Can Editors Compensate for the "Luck of the Reviewer Draw" Effect in Peer Review? An Agent-Based Model*

Flaminio Squazzoni[¤], Francisco Grimaldo[ǂ], Juan Bautista Cabotà[ǂ]

[¤] GECS-Department of Economics and Management, University of Brescia, Via San Faustino 74/B, 25122 Brescia, Italy, squazzon@eco.unibs.it

[ǂ] Departament d'Informàtica, Universitat de València, Avinguda de la Universitat, s/n, 46100 Burjassot-València, Spain, {francisco.grimaldo, juan.cabota}@uv.es

Abstract. This paper investigates the impact of various editorial options of author/referee matching and referee behaviour for the quality and efficiency of peer review. We built various scenarios where referees could behave randomly or follow different cheating strategies to outperform potential competitors, and editors could use different strategies for referee selection. In case of random behaviour of referees, any editorial option may have a negative effect. On the other hand, accurate matching by editors may reduce the negative effect of referee cheating and limit the negative effect of excessive competition.

Keywords: Peer review; referees; referee selection; author/referee matching; journal editors; agent-based model.

1 Introduction

Peer review has been recently under the spotlight for cases of misconduct that imply distorted allocation of funds and reputation in the science system (e.g., Crocker and Cooper 2011). Biased referee behaviour and inappropriate referee selection criteria (e.g., Couzin 2006), as well as the influence of the growing competition pressures on editors' and referees' judgment have been discussed (e.g., Smith 2006). In this situation, understanding the mechanics of the peer review process and the influence of motivations and behaviour of scientists on the quality and efficiency of the process is urgent (e.g., Leek, Taub and Pineda 2011; Squazzoni and Takács 2011; Squazzoni, Bravo and Takács 2013).

Empirical analysis can hardly understand the mechanics of peer review and its systemic implications (Edmonds et al. 2011). Our paper aims to contribute to recent agent-based research on peer review (e.g., Roebber and Schultz 2011; Thurner and Hanel 2011; Allesina 2012; Grimaldo and Paolucci 2013) by looking at the impact of referee behaviour and different editorial options for referee selection on the quality and efficiency of the process.

The structure of the paper is as follows: in the second section, we introduce the model and some simulation scenarios; in the third one, we discuss the simulation results and,

* This work was funded by the Department of Economics and Management of the University of Brescia, the Spanish MICINN and the European Commission FEDER funds, under grant TIN2009-14475-C04.

in the concluding section, we summarize the main findings and discuss limitations and developments.

2 The model

Following Squazzoni and Gandelli (2012, 2013), we assumed a population of *n* scientists (*n* = 200) randomly selected to fill one of two roles: author or referee. The task of an author was to submit an article with the goal of having it accepted to be published. The task of a referee was to evaluate the quality of author submissions. Authors and referees were randomly matched 1 to 1. As informed by the referees' opinion, only the best *p* submissions were published.

Each agent had $R \in \mathbb{R}$ resources, needed for submissions and reviewing, which were initially homogeneous (R(0)=1), then accumulated according to agent's publication score. Each agent had a expected quality that was dependent on agent resources, and it was calculated as follows:

$$\mu = \frac{v * R}{v * R + 1} \quad (1)$$

where *v* indicated the velocity at which the expected quality increased with the increase of agent resources. The quality of author submissions (Q) followed a normal distribution $N(\mu, \sigma)$, with a standard deviation σ proportion to the expected quality (μ), which was calculated as σ= μ*b, with b represents the bias percentage. This means that, with some probability, top scientists could write average or low quality submissions, and average scientists had some chance to write good submissions. Successful publication multiplied author resources by a value *m*, which varied between 1.5 for less productive published authors and 1 for more productive published authors. This was seen as mimicking reality, where publication is more important for scientists at the initial stages of their academic careers and cannot infinitely increase for top scientists. If not published, following the "winner takes all" rule characterizing science, we assumed that authors lost all resources invested prior to submitting.

The chance of being published was determined by evaluation scores assigned by referees. We assumed that reviewing was a resource-intensive activity and agents resources determined the cost to the reviewer (i.e., time lost for publishing their own work). The total expense *S* for a referee was calculated as follows:

$$S = \frac{1}{2} R_r [1 + (Q - \mu_r)] \quad (2)$$

where R_r was the referee's resources, Q was the real quality of the author's submission and μ_r was the referee's expected quality. Reviewing expenses grew linearly with the quality of authors' submissions. Top scientists would be expected to spend less time reviewing in general, as they have more experience and are better able to evaluate sound science than are average scientists. They will lose more resources than average scientists, however, because their time is more valuable than the latter.

To test the effect of the different editorial options in combination with the different referees behaviour, we first created four different scenarios of the way through which this behaviour was assigned: In the first scenario, called "*random behaviour*", referees had a fifty percent constant probability of being biased in their judgment. When reliable, they did the best they could to provide an evaluation which truly reflected the quality of the submission. In this case, they spent all needed resources for reviewing. When unreliable, referees under- or over- estimated author submissions and spent half of the resources ideally spent by reliable referees.

In the second scenario, called "*cheating*", referees estimated submission authors' resources (R_a) and identified each author with an expected R_a similar or higher than theirs as a competitor, following a threshold function. Referees tended to outperform potential competitors by systematically underrating their submission, even at their own expenses (e.g., resources spent for reviewing). In the third scenario, called "*local competition*", scientists detected possible competitors only in their own performance neighbourhood. This was to mimic certain fragmented scientific communities where scientists tend to compete locally within or across similar groups. In this case, we assumed that competitor's detection followed a normal distribution $N(R_r, \sigma_2)$ where R_r indicated referee resources and σ_2 was the standard deviation which was calculated as a proportion of R_r. Finally, in the last scenario, called "*glass ceiling*", scientists tried to outperform the less and the more productive colleagues. This was to mimic a situation where scientists protect against upstart scientists and outperform superior scientists. In this case, competitors' detection function followed a logistic shape.

We used these scenarios as a baseline to test different editorial options to match authors and referees. In the "editorial decision 0", authors and referees were randomly matched as if editors would lack knowledge of referee expertise. In the "editorial decision 1" scenario, authors were matched with referees of a similar productivity. In the "editorial decision 2" scenario, authors were matched with referees of higher productivity, while in the "editorial decision 3" authors were matched with referees of lower productivity.

The quality of peer review was measured as the percentage of errors made by referees by calculating the optimal situation, in which submissions were published according to their real value, and by measuring the discrepancy with the actual situation in each simulation step (evaluation bias). The efficiency was measured as the percentage of resources wasted by unpublished authors who deserved to be published (productivity loss) and the percentage of resources spent by referees compared with the resources invested by submitting authors (reviewing expenses).

3 Results

Data were averaged across 10 simulation runs of 200 simulation steps. For the shortage of space, here we reported only relevant results with strongly competitive publication selection rate ($p = 0.25$, which means 25% of submissions eventually published in each simulation step), though we explored different values of this parameter (e.g., $p = 0.50, 0.75$). In case of *random behaviour* and *local competition*, any editorial option of referee selection determined more evaluation bias, higher productivity lost and similar or higher reviewing expenses than the random matching. The situation was different with *cheating* and *glass ceiling* scenarios. In these cases, editorial decisions 1 and 2 could significantly lower evaluation bias and productivity loss compared with random matching. In terms of resources distribution, random behaviour and "local competition" generally generated higher inequality, except with editorial decision 1.

Scenario	Evaluation bias	Productivity loss	Review expenses	Gini index	Percentage of cheaters
Random behaviour					
Editorial dec. 0	29.42	15.00	29.42	0.47	
Editorial dec. 1	39.55	19.56	34.43	0.37	
Editorial dec. 2	32.99	16.22	30.87	0.43	
Editorial dec. 3	29.51	15.71	29.47	0.46	
Cheating					
Editorial dec. 0	70.86	34.72	35.24	0.28	0.27
Editorial dec. 1	51.97	25.69	35.19	0.33	0.25
Editorial dec. 2	61.95	29.81	34.60	0.30	0.19
Editorial dec. 3	73.00	36.92	34.86	0.29	0.32
Local competition					
Editorial dec. 0	31.04	15.63	30.13	0.45	0.20
Editorial dec. 1	57.87	28.61	35.70	0.31	0.41
Editorial dec. 2	36.54	17.74	31.85	0.41	0.22
Editorial dec. 3	33.47	17.37	30.06	0.44	0.18
Glass ceiling					
Editorial dec. 0	70.35	34.70	34.56	0.29	0.34
Editorial dec. 1	58.02	28.56	35.64	0.32	0.38
Editorial dec. 2	65.88	32.26	35.23	0.30	0.37
Editorial dec. 3	68.21	34.47	34.29	0.29	0.36

Table 1. The effect of different editorial decisions on the peer review process. Note that the Gini index measured the inequality of distribution of resource in the system.

It is worth noting that differences in the competitors' detection mechanism had a considerable effect on the percentage of cheaters. It was generally higher in the "*glass ceiling*" scenarios, while the highest value was reached in the "*local competition*" scenario combined with "editorial decision 1", i.e., when authors were matched with referees of similar qualities. Furthermore, it is worth noting that the evaluation bias in various scenarios was not univocally correlated with the number of cheaters in the population.

4 Conclusions

We investigated possible editorial counteractions to reduce misbehaviour of referees, such as matching authors and referees by looking at reputation, in our case ideally synthesized by agent resources. We found that, in case of complete randomness of

referee judgment, any editorial option may have even a negative effect. If referees behave strategically, bias tends to increase but certain editorial options might have a counteractive effect, such as matching referees of similar or higher quality than submission authors. The side-effect of exploiting referees' effort is to generate benefits to published authors, who might gain cumulative publication advantages.
Finally, we found that peer review outcomes are significantly sensitive to differences in the way scientists identify their competitors. A "man is man's wolf" competitive scenario increases the chances of referee bias. Our results showed that certain mechanisms, such as the stratification of scientists in local competing groups and the presence of niches of competition, might reduce the negative effect of cheating and excessive competition. On the other hand, if the competition between scientists is stratified and refers to local groups, the potentially positive effect of editorial options tends to decrease.

References

1. Allesina, S.: Modeling Peer Review: An Agent-Based Approach. Ideas in Ecology and Evolution, 5 (2), 27-35 (2012)
2. Couzin, J.: … and how the Problems Eluded Peer Reviewers and Editors. Science, 311, 614-615 (2006)
3. Crocker, J., Cooper, M. L.: Addressing Scientific Fraud. Science, 334 (6060), 1182 (2011)
4. Edmonds, B., Gilbert, N., Ahrweiler, P. and Scharnhorst, A.: Simulating the Social Processes of Science. JASSS- The Journal of Artificial Societies and Social Simulation, 14 (4) 4 <http://jasss.soc.surrey.ac.uk/14/4/4.html> (2011)
5. Grimaldo, F. and Paolucci, M.: A Simulation of Disagreement for Control of Rational Cheating in Peer Review. Advances in Complex Systems, forthcoming (2013)
6. Leek, J. T., Taub, M. A., Pineda, F. J.: Cooperation between Referees and Authors Increase Peer Review Accuracy. PLoS ONE, 6 (11): e26895: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0026895>.
7. Roebber, P. J. and Schultz, D. M.: Peer Review, Program Officers and Science Funding. Plos One, 6 (4), e18680, accessible at: <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0018680> (2011)
8. Smith, R.: Peer Review. A Flawed Process at the Heart of Science and Journals. Journal of the Royal Society of Medicine, 99, 759-760 (2006)
9. Squazzoni, F., Bravo, G., Takács, K.: Does Incentive Provision Increase the Quality of Peer Review? An Experimental Study. Research Policy, 42, 1, 287-294 (2013)
10. Squazzoni, F. and Gandelli, C.: Saint Matthews Strikes Again. An Agent-Based Model of Peer Review and the Scientific Community Structure. Journal of Informetrics, 6, 265-275 (2012)
11. Squazzoni, F. and Gandelli, C.: Opening the Black Box of Peer Review. An Agent-Based Model of Scientist Behaviour. JASSS- The Journal of Artificial Societies and Social Simulation, 16(2) 3: http://jasss.soc.surrey.ac.uk/16/2/3.html (2013)
12. Squazzoni, F. and Takács, K.: Social Simulation that 'Peers into Peer Review'. JASSS 14 (4) 3 <http://jasss.soc.surrey.ac.uk/14/4/3.html> (2011)
13. Thurner, S. and Hanel, R.: Peer Review in a World with Rational Scientists: Toward Selection of the Average. The European Physical Journal B, 84, 707-711 (2011)