DO EDITORS HAVE A SILVER BULLET? AN AGENT-BASED MODEL OF PEER REVIEW

Juan Bautista Cabotà
Departament d'Informàtica
Universitat de València
Avinguda de la Universitat, s/n
46100 Burjassot-València
Email: juan.cabota@uv.es

Francisco Grimaldo
Departament d'Informàtica
Universitat de València
Avinguda de la Universitat, s/n
46100 Burjassot-València
Email: francisco.grimaldo@uv.es

Flaminio Squazzoni
Department of Economics and
Management
University of Brescia, Via San
Faustino 74/B 25122 Brescia
Email: squazzon@eco.unibs.it

KEYWORDS

Peer review; editor; referee-author matching policy; referee behaviour; agent-based modelling.

ABSTRACT

This paper presents an agent-based model of peer review that looks at the effect of different editorial policies of referee selection. We tested four author/referee matching scenarios as follows: random selection of referees, selection of referees with a similar status to submission authors, selection of higher-skilled and lower skilled referees. We tested these scenarios against three types of referee behaviour, i.e., fair, unreliable and strategic and measured their implications for the quality and efficiency of the process. Results show that in case of randomness of referee judgment, any editorial policy is detrimental for peer review. If referees behave strategically, certain matching policies, such as selecting referees of good quality, might counteract possible bias.

INTRODUCTION

Peer review is essential to guarantee the quality of scientific publications (e.g., Squazzoni 2010). Previous studies have tried to measure the effect of peer review on the quality of publications and referees' reports as well as time and costs of different review processes (Jefferson et al. 2002). Other studies have examined the impact of peer review on authors' satisfaction (Weber et al. 2002), the motivation of referees (Squazzoni et al. 2013) and the editors' approach (Neff and Olden 2006; Kravitz et al. 2010; Newton 2010).

A recent large-scale international online survey (Mulligan et al. 2012) called for initiatives to increase the quality of of the process. It also indicated the importance of understanding the effect of referee behaviour under different author-referee matching circumstances, such as junior referees reviewing the submissions by senior authors or judgment bias on submissions by young and non prestigious scientists. An analysis on how papers are assigned to referees in a journal (Hamermesh 1994) suggested that except for a very few superstar authors, editors do not usually match authors with referees of similar quality. Another empirical analysis (Callaham and Tercier 2007) showed that academic rank, formal training or status as principal investigator could not help to predict higher-quality reviews. Some previous work tested a simple selection mechanism based on disagreement control that allows to reduce bias due to lower-skilled or unfair referees (Grimaldo et al. 2011; Grimaldo and Paolucci 2013). Unfortunately, there is no systematic study of the editor's role in referee selection. This paper tries to fill this gap by simulating the effect of different kinds of author-referee assignments on the quality and efficiency of the process. On the one hand, our model explores different editorial policies to match referees and authors based on their academic status. On the other hand, it also considers different types of referee behaviour, which may be fair, strategic or random and looks at their implications in terms of productivity, efficiency and resource distribution.

THE MODEL

This section follows the ODD protocol for model documentation (Polhill et al. 2008; Grimm et al. 2010).

Purpose

Our model aims to explore the effects of different referee/author matching policies by journal editors. It also explores interaction effects between these policies and different behaviour of referees. It follows previous studies by Squazzoni and Gandelli (2012, 2013) and Cabotà et al. (2013).

By referee behaviour, we mean the way scientists carry out their reviewing task, i.e. their reviewing effort and their possible intentionality. By analysing different editorial matching policies, we aim to explore possible options that editors might consider to improve the quality of the peer review process.

Entities, state variables and scales

The only agent entity in the model is the scientist. Scientists play one of two possible roles in each simulation step: author or referee. The task of an author is to submit an article with the goal of having it accepted for publication, whereas the task of a referee is to evaluate the quality of the submission she/he is matched to. Table 1 summarizes the attributes that parameterize the characteristics of scientists in the model. The editorial policy is modelled through a pair of state variables referring to the review and publication

processes that are represented by the parameters in Table 2.

Table 1: Scientist Attributes

Attr	Name	Brief description
Id	Identifier	Unique agent identifier
R	Resources	Amount of resources accumulated by scientists. They are a proxy of their academic status, position, experience and scientific achievement
Rl	Role	Role played by agents: author/referee
B	Behaviour	The scientist behaviour as referee: unreliable, fair or cheater

Table 2: Editorial Policy Parameters

Attr	Name	Brief description
p	Publication rate	Percentage of submissions that are published.
M	Matching policy	The way editors assign referees to authors: random, similar quality, higher-skilled or lower-skilled.

The model includes different spatial and temporal scales. The spatial scale indicates the number of scientists and the temporal scale the number of simulation steps. One simulation step represents one round in the peer review process, i.e. the submission of papers by authors, the review of all submissions and the selection of publications (see detail below).

Process overview and scheduling

Here, we describe the different processes carried out by the model in each simulation step, in which half of the scientists are randomly selected to play the role of author, while the others act as referees.

Matching authors and referees
Authors and referees are matched on a one to one basis. Therefore, multiple submissions and reviews are not possible and the reviewing effort is equally distributed among the population. We consider four different author/referee matching policies (*M*) as follows: random, peer, higher-skilled and lower-skilled referees. By matching policy we mean the way editors assign referees to authors (for details see the submodels section).

Authors' submissions
The quality of an author's submission ($Q \in [0, 1]$) is a random number from a normal distribution $N(E,\sigma_a)$ with the tails cut off. The mean of this distribution, $E \in [0, 1]$, is the expected quality of the scientist, which is dependent on agent resources as shown in Equation 1.

The parameter $v \in [0, 1]$ indicates the velocity at which the expected quality increases with the increase of the agent's resources. For instance, for a value of $v = 0.1$, each agent needs $R = 10$ to reach a medium-sized expected quality ($E = 0.5$).

$$E = \frac{v*R}{v*R+1} \qquad (1)$$

The standard deviation of submission quality is calculated as a proportion of the expected quality using Equation 2, where the parameter $av \in [0, 1]$ indicates the level of quality variability when preparing a submission. Indeed, while top scientists could write average or low quality submissions, an average scientist can also write good submissions.

$$\sigma_a = E * av \qquad (2)$$

Note that, by calculating σ_a as a proportion of the author's expected quality, the variability in the quality of submissions depends on the scientist's resources (e.g., it will be higher for well established scientists). Indeed, as senior researchers are usually involved in several research lines (that can be at different phases of development) and collaborate with heterogeneous groups (ranging from PhD students to other senior colleagues), it is reasonable to assume that they can produce submissions of different quality.

Referee behaviour
In the role of referees, scientists can be:
- Unreliable, i.e., they do not take reviewing seriously (e.g. due to lack of time or interest) so that their evaluation does not reflect the actual quality of submissions under evaluation.
- Fair, i.e., they consider reviewing seriously and provide an accurate evaluation, which is likely to approximate the actual value of the submission.
- Cheaters, i.e., they consider reviewing seriously to outperform potential competitors by underrating their submissions, even at their own expenses (i.e. resources spent to justify their negative evaluation).

Submission evaluation
To evaluate a submission, referees first estimate the authors' amount of resources. Although a referee can not know the actual amount of resources of an author (R_a), we assume that she/he can have a perception of this, which we called R'_a. R'_a is calculated as a sample value taken from the normal distribution $N(R_a, \sigma_r)$. The mean of this distribution are the author's actual resources, while the standard deviation is a proportion of this mean obtained following Equation 3, where the parameter $rv \in [0, 1]$ indicates the level of variability when perceiving others' resources. This variability mimics the typical knowledge and information asymmetry between authors and referees that characterize peer review.

$$\sigma_r = R_a * rv \qquad (3)$$

The quality of the submission as judged by the referee, i.e., $E' \in [0, 1]$, depends on referee behaviour. Unreliable referees can fall into two types of errors: I) recommending submissions of low quality to be published or II) recommending against the publication of submissions which should have been published. We assume that unreliable referees have a fifty percent probability of falling into type I or type II error. If the referee falls into a type I error, E' is calculated as in Equation 4, with $o \in]1, 2]$ as the model's overrating factor. The minimum value of o determines a certain degree of overestimation whereas the maximum would make the referee double estimate the author's resources.

$$E' = \frac{v*R'_a*o}{v*R'_a*o+1} \qquad (4)$$

When referees fall into a type II error, referees apply an underrating factor to the perceived author's resources as shown in Equation 5. The model parameter $u \in [0,1[$ indicates the percentage of underestimation made by the referee.

$$E' = \frac{v*R'_a*u}{v*R'_a*u+1} \qquad (5)$$

Fair referees evaluate authors' submission using Equation 6, so that evaluation scores approximate the actual paper quality.

$$E' = \frac{v*R'_a}{v*R'_a+1} \qquad (6)$$

Finally, referees behaving as cheaters always commit a type II error, as defined for unreliable referees (see Equation 5).

Publication
The publication rate (*p*) determines the percentage of acceptance so that only submissions getting the highest evaluations are eventually published.

Resource expenses and accumulation
When playing the role of an author, scientists invest all their resources for the submission. If the paper is not published, following the "winner takes all" rule characterising science, authors lose all invested resources. If published, authors accumulate resources according to their publication score, which leads to subsequent submissions of presumably higher quality. The guiding principle is that the more scientists publish, the more resources they have access to, and thus the higher their academic status and position is (e.g., Squazzoni and Gandelli 2012).
We assume that a successful publication increments author resources through a linearly variable multiplication factor in the range [1, *m*], where *m* is a parameter of the model. We use higher values for less established authors (i.e. those with less amount of accumulated resources) and we approximate 1 for more established authors. This mimicks reality as 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 (e.g., Squazzoni and Gandelli 2012).
When acting as referees, scientists allocate resources to reviewing as shown in Equation 7, where $S \in R$ is the amount of resources that are consumed. The cost of reviewing grows linearly with the quality of author submissions and is proportionally dependent on the referee's resources. If referees are matched with a submission of a quality close to a potential submission of their own, they allocate 50% of their available resources towards reviewing. Accordingly, they spend either less resources when matched with lower quality submissions or more resources when matched with higher quality submissions. Even though top scientists are generally expected to spend less time in reviewing, as they have more experience and are better suited to evaluate sound science than are average scientists, they lose more resources than average scientists because their time is more scarce and valuable (Squazzoni and Gandelli 2012).

$$S = \left(\frac{1}{2}R_r\left(1 + (Q - E_r)\right)\right)*s \qquad (7)$$

Unreliable referees spend less resources than fair and cheating referees as they do not put much effort in reviewing. As indicated by Equation 7, a multiplication factor $s \in [0, 1]$ is applied to reviewing costs. This parameter indicates the percentage of resources saved by unreliable referees, while it is set to 1 in the other cases.

Indexes calculation
In order to measure the quality and efficiency of the peer review process, we define four indexes that are calculated at the end of each simulation step, namely: the evaluation bias, the productivity loss, the reviewing expenses and the Gini index (e.g., Squazzoni and Gandelli 2012).
The evaluation bias (*EB*) indicates the quality of the peer review process by comparing the optimal situation, in which submissions would have been published according to their quality, to the actual situation in which publication depends on the referee opinion. As shown in Equation 8, we calcuate *EB* as the ratio between the publication errors (*PE*), i.e. the number of unpublished articles that should have been published and the total number of published articles (*PA*).

$$EB = \frac{PE}{PA}*100 \qquad (8)$$

The productivity loss (*PL*) measures the percentage of resources wasted by unpublished authors who deserved to be published. Equation 9 calculates this metric as the percentage of the difference of the quality from the submissions that should have been published (*BPQ*, from Best Publications Quality) minus the quality from the submissions that were actually published (*PQ*, from

Published Quality) with respect to the quality from submissions that should have been published (BPQ).

$$PL = \frac{BPQ - PQ}{BPQ} * 100 \qquad (9)$$

The reviewing expenses (*REx*) is the ratio between the total resources spent by referees and the total resources invested by authors. As indicated by Equation 10, they are measured as the sum of resources used by referees in the evaluation process (*RevSp*) divided by the sum of resources invested by authors in their submissions (*AuthSp*).

$$REx = \frac{RevSp}{AuthSp} * 100 \qquad (10)$$

Finally, the Gini index measures the inequality in the allocation of resources at the system level; it takes 0 when there is complete equality in the resource distribution and 1 when a single agent has everything. We calculate this index by considering the difference of resources for each pair of agents as in Equation 11, where *n* is the number of scientists and $\bar{R}$ is the mean of the amount of resources of everyone.

$$Gini = \frac{\sum_{i,j=0}^{n} |R_i - R_j|}{2 * \bar{R} * n^2} \qquad (11)$$

Submodels

Depending on the type of author/referee matching policy and reviewing behaviour, we developed sixteen submodels that include certain typical situations of peer review.

Matching policy scenarios
We distinguish four types of author/referee matching policy (*M*) scenarios as follows: random, peer, higher-skilled and lower-skilled.
In the "random" matching policy (RMP), authors and referees are randomly matched as if editors would lack knowledge of the scientists' expertise. This mimics the "luck of the reviewer draw" situation where good quality authors can be matched to lower quality referees and vice-versa. In the other scenarios, we assume that editors have full information on the potential quality of their pool of referees that can be used for referee selection.
In the "peer" matching policy (PMP), authors are matched to referees with similar skills. We arranged authors and referees in two lists that are sorted in descending order by the amount of resources. By following this ordering, authors are paired with their corresponding referees.
In the "higher-skilled" matching policy (HSMP), authors are matched to referees of higher expertise. In this case, for each author, there are two lists of referees: one including referees with equal or greater amount of resources, the other one including referees with lower resources than the author. Then, the author is assigned a random referee from the first list unless there is no one

left (i.e., all of them have been previously matched to other authors). In the latter case, a random referee from the second list is selected.
Finally, in the "lower-skilled" matching policy (LSMP), authors are matched with referees of lower prestige. The logic is the same as in the "higher-skilled" matching policy.
It is worth noting that the last two scenarios mimic situations in which editors could exploit the willingness of high quality scientists to contribute to the reviewing process (i.e. the "higher-skilled" matching policy) or where young scholars (typically PhD students and post-doc researchers) are more frequently involved. It is also worth noting that, in the higher- and lower-skilled scenarios, the success in the application of the matching policy is influenced by the concrete availability of the required referees. For instance, matching an author to a referee with higher resources would not be possible in a situation in which good referees had been already assigned. To check for this, we measured how many times these situations occurred in our simulations. On average, these situations occurred for about 20% of the matchings. On the one hand, this would mimick a realistic constraint of peer review, as editors cannot always find an optimal matching. On the other hand, this constraint has a limited impact on our results.

Reviewing behaviour scenarios
We designed four scenarios for evaluating the effect of unreliable, fair and cheating behaviour of referees.
In the "random behaviour" scenario, there is no room for cheating strategies but only for random behaviour. This is a baseline for studying the effect of the cheating behaviour. During the initialisation of this scenario, we set the scientists behaving unreliably by means of the model parameter $up \in [0, 1]$, which indicates the probability of being unreliable. Then, referee behaviour does not change during the simulation, as if there were no learning or influence from the context.
In the "cheating" scenario, we assume a fixed number of unreliable referees (determined by the *up* parameter) while the rest of referees behave according to the following criterion: if referees perceive that authors they are matched to have similar or higher resources, they see them as competitors and behave as cheaters; otherwise, they behave fairly. To do so, referees estimate authors' resources (R'_a) using a normal distribution of the form $N(R_a, \sigma_r)$, where R_a is the actual author's resources and σ_r follows equation 3.
In the "local competition" scenario, we assume a fixed number of unreliable referees (in accordance with the *up* parameter), while the rest of referees detect possible competitors only in their own resources neighbourhood. This scenario mimics a situation where referees underrate submissions by authors that have similar resources, while not caring about others. In these cases, their evalutation is fair. To this purpose, we use a Gaussian neighbourhood function of the form $N(R_r, \sigma_c)$ from which we obtain a sample value Υ. The mean of this normal distribution is the referee's amount of

resources (R_r), while the standard deviation follows Equation 12 using the parameter $cd \in [0, 1]$, which indicates the distance needed to consider an author as a competitor. Then, referees adopt cheating behaviour when the perceived author's resources ($R'_a \in N(R_a, \sigma_r)$) are within the interval $[R_r-(|R_r-\Upsilon|),\ R_r+(|R_r-\Upsilon|)]$. Otherwise, referees behave fairly.

$$\sigma_c = R_r * cd \qquad (12)$$

In the "glass ceiling" scenario, we assume that there is a fixed number of unreliable referees (again set through the *up* paremeter) while the rest of referees try to outperform both the less and the more productive colleagues. Therefore, given a referee's amount of resources (R_r), the probability of cheating increases when the perceived author's resources ($R'_a \in N(R_a, \sigma_r)$) approach R_r and it is higher when they are greater than R_r. To model the probability of cheating ($P(B=cheating)$), we used the logistic function shown in Equation 13, where d indicates the absolute distance between the author's and the referee's resources (i.e. $d = |R'_a-R_r|$).

$$P(B = cheating) = \frac{1}{(e^{-(\beta_1 * d + \beta_0)}+1)} \qquad (13)$$

Constants β_0 and β_1 determine the shape of the curve as shown in Figure 1 and are calculated as in Equations 14 and 15, respectively. These equations are related to the three model parameters as follows: k_1 indicates the probability of cheating when both referee and author have the same amount of resources (i.e. $d = 0$), and k_2 indicates the probability of cheating when the distance between the author's and the referee's resources is equal to k_3.
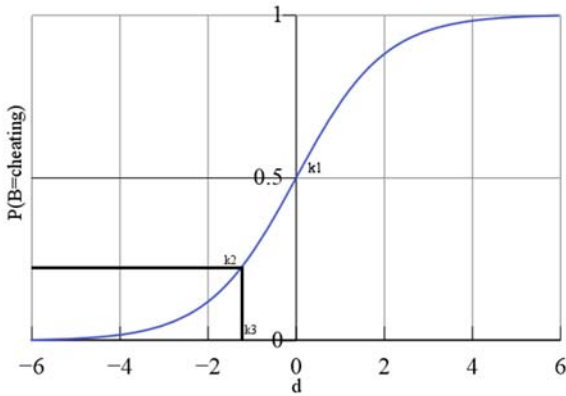

Figure 1: The logistic function used for detecting competitors in the "glass ceiling" scenario.

$$\beta_0 = -ln\left(\frac{1-k_1}{k_1}\right) \qquad (14)$$

$$\beta_1 = \frac{ln\left(\frac{1-k_2}{k_2}\right)+\beta_0}{k_3} \qquad (15)$$

Global parameters of the model
Table 3 shows the model parameters also including the number of scientists (n) and the initial amount of resources for each of them (r_0). It also specifies the range and values that we used to obtain the results shown in the next section.

Table 3: Global parameters of the model

Par.	Description	Range	Values
n	Number of scientists	$\mathbb{N}$	200
r_0	Initial resources	$\mathbb{R}$	1
f	Fixed amount of resources gained per agent in each simulation step	$\mathbb{R}$	1
p	Publication rate	[0, 1]	{0.25, 0.5, 0.75}
m	Publication multiplier	$\mathbb{R}$	1.5
up	Probability of being unreliable	[0, 1]	0.5
s	Reviewing expenses factor for unreliable referees	[0, 1]	0.5
o	Overrating factor in type I errors	]1, 2]	1.9
u	Underrating factor in type II errors	[0, 1[	0.1
v	Submission quality velocity	[0, 1]	0.1
av	Author variability	[0, 1]	0.1
rv	Reviewing variability	[0, 1]	0.1
cd	Competitor distance in the local competition scenario	[0, 1]	0.5
k_1	Cheating probability when resources are equal in the glass ceiling scenario	[0, 1]	{0.75, 0.85, 1}
k_2	Cheating probability when resources differ a value of k_3 in the glass ceiling scenario	[0, 1]	{0.4, 0.5, 0.6}
k_3	Resource distance to have a k_2 cheating probability in the glass ceiling scenario	$\mathbb{R}$	{1, 2, 5, 10}

RESULTS

This section shows the results for all possible combinations of the matching policy and referee behaviour scenarios presented above. For each setting, we averaged the results for 10 runs of 200 simulation steps.

Table 4 shows the results for the most competitive publication rate ($p = 0.25$) which means 25% of submissions eventually published in each simulation step. Other less competitive policies were tested (e.g., p

= 0.50 and p = 0.75) that yielded similar results which, for the sake of shortness, we did not report here.

In the case of "random" and "local competition" referee behaviour, any editorial matching policy determined more evaluation bias, higher productivity lost and similar or higher reviewing expenses than the random matching. The situation was different in the "cheating" and "glass ceiling" scenarios. In these cases, the "peer" and "higher-skilled" matching policies significantly lowered evaluation bias and the productivity loss compared to the "random" matching policy. In terms of resource distribution, "random behaviour" and "local competition" scenarios generally generated higher values for the Gini index, unless the "peer" matching policy was applied.

Table 4: The effects of editorial matching policies on the peer review process.

	Eval. bias	Prod. loss	Rev. exp.	Gini. index	Cheat. perc.
Random behaviour					
RMP	29.42	15.00	29.42	0.47	NA
PMP	39.55	19.56	34.43	0.37	NA
HSMP	32.99	16.22	30.87	0.43	NA
LSMP	29.51	15.71	29.47	0.46	NA
Cheating					
RMP	70.86	34.72	35.24	0.28	0.27
PMP	51.97	25.69	35.19	0.33	0.25
HSMP	61.95	29.81	34.60	0.30	0.19
LSMP	73.00	36.92	34.86	0.29	0.32
Local competition					
RMP	31.04	15.63	30.13	0.45	0.20
PMP	57.87	28.61	35.70	0.31	0.41
HSMP	36.54	17.74	31.85	0.41	0.22
LSMP	33.47	17.37	30.06	0.44	0.18
Glass ceiling					
RMP	70.35	34.70	34.56	0.29	0.34
PMP	58.02	28.56	35.64	0.32	0.38
HSMP	65.88	32.26	35.23	0.30	0.37
LSMP	68.21	34.47	34.29	0.29	0.36

It is worth noting that the difference of the competitors' detection mechanism had a considerable effect on the average percentage of cheaters in the population. Generally, this was higher in the "*glass ceiling*" scenario, although the highest value was reached when the "peer" matching policy was applied in the "*local competition*" scenario, i.e., when authors were matched with referees of similar quality. Furthermore, it is worth noting that the evaluation bias is not univocally correlated with the number of cheaters in the population. This reflects the importance of the editorial matching policy to influence the peer review process.

Figure 2 shows the indexes equilibrium for the "*local competition*" scenario. In this situation, applying a "peer" matching policy is detrimental as it leads to higher evaluation bias, productivity loss and reviewing expenses. On the other hand, other editorial matching policies lead to better quality, productivity and efficiency of peer review. Though, this came at a price of having a more unequal distribution of resources, where it is assumed that the best published scientists get the most (e.g., Squazzoni and Gandelli 2013).
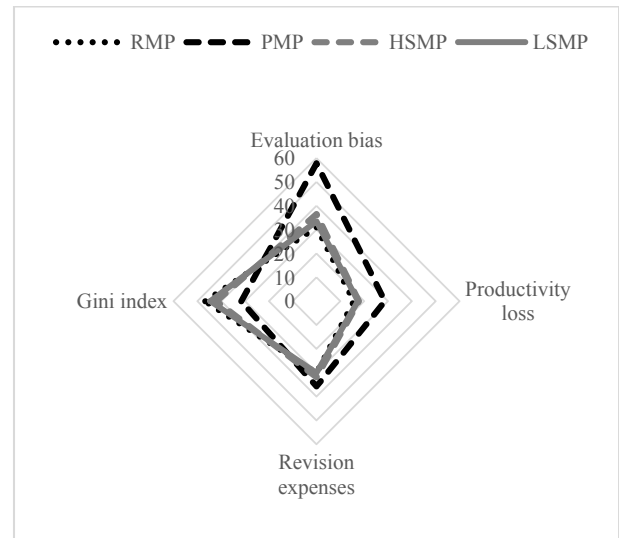

Figure 2: Indexes equilibrium for the different editorial matching policies in a "*local competition*" behavioural scenario. Note that the Gini index in this figure has been drawn as a percentage.

CONCLUSIONS

This paper aimed to contribute to a growing stream of literature that looks at peer review through agent-based models (e.g., Thurner and Hanel 2011; Allesina 2012; Paolucci and Grimaldo 2014). Given the lack of empirical data on internal processes of peer review, the use of agent-based models can allow us to understand implications of scientist behaviour in idealised situations (Squazzoni and Takacs 2011).

Our results show possible editorial counteractions to reduce the impact of referee misbehaviour, such as matching authors and referees by looking at their reputation. We found that, in case of complete randomness of referee judgment, any editorial matching policy may have even a negative effect. If referees behave strategically, certain matching policies, such as assigning referees of similar or higher quality than submission authors, might counteract referee bias. Besides certain implementation constraints, e.g., the scarce availability of reliable referees, there is also a side-effect of exploiting reliable referees, i.e., generating benefits to published authors who might gain cumulative publication advantages.

Furthermore, 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 matching policies tends to decrease.

REFERENCES

Allesina, S. 2012. Modeling Peer Review: An Agent-Based Approach". *Ideas in Ecology and Evolution*. 5(2), pp. 27-35. Doi:10.4033/iee.2012.5b.8.f.

Cabotà, J.B.; Grimaldo, F.; and Squazzoni, F. 2013. "When Competition is Pushed too Hard. An Agent-Based Model of Strategic Behaviour of Referees in Peer Review". *Proceedings 27th European Conference on Modelling and Simulation.*

Callaham, M. L. and Tercier, J. 2007. "The Relationship of Previous Training and Experience of Journal Peer Reviewers to Subsequent Review Quality". PLoS Med 4(1): e40. Doi:10.1371/journal.pmed.0040040.

Grimaldo, F; Paolucci, M.; Conte, R. 2011. "Agent simulation of peer review: the PR-1 model". *Proceedings 12th International Workshop on Multi-Agent Based Simulation*, LNCS 7124, Pages 1-14.

Grimaldo, F; Paolucci, M. 2013. "A simulation of disagreement for control of rational cheating in peer review". *Advances in Complex Systems*. Doi: 10.1142/S0219525913500045.

Grimaldo, F. and Paolucci, M. 2014. "Mechanism change in a simulation of peer review: from junk support to elitism". *Scientometrics*. Doi: 10.1007/s11192-014-1239-1.

Grimm, V; Berger, U; DeAngelis, D. L.; Polhill, J. G.; Giske, J.; and Railsback, S. F. 2010. "The ODD Protocol: A review and first update". *Ecological Modelling*. Volume 221, Issue 23, 24 November 2010, Pages 2760-2768.

Hamermesh, D.S. 1994. "Facts and Myths about Refereeing". *Journal of Economic Perspectives*. Volume 8, Number 1, 153-163.

Jefferson, T.; Wager, E.; and Davidoff, F. 2002. "Measuring the Quality of Editorial Peer Review". *The Journal of the American Medical Association*. Vol. 287, No. 21, 2786-2790.

Kravitz, R.L.; Franks, P.; Feldman, M.D.; Gerrity, M.; Byrne, C.; and Tierney, W.M. 2010. "Editorial Peer Reviewers' Recommendations at a General Medical Journal: Are They Reliable and Do Editors Care?". *PLoS ONE* 5(4): e10072. Doi:10.1371/journal.pone.0010072.

Mulligan, A.; Hall, L.; and Raphael, E. 2013. "Peer Review in a Changing World: An International Study Measuring the Attitudes of Researchers". *Journal of the American Society for Information Science and Technology* 64(1):132-161.

Neff, B.D. and Olden, J.D. "Is Peer Review a Game of Chance?" *BioScience,* 56(4):333-340, April 2006.

Newton, D.P. 2010. "Quality and Peer Review of Research: An Adjudicating Role for Editors". *Accountability in Research: Policies and Quality Assurance.* 17:3, 130-145.

Polhill, J. G; Parker, D; Brown, D.; and Grimm, V. 2008. "Using the ODD Protocol for Describing Three Agent-Based Social Simulation Models of Land-Use Change". *Journal of Artificial Societies and Social Simulation.* Vol. 11, no. 2 3.

Squazzoni, F. 2010. "Peering into Peer Review". *Sociologica,* 3. DOI:10.2383/33640.

Squazzoni, F., Bravo, G., Takács, K. 2013. "Does Incentive Provision Increase the Quality of Peer Review? An Experimental Study". *Research Policy*, 42(1), pp. 287-294.

Squazzoni, F. and Gandelli, C. 2012. "Saint Matthews Strikes Again. An Agent-Based Model of Peer Review and the Scientific Community Structures". *Journal of Informetrics*, 6: 265-275.

Squazzoni, F. and Gandelli, C. 2013. "Opening the Black Box of Peer Review. An Agent-Based Model of Scientist Behaviour. *Journal of Artificial Societies and Social Simulation*, 16(2) 3: http://jasss.soc.surrey.ac.uk/16/2/3.html

Squazzoni, F. and Takacs K. 2011. "Social Simulation That 'Peers into Peer Review'". *Journal of Artificial Societies and Social Simulation*, 14(4) 3: http://jasss.soc.surrey.ac.uk/14/4/3.html.

Thurner, S. and Hanel, R. 2011. "Peer review in a World with Rational Scientists: Toward Selection of the Average", *The European Physical Journal B*, 84, pp. 707-711. Doi: 10.1140/epjb/e2011-20545-7.

Weber, E.J.; Katz, P. P.; Waeckerle, J.F.; and Callaham, M.L. 2002. "Author Perception of Peer Review". *The Journal of the American Medical Association.* Vol. 287, No. 21, 2790-2793.

AUTHOR BIOGRAPHIES

JUAN BAUTISTA CABOTÀ is PhD. Student at the University of Valencia. His research interests include agent-based modelling and simulation and intelligent decision-making support systems. His email address is: juan.cabota@uv.es.

FRANCISCO GRIMALDO is associate professor at the University of València. His research is focused on agent-based modelling and simulation. He is member of the HiPEAC network of excellence and the IEEE Systems, Man & Cybernetics Society. His email is francisco.grimaldo@uv.es and his webpage can be found at: http://www.uv.es/grimo/.

FLAMINIO SQUAZZONI leads the GECS-Research Group on Experimental and Computational Sociology. He is President of *ESSA* The European Social Simulation Association. His e-mail is: flaminio.squazzoni@unibs.it and his web-page can be found at: www.eco.unibs.it/gecs/Squazzoni.html.