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

Flaminio Squazzoni*, Francisco Grimaldo‡, Juan Bautista Cabotà‡

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

‡ 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 modelled the peer review process as an interaction of editors, authors and referees characterized by intrinsic knowledge asymmetry and subject to evaluation bias. We built various scenarios where referees could behave randomly or follow cheating strategies to outperform potential competitors and editors could use different strategies for referee selection. We also tested different ways through which referees classify authors as potential competitors. We found that there is no editorial best way 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 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 various cases of misconduct that imply dramatically distorted allocation of funds and reputation in the science system (e.g., Suls and Martin 2009; Bornmann 2011; Crocker and Cooper 2011). Among the most important sources of disappointment, analysts included biased referee behaviour and inappropriate referee selection criteria (e.g., Couzin 2006; Bailey, Hermanson and Louwers 2008), not to mention the influence of the growing competition pressures on editors' and referees' judgment bias (e.g., Smith 2006). The popularized idea of peer review as a "black box" or the widely acknowledged "the luck of the reviewer drawn" principle have posed serious concerns on the quality and legitimization of the peer review process (e.g., Cole, Cole and Simon 1981). This made the need for better understanding the mechanics of the peer review process and the influence of motivations and behaviour of scientists an urgent matter (e.g., Squazzoni and Takács 2011).

Most of the problems depend on the fact that the peer review process is a complex strategic interaction where behaviour of scientists in various roles (e.g., journal editors, authors and referees) may generate different and unpredictable collective outcomes in terms of quality and efficiency of the evaluation process, which are strongly sensitive to interactional and contextual factors (e.g., Leek, Taub and Pineda 2011, Squazzoni, Bravo and Takács 2013). Unfortunately, empirical analysis can hardly understand the mechanics of this interaction and examine its implications at the system level (Edmonds *et al.* 2011).

To fill this gap, a growing literature has recently grown around agent-based models of peer review that aimed at understanding collective outcome of peer review starting from hypothesized scientist behaviour and simulating complex interaction situations (e.g., Roebber and Schultz 2011; Thurner and Hanel 2011; Allesina 2012; Grimaldo and Paolucci 2013).

Our paper aims to contribute to this line of research by presenting an agent-based model of peer review that examines the impact of referee behaviour and different editorial options for referee selection on the quality and efficiency of the peer review process. We also looked at the influence of competitive spirits of scientists. We started from the idea that peer review is intrinsically subject to knowledge asymmetry and imperfection, but also strongly sensitive to scientist strategies and decisions. The combination of these strategies, as well as editorial decisions that often match authors and referees independently of their mutual level of expertise, might be an important source of bias and influence the resource allocation in the science system.

The rest of the paper is structured as follows: in the second section, we introduce the model, in the third, we illustrate some simulation scenarios, while in the fourth one, we discuss the simulation results. 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 each 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. As informed by the referees' opinion, only the best submissions were published (i.e., those exceeding the publication rate).

We gave each agent a set of resources which were initially homogeneous ($R_a$=1). We assumed that resources were needed both to submit and review an article. With each simulation step, agents were endowed with a fixed amount of resources $f$, equal for all (e.g., common access to research infrastructure and internal funds, availability of PhD. students, etc.). They then accumulated resources according to their publication score.

We assumed that the quality of submissions $\mu$ varied and was dependent on agent resources. Resources should be viewed as ideal academic status, reputation, and career among scientists. Each agent had $R_a \in \mathbb{R}$ resources, from which we derived an expected submission quality as follows:

$$\mu = \frac{v * R_a}{v * R_a + 1} \qquad (1)$$

where $v$ indicated the velocity at which the quality of the submission increased with the increase of author resources. For instance, this means that for $v$ = 0.1 each agent needed $R_a$ = 10 to reach a medium-sized quality submission ($\mu$ = 0.5).

We assumed that authors varied in terms of the quality of their output depending on their resources. More specifically, the quality of submissions by authors followed a standard deviation $\sigma$ which proportionally varied according to agent resources and followed a normal distribution $N(\mu, \sigma)$. 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. We assigned a heterogeneous value of $m$ after various explorations of the parameter space. This was seen as mimicking reality, where publication is crucial in explaining differences in scientists' performance, but is more important for scientists at the initial stages of their academic careers and cannot infinitely increase for top scientists. Thus, the resources of published authors grew accordingly, leading to subsequent submissions of presumably higher quality. 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. The value of author submissions was therefore not objectively determined (i.e., it did not perfectly mirror the real quality of submissions), but was instead dependent on the referees’ opinion. We assumed that reviewing was a resource-intensive activity and that agent resources determined both the agent’s reviewing quality and the cost to the reviewer (i.e., time lost for publishing their own work). The total expense $S$ for any referee was calculated as follows:

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

where $R_r$ was the referee’s resources, $Q_a$ was the real quality of the author’s submission and $\mu_r$ was the referee’s expected quality. This last was calculated as in equation (1). It is worth noting that, when selected as referees, agents not only needed to allocate resources toward reviewing but also potentially lost additional resources as a result of not being able to publish their own work in the meantime.

We assumed that authors and referees were randomly matched 1 to 1 so that multiple submissions and reviews were not possible and the reviewing effort was equally distributed among the population. We assumed that reviewing expenses grew linearly with the quality of authors’ submissions. If referees were matched with a submission of a quality close to a potential submission of their own, they allocated 50% of their available resources toward reviewing. They spent fewer resources when matched with lower quality submissions, more when matched with higher quality submissions. Reviewing expenses, however, were proportionally dependent on agent resources, meaning that 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.

## 3. Simulation scenarios

In the first baseline scenario, called “*random behaviour*”, we assumed that referees had a constant probability of being biased in their judgment. When fair, referees provided a consistent and unequivocal opinion which truly reflected the quality of the submission. In this case, they did the best they could to provide an accurate evaluation and spent all needed resources for reviewing. We assumed that referees estimated the authors’ resources following a normal distribution of the actual authors’ resources and a narrow standard deviation ($\sigma = R_a/10$). Then, they estimated the author submissions’ quality as in (1). This meant that the evaluation scores by fair referees were likely to approximate the real value of author submissions.

In order to mimic typical knowledge and information asymmetries between authors and referees which characterize peer review, we assumed that even in any case there was a chance for some evaluation bias ($b$ = *0.1*), and that $b$ increased in proportion to the difference between the submission’s estimated quality by referees and the actual submission quality.

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.

In the case of unfairness, referees fell into type I and type II errors: recommending submissions of low quality to be published or recommending against the publishing of submissions which should have been published (e.g., Laband and Piette 1994). More specifically, when unfair, referees spent fewer resources than did fair referees, and under- or over- estimated author submissions (i.e., $u = 0.1$ and $o = 1.9$).

In the second baseline scenario, called "*cheating*", we assumed that referees tended to outperform potential competitors by systematically underrating their submission, even at their own expenses (e.g., resources spent for reviewing). More specifically, we assumed that referees were capable to estimate submission authors' resources ($R_a$) and identify each author with an expected $R_a$ similar or higher than his/her $R_r$ as a competitor.

We used these two 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. This perfectly mimics the "luck of the reviewer drawn" situation, where good quality authors could be matched with more or less fair reviewers by mere chance. Then, we assumed that editors had full information on the potential quality of their pool of referees and used this information to decide how to match authors and referees. 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. These last two scenarios mimicked situations where editors could exploit the willingness of high quality scientists to contribute to the reviewing process (i.e., "*editorial selection 2*") or where young scholars (typically PhD students and post-doc researchers) are more frequently involved.

It is worth noting that, in order to understand implications of referee behaviour in different matching conditions, we tested each scenario with different referee behaviour, i.e., "random behaviour" and "cheating".

## 4. Results

Tab. 1 shows the impact of these different editorial decisions on the quality and efficiency of the peer review process. Data were averaged across 10 simulation runs of 200 simulation steps. For the shortage of space, here we reported only results in case of 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., $p = 0.50, 0.75$).

Results showed that, in case of random behaviour of referees, any editorial option of referee selection determined more evaluation bias than the random matching of authors and referees. The random matching also determined higher productivity lost by unpublished authors who deserved to be published and higher reviewing expenses by referees (see Tab. 1). The situation was different with cheating. In this case, except "*Editorial decision 3*", where authors were matched with referees of lower productivity, editorial options of referee selection could significantly lower evaluation bias compared with random matching (see Tab. 2). When referees had lower productivity than authors, the number of cheaters among referees increased as less advanced scientists tried to outperform more established colleagues by providing unfair judgment and underrating the quality of competitors' submissions (see Tab. 3).

| Scenario | Evaluation bias | Productivity loss | Reviewing expenses |
|---|---|---|---|
| Editorial decision 0 | 29.42 | 15.00 | 29.42 |
| Editorial decision 1 | 39.55 | 19.56 | 34.43 |
| Editorial decision 2 | 32.99 | 16.22 | 30.87 |
| Editorial decision 3 | 29.51 | 15.71 | 29.47 |

**Tab. 1.** The effect of different editorial decisions on the quality and efficiency of the peer review process with random behaviour of referees.

| Scenario | Evaluation bias | Productivity loss | Reviewing expenses |
|---|---|---|---|
| Editorial decision 0 | 70.86 | 34.72 | 35.24 |
| Editorial decision 1 | 51.97 | 25.69 | 35.19 |
| Editorial decision 2 | 61.95 | 29.81 | 34.60 |
| Editorial decision 3 | 73.00 | 36.92 | 34.86 |

**Tab. 2.** The effect of different editorial decisions on the quality and efficiency of the peer review process with cheating by referees...

| Scenario | Percentage of cheaters |
|---|---|
| Editorial decision 0 | 0.27 |
| Editorial decision 1 | 0.25 |
| Editorial decision 2 | 0.19 |
| Editorial decision 3 | 0.32 |

**Tab. 3.** Percentage of cheaters among referees in the editorial scenarios based on “cheating” baseline..

We next calculated the resources of all agents at the end of the simulation run in all scenarios and measured their distribution over the population. Figures 1 and 2 show the resource distribution in scenarios with random behaviour and cheating by referees, respectively. Results showed that cheating generated higher asymmetry in resource distribution. In this case, random matching of authors and referee implied less asymmetry than other matching scenarios. The situation was different with random behaviour of referees. In this case, the distribution was more asymmetric, especially in the random matching scenario. This meant that, if referees behave randomly, more accurate and competent editorial decisions of referee selection might moderate asymmetric distribution of resources.

We also considered inequality of resource distribution by calculating a Gini index, which measured the inequality in the allocation of resources at the system level. Results showed that scenarios with cheating implied less inequality, with the only exception of “*editorial decision 1*”, where authors were matched with referees of similar quality, which approximated results of random scenarios (see Tab. 4).

**Fig. 2.** Productivity distribution in scenarios with random behaviour of referees.

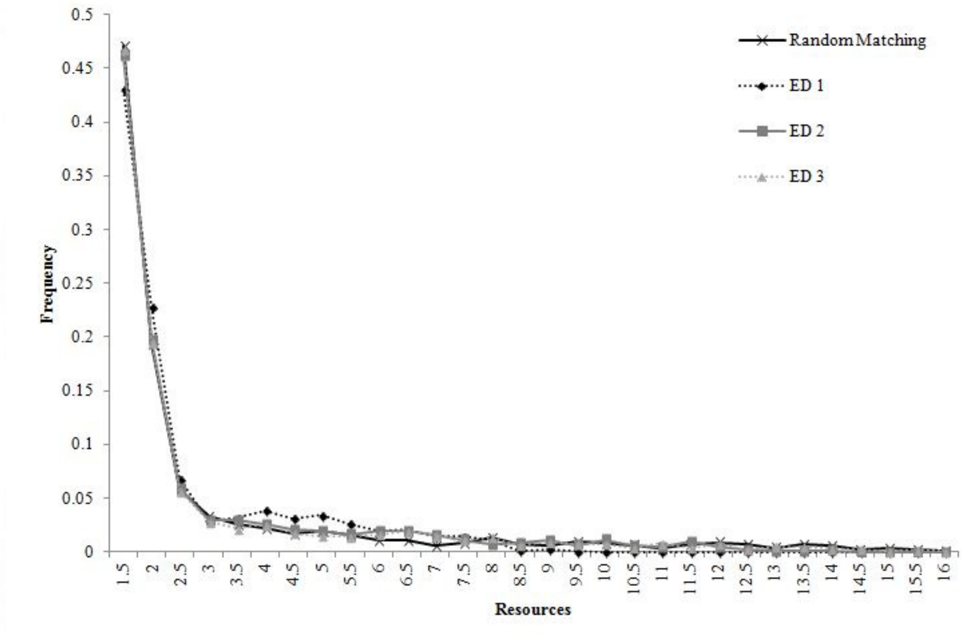
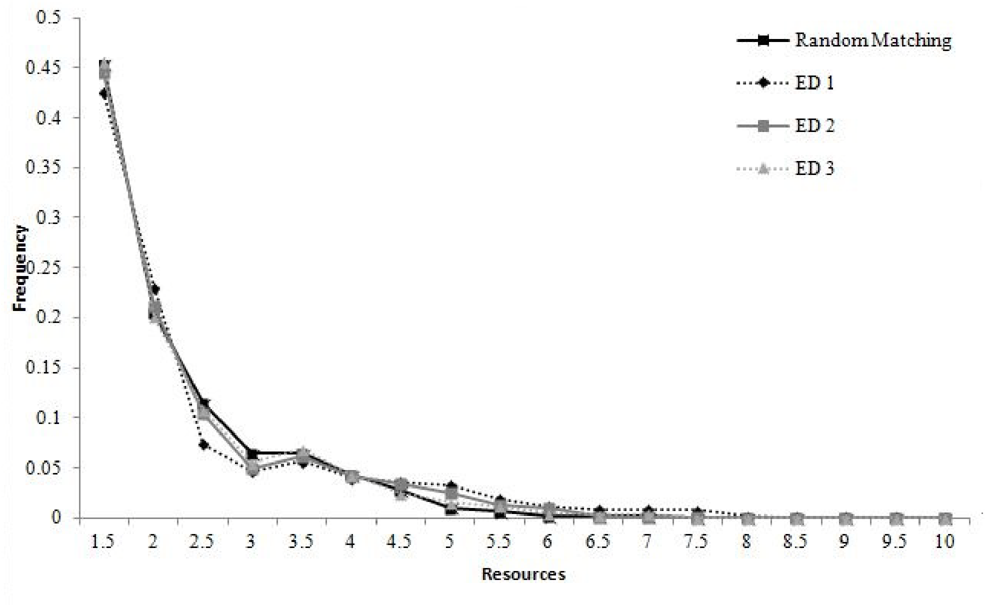



**Fig. 2.** Productivity distribution in scenarios with cheating.



| Scenario | Gini index | |
|---|---|---|
| | Random behaviour of referees | Cheating |
| Editorial decision 0 | 0.47 | 0.28 |
| Editorial decision 1 | 0.37 | 0.33 |
| Editorial decision 2 | 0.43 | 0.30 |
| Editorial decision 3 | 0.46 | 0.29 |

**Tab. 4.** The Gini index (values calculated at the end of the simulation). The index takes 0 when there was complete equality in resource distribution among agents and 1 when a single agent had everything.

To test differences in the way scientists could detect their potential competitors and its combination with editorial options of referee selection, we then created two supplementary scenarios where we modified the way through which cheaters identified competitors. Here, the idea was that in certain scientific communities, especially among the so-called "hard sciences", widely shared objective measures exist that help everyone to precisely commensurate his/her respective performance with that of others. This means that competitors might be precisely identified across the whole population. On the other hand, this does not hold in other scientific communities, especially those revolving around the humanities, where these standards do not exist and are even widely contested. In these cases, the definition of potential competitors depends on the stratification of scientists in local groups, with the prevalence of disciplinary or group specificities of standards (e.g., Laudel and Gläeser 2006).

More specifically, unlike the previous "*cheating*" scenarios, which followed a threshold function to detect possible competitors, we tested a "*local competition*" scenario, where this function followed a Gaussian shape and a "*glass ceiling*" scenario, where this function followed a logistic shape. This means that, in the first case, we assumed that 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. More specifically, we assumed that competitor's detection followed a normal distribution $N(R_r, \sigma_2)$ where $R_r$ was the resources of the referee and $\sigma_2$ was the standard deviation which was calculated as a proportion of $R_r$. In the second case, i.e., "the glass ceiling scenario", we assumed that scientists tried to similarly outperform the less and the more productive colleagues. This was to mimic a situation where scientists protect against upstart scientists and outperform superior scientists.

Results showed that "local competition" ensured significantly better quality when referee selection was random, also ensuring less productivity loss and reviewing expenses (see Tab. 5), while "glass ceiling" generated consistent bias independently of the editorial options. It is also worth noting that, when authors and referees were matched among scientists of similar productivity, the "cheating" scenario comparatively ensured less bias and less inefficiency, although at a significant absolute value (e.g., 51% of bias).

Tab. 6 shows the number of cheaters in the population in all scenarios. It is worth noting that differences in the competitors' detection mechanism had a considerable effect on the number of cheaters at the population level. The number of cheaters was generally higher in the "glass ceiling" scenarios, while the highest number of cheaters 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 number of evaluation bias in various scenarios was not univocally correlated with the number of cheaters in the population (compare Tab. 5 and 6).

It is important to mention that the various options of referee selection were influenced by the concrete availability of required referees in the population in each simulation step. For instance, in some cases, required matching conditions (e.g., matching between an author of a given quality and a referee of higher quality) could not be satisfied as required referees were run out by previous matching. We performed detailed analysis in all scenarios to conclude that these cases influenced around 30-40% of matching in worst cases. However, this also characterizes real constraints faced by editors in the referee selection process, as desired referees are scarce in the population appropriate allocation cannot always take place and often editors might be induced to select non-optimal referees.

| Scenario | Evaluation bias | Productivity loss | Reviewing expenses |
|---|---|---|---|
| *Editorial decision 0* | | | |
| Cheating | 70.86 | 34.72 | 35.24 |
| Local competition | 31.04 | 15.63 | 30.13 |
| Glass ceiling | 70.35 | 34.70 | 34.56 |
| *Editorial decision 1* | | | |
| Cheating | 51.97 | 25.69 | 35.19 |
| Local competition | 57.87 | 28.61 | 35.70 |
| Glass ceiling | 58.02 | 28.56 | 35.64 |
| *Editorial decision 2* | | | |
| Cheating | 61.95 | 29.81 | 34.50 |
| Local competition | 36.54 | 17.74 | 31.85 |
| Glass ceiling | 65.88 | 32.26 | 35.23 |
| *Editorial decision 3* | | | |
| Cheating | 73.00 | 36.92 | 34.86 |
| Local competition | 33.47 | 17.37 | 30.06 |
| Glass ceiling | 68.21 | 34.47 | 34.29 |

**Tab. 5.** The effect of different editorial decisions on the quality and efficiency of the peer review in scenarios with different competitors' detection mechanisms.

| Scenario | Percentage of cheaters |
|---|---|
| *Editorial decision 0* | |
| Cheating | 0.27 |
| Local competition | 0.20 |
| Glass ceiling | 0.34 |
| *Editorial decision 1* | |
| Cheating | 0.25 |
| Local competition | 0.41 |
| Glass ceiling | 0.38 |
| *Editorial decision 2* | |
| Cheating | 0.19 |
| Local competition | 0.22 |
| Glass ceiling | 0.37 |
| *Editorial decision 3* | |
| Cheating | 0.32 |
| Local competition | 0.18 |
| Glass ceiling | 0.36 |

**Tab. 6.** Percentage of cheaters among referees in different competitors' detection scenarios.

As for the previous scenarios, we next calculated the resources of all agents at the end of the simulation run in the "local competition" and "glass ceiling" scenarios and measured their distribution over the population. Figures 3 and 4 show that cheating by referees implied higher asymmetry in the resource distribution in the "local competition" scenarios. In both cases, the "Editorial decision 2" scenarios (where authors were matched with higher quality referees) tended to determine more asymmetric distribution, as published authors could benefit from cumulative advantages of being evaluated by high quality referees.

**Fig. 3.** Productivity distribution in the "local competition" scenario.

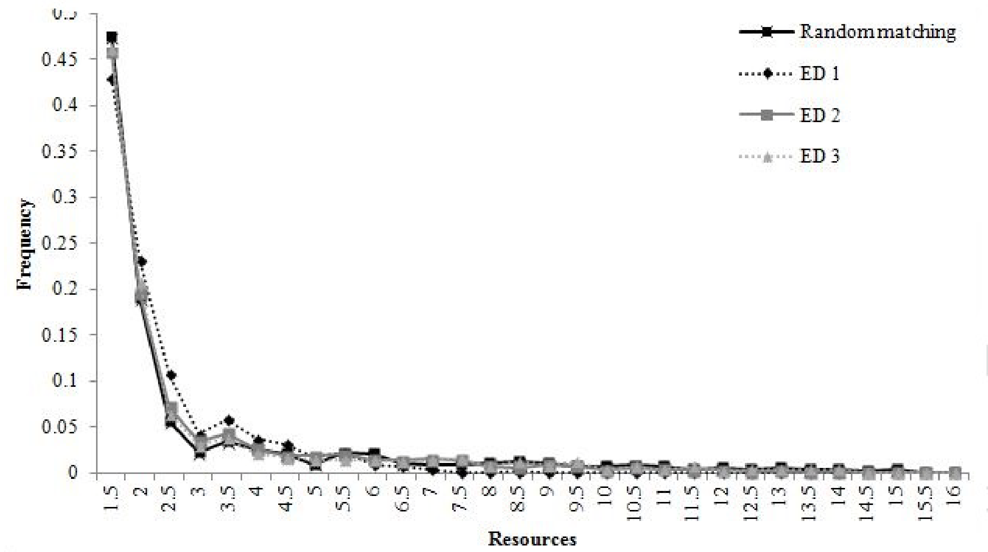
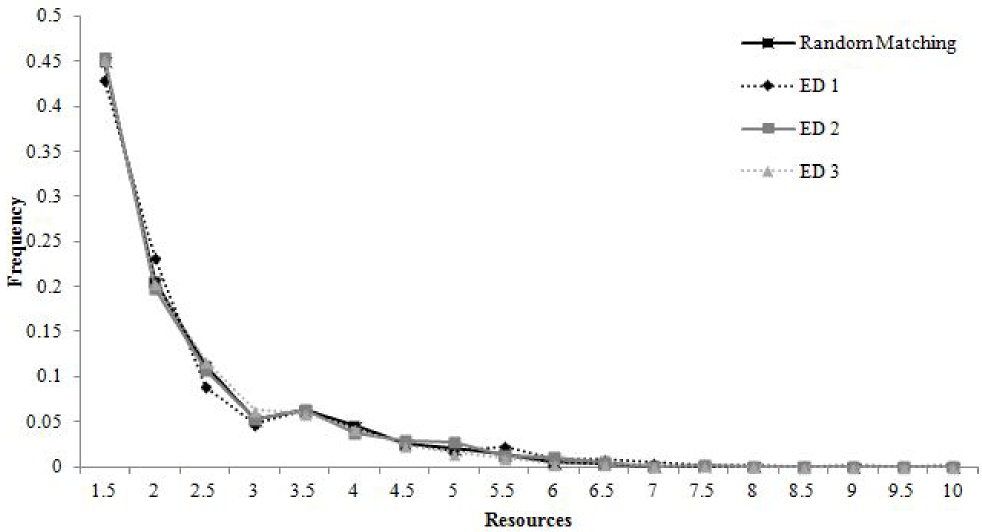



**Fig. 4.** Productivity distribution in the "glass ceiling" scenario.




We considered inequality of resource distribution by calculating a Gini index also in these scenarios. Results showed that generally "local competition" generated higher inequality independently from the author/referee matching. In this case, higher inequality was found in the "Editorial decision 3", where authors of high quality were matched with referees of lower quality. With the exception of "local completion" scenario, random matching of authors and referees generated similar or less inequality than more sophisticated editorial matching strategies.

| Scenario | Gini index | | |
|---|---|---|---|
| | Cheating | Local competition | Glass ceiling |
| Editorial decision 0 | 0.28 | 0.45 | 0.29 |
| Editorial decision 1 | 0.33 | 0.31 | 0.32 |
| Editorial decision 2 | 0.30 | 0.41 | 0.30 |
| Editorial decision 3 | 0.29 | 0.44 | 0.29 |

**Tab. 7.** The Gini index in different competitors' detection scenarios (values calculated at the end of the simulation). The index takes 0 when there was complete equality in resource distribution among agents and 1 when a single agent had everything.

To sum up, our results showed that there is no best way for editorial referee selection. In cases of prevalence of random behaviour of referees, any editorial option of matching may have a negative effect compared with random matching. On the other hand, accurate matching by editors may reduce the negative effect of referee cheating and limit the effect of excessive competition. Indeed, when referees are inclined towards cheating to outperform potential competitors, evaluation bias of peer review could be reduced by competent matching by editors. This said, if strong competitive spirits take place among scientists that bring them to strongly bias their judgment, any good editorial decision is poorly effective.

## 5. Conclusions

Previous studies suggested that even a small proportion of cheaters among the referees may dramatically distort the publication quality (e.g., Thurner and Hanel 2011). This is confirmed by our results, while we added insights on the consequences of this on the system's resource allocation in terms of growth and inequality of distribution. We also investigated possible editorial counteractions to reduce misbehaviour of referees, such as matching authors and referees by looking at competence and reputation, in our case synthesized by agent resources. We found that, in case of complete randomness of referee judgment, any editorial option may have even a negative effect. The situation is slightly different if referee behaviour is sensitive to interaction and follows strategic reasoning, i.e., referees are influenced by certain features of authors (e.g., comparatively higher quality) and might be tempted to provide unfair judgment to outperform competitors. Although in this case the number of potential bias tends to systemically increase, caused by competitive spirits of referees, certain editorial options might have a potential counteractive effect. For instance, matching referees of similar or higher quality than submission authors can reduce bias. This has the side-effect to exploit referees' effort and generate benefits especially for published authors who might gain cumulative advantages exploiting referees' competence and fairness.

Furthermore, we found that peer review outcomes are significantly sensitive to differences in the way scientists identify their competitors. It is widely acknowledged that competition pressures dramatically increased in recent times, with scientists harshly competing for funds at the local, national and international levels. This "man is man's wolf" scenario increases the chances of referee bias, not to mention the likelihood of authors' misbehaviour. 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.

## Acknowledgements

Financial support was provided by the Department of Economics and Management of the University of Brescia as well as by the Spanish MICINN and the European Commission FEDER funds, under grant TIN2009-14475-C04. Usual disclaimers apply.

## References

Allesina, S.: Modeling Peer Review: An Agent-Based Approach. Ideas in Ecology and Evolution, 5 (2), 27-35 (2012)

Bailey, C. D., Hermanson, D. R., Louwers, T. J.: An Examination of the Peer Review Process in Accounting Journals. Journal of Accounting Education, 26, 55-72 (2008)

Bornmann, L.: Scientific Peer Review. Annual Review of Information Science and Technology, 45, 199-245 (2011)

Cole, S., Cole, J. R., Simon, G. A.: Chance and Consensus in Peer Review. Science, 214 (4523), 881-886 (1981)

Couzin, J.: … and how the Problems Eluded Peer Reviewers and Editors. Science, 311, 614-615 (2006)

Crocker, J., Cooper, M. L.: Addressing Scientific Fraud. Science, 334 (6060), 1182 (2011)

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)

Grimaldo, F. and Paolucci, M.: A Simulation of Disagreement for Control of Rational Cheating in Peer Review. Advances in Complex Systems, forthcoming (2013)

Laband, D. N. and Piette, J. M.: Favoritism Versus Search of Good Papers. Empirical Evidence Regarding the Behavior of Journal Editors. Journal of Political Economy, 102, 194-203 (1994)

Laudel, G. and Gläeser, J. Tensions between Evaluations and Communication Practices. Journal of Higher Education Policy and Management, 28, 289-295 (2006)

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>.

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)

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)

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)

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)

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)

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)

Suls, J., Martin, R.: The Air We Breathe: A Critical Look at Practices and Alternatives in the Peer Review Process. Perspectives on Psychological Science, 4, 40-50 (2009)

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)