic observations. It is an initiative towards theorizing the no-shirking conditions for the referees in a peer-review process, which has so far remained a less explored area in the literature.

References

Atal, V. (2010) Do journals accept too many papers?, *Economics Letters*, 107(2), 229-232.

Blank, R.M. (1991) The effects of double-blind versus single-blind reviewing: experimental evidence from the American Economic Review, *American Economic Review*, 81(5), 1041-1067.

Ellison, G. (2002) The slowdown of the economics publishing process, *Journal of Political Economy*, 110(5), 947-993.

Gans, J.S. and Shepherd, G.B. (1994) How are the mighty fallen: rejected classic articles by leading economists, *Journal of Economic Perspectives*, 8(1), 165-179.

Hamermesh, D.S. (1994) Facts and myths about refereeing, *Journal of Economic Perspectives*, 8(1), 153-163.

Mrowinski, M.J., Fronczak, A., Fronczak, P., Nedic, O. and Ausloos, M. (2016) Review time in peer review: quantitative analysis and modelling of editorial workflows, *Scientometrics*, 107(1), 271-286.