

PEER REVIEW UNDER THE MICROSCOPE. AN AGENT-BASED MODEL OF SCIENTIFIC COLLABORATION

Flaminio Squazzoni

University of Brescia
Via San Faustino 74/B
25122 Brescia, ITALY

Claudio Gandelli

University of Brescia
Via San Faustino 74/B
25122 Brescia, ITALY

ABSTRACT

This paper investigates whether the quality and efficiency of peer review is more influenced by scientists' behaviour or by the type of scientific community structure (homogeneous vs. heterogeneous). We looked especially at the importance of reciprocity and fairness to ensure cooperation between everyone involved and the role of evaluation standards to reduce parochialism. We modelled peer review as a process based on knowledge asymmetries and subject to evaluation bias. We found that reciprocity can have a positive effect on peer review only when agents are not driven by self-interest motivation and are inspired by standards of fairness. Secondly, we found that in a strong competitive scientific landscape, high quality of peer review can be achieved when shared evaluation standards are supported by normative standards of conduct. Finally, we found that unequal resource allocation in science (e.g., reputation and funds) is the consequence of good peer review standards.

1 INTRODUCTION

Peer review is under the spotlight again. Recent cases of misconduct and fraud, such as the stem cell scandal in Science 2005 or the more recent Stapel scandal, called for a reconsideration of the rigour and reliability of evaluation (Crocker and Crooper 2011; Gewin 2012). Most observers emphasised that this essential institution has been dramatically under-investigated (e.g., Couzin 2006; Smith 2006; Alberts, Hanson and Kelner 2008) and that we lack sound experimental knowledge on peer review mechanisms (Bornmann 2011).

Recently, two attempts of explanation have been suggested that account for certain peer review problems. Some authors have emphasised the weakness of (material and reputational) incentives for referees (e.g., Chang and Lai 2001; Squazzoni, Bravo and Takács 2012). Indeed, while journal editors and submission authors can benefit from reputational rewards, referees' incentives and motives are less clear. On the other hand, other authors have emphasised the importance of evaluation standards and the radical uncertainty intrinsic to any academic judgement, especially for multi-disciplinary research and in non-paradigmatic sciences (Lamont 2009).

The first explanation looked at behaviour, norms and incentives of everyone involved, without considering disciplinary specifications and the variety of epistemic communities. These studies viewed peer review as a typical cooperation problem, where agents (e.g., authors, referees and editors) can behave strategically and various equilibria might result from their interaction (e.g., Northcraft and Tenbrunsel 2011). They especially focused on the importance of reciprocity motives. Referees would cooperate with journal

editors in ensuring the quality of evaluation as they are concerned about protecting the prestige of the journal in case of (i) previous publication, as a means to protect their own impact and/or (ii) future publication, as a means to build positive reputation with editors. On the other hand, they could also be motivated to cooperate with authors by providing a fair evaluation and constructive feedback as they are interested to establish good standards of reviewing in prospect of benefiting from other referees when they will be authors (Seinen and Schram 2004; Squazzoni, Bravo and Takács 2012).

In this view, potential traps would easily occur as reciprocity motives are fragile and extremely sensitive to interaction detail, especially in a context of increasing competition of scientists and dramatic fragmentation of the scholarly publication market. These conditions would provide room for cheating temptations as monitoring and sanction costs against unfair behaviour would be unsustainable. This is even more complicated by the fact that referees are dramatically overexploited and this could increase their cheating temptations, e.g., reducing time for reviewing and ensuring less quality. It is worth noting that recent surveys estimated that peer evaluation is applied to about 1.5 million journal articles per year, not to mention conferences, research proposals, fellowships and university/dept/institute productivity evaluation (Björk, Roos and Lauri 2009).

The second explanation looked at peer review less generally and from the point of view of scientific community structure and historical development of disciplines. These studies emphasised the diversity of disciplinary evaluative cultures and suggested that the problem is the intrinsic complexity of any evaluation process of scientific products, due to highly specific knowledge and increasingly complex research technologies. These features of peer review would even question the possibility of shared evaluation standards across disciplines, fields and epistemic cultures (Abbott 2001; Lamont 2009).

To make the situation even more complicated, another constraint should be considered, i.e., the difficulty of studying peer review empirically, at least at the appropriate level needed to look at interaction and processes in detail. Our paper is an attempt to contribute to understand peer review by proposing a modelling approach, which starts from the idea that manipulating agent behaviour and interaction in a computer simulation model can allow us to observe resultant peer review aggregate dynamics that are difficult to observe empirically (e.g., Squazzoni and Takács 2011).

Recently, some studies have shown the potential of this approach. Martins (2010) examined the role of social influence in determining the convergence of scientists' opinions towards common paradigms. His results showed that inter-generational replacement was pivotal to ensure less parochialism and a widely acceptance of innovative research. Thurner and Hanel (2011) studied the effect of peer review bias on the quality of publication by modelling authors and referee interaction. Their results showed that a small fraction of unfair referees is sufficient to considerably reduce the quality of published work. Roebber and Schultz (2011) modelled the research funding process and the interaction of program officers and referee behaviour to look at proposal-submission strategies. They found that program officers who use more referees and require consensus allow funding agencies to reduce the proposal submission, but this is fragile against referee selfish or negligent behaviour. In view of this and in contexts of strong competition for funds, it is expected that proposers continue to submit many proposals, overloading program officers and referees.

Our aim was to understand the impact of strategic behaviour of scientists on the quality and efficiency of peer review and to test the influence of the community structure. This was to integrate certain behavioural and structural aspects emphasised by the two explanations mentioned above. We wanted to understand under which conditions reciprocity strategies can sustain cooperation, in terms of quality and efficiency of peer review, and whether shared evaluation standards could compensate for unfair behaviour of scientists.

The structure of the paper is as follows. Sect 2 will introduce the model, while in Sect 3 we will present various simulation scenarios and the simulation parameters. In Sect 4, we will illustrate our simulation results, while in the concluding section, the main findings will be summarized.

2 THE MODEL

We assumed a population of N scientists ($N = 200$) randomly selected each tick to play one of two roles: authors or referees. The task of an author was to submit an article and have it published. The task of a referee was to evaluate the quality of submissions by authors. Depending on the referees' opinion, only the best submissions were published (i.e., those exceeding a given quality threshold).

We gave each agent a parameter for individual productivity, which was initially homogeneous. Productivity was a measure of academic status, position, experience and scientific achievement of scientists. The principle was that the more scientists published, the more resource they had and the higher their academic status and position were. We assumed that resources were needed both to submit and review an article. With each simulation tick, agents were endowed with a fixed amount of resources, equal for all (e.g., common access to research infrastructure and internal funds, availability of PhD. students). Then, they cumulated resources according to their publication score.

We assumed that the quality of submissions was dependent on agent productivity. Each agent had resources $R_a \in N$ from which we derived an expected submission quality as follows:

$$\mu_{ar} = \frac{v \cdot R_a}{v \cdot R_a + 1}$$

where v was the velocity through which agent submissions can approximate the best quality ($v = 0.1$). We assumed that authors varied in terms of quality output. The quality of submissions by authors followed a standard deviation σ which proportionally varied according to agent productivity and followed a normal distribution $N(R_a, \sigma)$. This meant that, with some probability, top scientists could write average or low quality submissions, and average scientists had some chance to write good submissions.

The chance of being published was determined by evaluation scores assigned by referees (see below). If published, agents earned resources proportionally to the article's quality assigned by referees as follows:

$$P = \frac{Q}{v * (1 - Q)}$$

where P was the productivity gained and Q the submission quality score assigned by referees and v the velocity of the best quality approximation mentioned above.

We assumed that publication multiplied author resources of a M value, which gradually varied between 1 for more productive published authors to 1.5 for less productive published authors. This was to mimic reality where productivity gain from publication is crucial to explain differences in scientists' performance, but is higher for scientists at their initial steps and cannot exponentially increase for top scientists. More specifically, at the end of each tick, we graded all published authors i per resources according to the A_n sequence, where A_0 was the published author with less resources and $A_{(i-1)}$ the published authors with more resources. We defined g as the highest multiplier value possible ($g = 1.5$). The multiplier M for the n -agent was calculated as follows:

$$M_{(A_n)} = g - \left(\frac{n}{i-1}\right) * (g-1) = \frac{g(i-n-1) + n}{i-1}$$

If not published, following the "winner takes all" rule characterizing science, we assumed that authors lost all the resources invested for submitting. Therefore, the value of author submissions was not objectively determined (i.e., it did not perfectly mirror the real quality of submissions), but was dependent on the referees' opinion. When selected as referees, agents needed to invest a given amount of resources (see below) for reviewing but simultaneously lost them as they could not publish in the meantime. We as-

sumed 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 on the population.

We assumed that reviewing was a resources-demanding activity and that agent productivity determined both the agent's reviewing quality and its cost (i.e., time lost for publishing). The total expense S for any referee was calculated as follows:

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

where R_r was referee's resources, Q_a was the real quality of the author's submission and μ_r was the referee's expected quality. Furthermore, we assumed that reviewing expenses grew linearly with the quality of authors' submissions. We assumed that if referees were matched with a submission of a quality close to a potential submission of their own, they spent 50% of their resources for reviewing. They spent less when matched with lower quality submissions, more when matched with higher quality submissions. However, reviewing expenses depended proportionally on agent productivity. This means that top-scientists will waste less time for reviewing in general, as they have more experience and ability to evaluate good science than average scientists. However, they will lose more resources than average scientists because their time is more costly than the latter.

We assumed two types of referee behaviour, namely *reliable* and *unreliable*. For reliability, we meant the capacity of referees to provide a consistent and unequivocal opinion that truly reflects the quality of the submission. In case of reliability, referees did the best they could to provide an accurate evaluation and spent resources for reviewing close to their expected quality level. In this case, we assumed a normal distribution of the referees' expected quality, which depended on their productivity, and a narrow standard deviation of their evaluation score from the real value of the submission ($\sigma = R_a/100$). This meant that the evaluation scores by reliable referees were likely to approximate the real value of author submissions. However, we assumed that, also in case of reliability, the evaluation bias increased proportionally to the difference between referees' expected quality and author submission quality. This was to represent the knowledge and information asymmetries between authors and referees that characterize peer review in science.

In case of unreliability, referees fell into type I and type II errors: recommending to publish submissions of low quality or recommending not to publish submissions that should be published (Laband and Piette 1994). More specifically, unreliable referees spent less resources than reliable referees, and under or over estimated author submissions. To avoid that referees eventually assigned the real value by chance to submissions, we assumed that, when they underrated a submission, the evaluation score took a standard deviation around -90% of the real quality of the submission. The opposite sign was assigned in case of overrating (i.e., $+90\%$).

Finally, all simulation parameters are shown in Table 1. Agent resources were set at the beginning of the simulation at 0 for all. At the first tick, 50% of agents were published randomly. Subsequently, everyone had a fixed productivity gain each tick. If published, agents had the value of their publication multiplied by the parameter M [1, 1.5] and so their resources grew accordingly. This meant that the quality of their subsequent submission was presumably higher.

Table 1: The Simulation Parameters

Parameters	Value
Initial agent resources	0
Fixed productivity gain	1
Publication selection rate	[0.25, 0.50, 0.75]
Publication productivity multiplier	[1, 1.5]
Unreliability probability	[0, 1]
Evaluation bias by default	0.1
Author investment for publication	1
Reviewing expenses of unreliable referees	0.5
Underrating by unreliable referees	0.1
Overrating by unreliable referees	1.9
Velocity of best quality approximation	0.1

3 SIMULATION SCENARIOS

We built various simulation scenarios to test the impact of scientists' behaviour on the quality and efficiency of peer review. For quality, we meant the capability of peer review to ensure that only the best submissions were eventually published. For efficiency, we meant the capability of peer review of achieving quality by minimizing productivity lost by authors and reviewing expenses by referees.

Each scenario was tested in two types of scientific community, i.e., homogeneous vs. heterogeneous structures. For homogeneous community, we meant a community where shared evaluation standards permit a competent evaluation of author submissions, irrespectively of the author's position in the community. For heterogeneous community, we meant the co-presence of various sub-communities of scientists who have heterogeneous features, such as various scientific paradigms and there is no common evaluation standard. In this situation, we assumed that agents could fairly evaluate only work by similar other, while they tended to proportionally underrate agents more distant from them. As mentioned above, to calculate nearness and distance between agents, we considered the agent position in the productivity trajectory, which was a proxy of the agent position and status in the community.

In the first scenario (called "*no reciprocity*"), we assumed that, when selected as referees, agents had a random probability of behaving unreliably, which was constant over time and not influenced by past experience. When selected as authors, agents invested all their resources for publication, irrespectively of positive or negative past experience with their submission. In the second scenario (called "*indirect reciprocity*"), we assumed that agents, when selected as referees, in case of previous publication, reciprocated by providing reliable evaluations in turn when selected as referees. In the third scenario (called "*fairness*"), when selected as authors, agents measured the fairness of the referee opinion by comparing the real quality of their submission and the evaluation rate received by the referee. If the referee evaluation approximated the real value of their submission (i.e., $\geq -10\%$), they concluded that the referee was reliable and did a good job. In this case, when selected as referees, they reciprocated positively with other authors irrespectively of their past publication or rejection. This meant that now indirect reciprocity was not based on self-interest but on normative standards of conducts.

Finally, the last two scenarios (called "*self-interested authors*" and "*fair authors*") extended the previous two by manipulating also author behaviour. In "*self-interested authors*", we assumed that, when published, authors reacted positively and continued to invest all their resources for their next submission. In case of rejection, they reacted negatively and invested less in the subsequent round (i.e., only the 10% of their resources). This reaction was independent from the pertinence of the referee evaluation. In "*fair*

authors”, in case of a pertinent referee evaluation received when authors, they reinforced their confidence on the good quality of peer review and continued to invest to send good quality submissions irrespectively of the fate of their submission. In case of non-pertinent evaluation (see above), they invested less in the subsequent round (i.e., only the 10% of their resources) and accumulate resources for the subsequent round irrespectively of their previous publication. Therefore, in this case, authors inferred the bad functioning of peer review and strategically disinvested in their publication quality.

To measure the consequences of parameter manipulation, we built the following indexes: (1) evaluation bias, (2) productivity loss, (3) reviewing expenses and (4) career inequality. The first dealt with quality of peer review, the second and third with efficiency and the last one with the unequal distribution of resources in the system. The first measured the percentage of evaluation errors made by referees each tick. We calculated the optimal situation, where submissions were published according to their real value and measured the discrepancy with the actual situation each tick. We did the same with the second index to measure the resources wasted by (unpublished) authors compared with the optimal solution. The third index measured the resources spent by agents for reviewing compared with authors. The fourth measured inequality in agent resources through a Gini index. Inequality here meant an unequal allocation of science outputs, such as productivity, academic status, and career.

4 RESULTS

Tab. 2 shows the impact of referee behaviour on the quality and efficiency of peer review in high and low competitive environments and in homogeneous and heterogeneous community structures. Data were averaged on 200 simulation run in any parameter condition. Results showed, first, that reciprocity motives of referees *per se* had a negative effect on the quality and efficiency of peer review in a strongly selective science environment, while improved only minimally the situation in less selective environments. It is worth noting that these minimal improvements were achieved only by considerably draining referees’ resources. Although increasing selection generally implied more evaluation bias, “fairness” determined lower bias and lower productivity loss by authors, but only when the scientific structure was homogeneous. In this case, the positive effect of reciprocity and fairness on peer review was possible only at growing expenses of referees. With the heterogeneous community, a stronger selection doubled or tripled the percentage of bias.

Tab. 3 shows the impact of reciprocal behaviour of authors in various selection rate environments. Results showed that reciprocity of authors improved peer review only when associated with fair criteria of judgment of the fate of their submission. The difference was significant especially in case of homogeneous structures. Furthermore, it is worth noting that in the case of heterogeneous structures, the reviewing expenses drastically grew even overcoming those of authors.

Table 2: The impact of referee behaviour on the quality and efficiency of peer review in various selective environments and different scientific community structures (values in percentage).

Scenario	Evaluation bias		Productivity loss		Reviewing expenses	
	Homogeneous community	Heterogeneous Community	Homogeneous community	Heterogeneous Community	Homogeneous community	Heterogeneous Community
<i>Weak selection (75% of submissions eventually published)</i>						

No reciprocity	14.10	20.64	5.69	8.90	23.47	29.88
Indirect reciprocity	12.58	22.21	6.51	9.03	44.16	86.30
Fairness	13.14	22.98	7.48	11.28	40.61	46.19
<i>Strong selection (25% of submissions eventually published)</i>						
No reciprocity	28.00	68.56	15.01	39.90	29.47	49.66
Indirect reciprocity	43.12	50.64	16.92	22.67	33.39	50.07
Fairness	19.52	65.36	8.32	35.68	38.29	63.26

Table 3: The impact of author reciprocal behaviour on the quality and efficiency of peer review in various selective environments and different scientific community structures (values in percentage).

Scenario	Evaluation bias		Productivity loss		Reviewing expenses	
	Homogeneous community	Heterogeneous Community	Homogeneous community	Heterogeneous Community	Homogeneous community	Heterogeneous Community
<i>Weak selection (75% of submissions eventually published)</i>						
Self-interested authors	15.30	25.30	12.63	17.78	46.07	107.63
Fair authors	14.85	25.76	5.10	16.64	29.55	219.19
<i>Strong selection (25% of submissions eventually published)</i>						
Self-interested authors	45.04	71.44	38.31	51.22	47.13	125.23
Fair authors	14.24	65.52	4.00	49.23	15.96	110.22

Then, we calculated the system productivity in each scenario, by averaging the resources of all agents at the end of the simulation. Considering “no reciprocity” as a benchmark, in the case of homogeneous community structures, “indirect reciprocity” implied a loss of 20% of system resources, “fairness” showed a loss of 7% while “self interested authors” doubled the productivity and “fair authors” determined an exponential growth of resources. The heterogeneous scientific community provided a different picture. In this case, whereas all other scenarios determined a significant loss of resources compared with the benchmark, “fair authors” implied a growth of resources of about 40% compared with the benchmark.

Looking at resource accumulation, it is worth noting that, in weakly selective environments, the impact of the community structure was significant in determining higher variation of “fair authors” scenarios (see Figures 1 and 2). Indeed, while in case of community homogeneity fairness caused an exponential growth of resources, the heterogeneous structure implied that resources stabilised around fixed values in all scenarios, although a significant volatility was shown in “fair authors”. However, the most important impact of the community structure in “fair authors” was evident when competition increased. In case of structure homogeneity, stronger competition determined an exponential growth of resources. This was not true in case of heterogeneous structures, where the level of resources only doubled in “self interested authors” (see Figures 3 and 4).

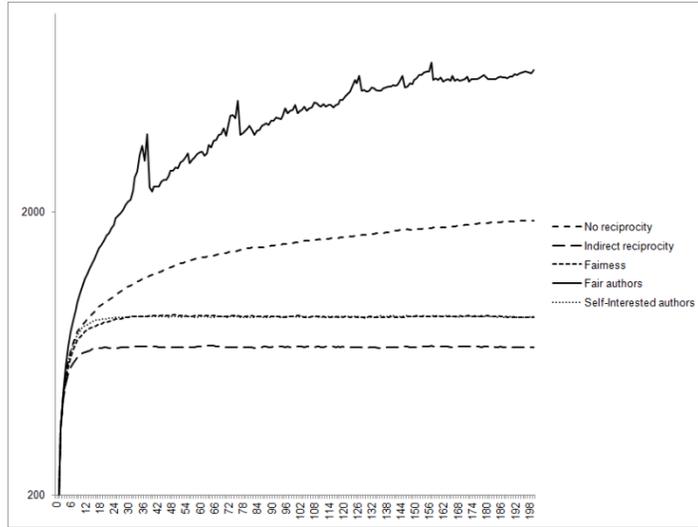


Figure 1: The impact of agent behaviour on system resource accumulation in weakly selective environments and homogeneous community structures. In the x-axis, the number of simulation run

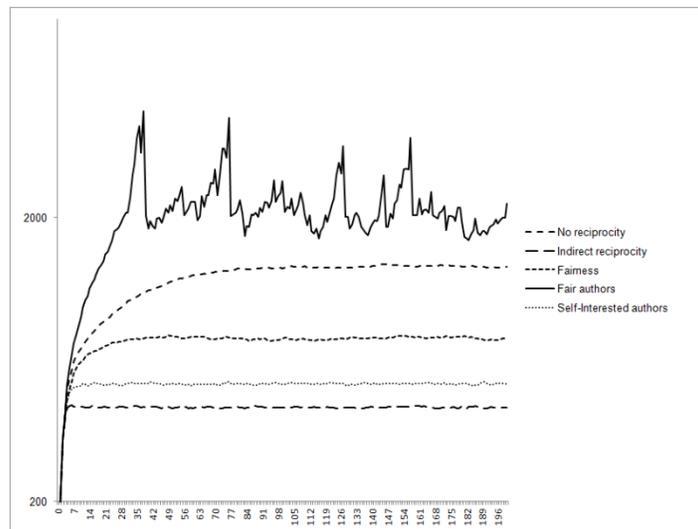


Figure 2: The impact of agent behaviour on system resource accumulation in weakly selective environments and heterogeneous community structures. In the x-axis, the number of simulation run.

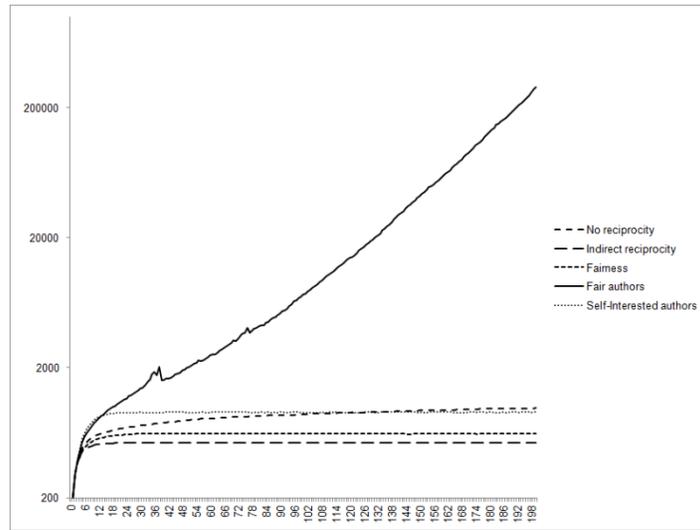


Figure 3: The impact of agent behaviour on system resource accumulation in strongly selective environments and homogeneous community structures. In the x-axis, the number of simulation run.

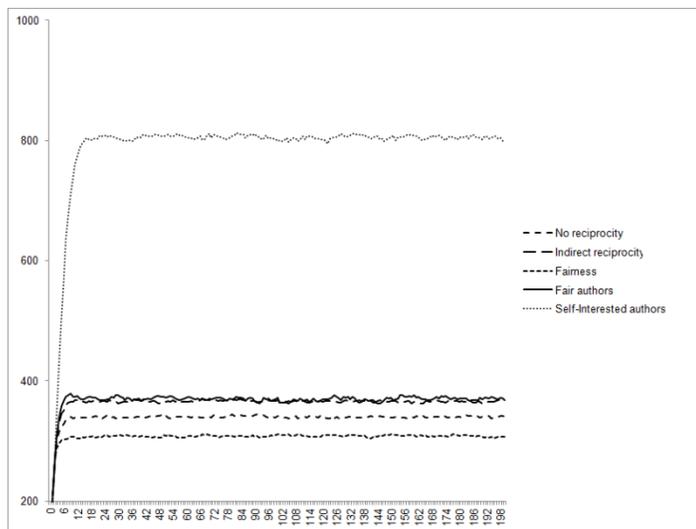


Figure 4: The impact of agent behaviour on system resource accumulation in strongly selective environments and heterogeneous community structures. In the x-axis, the number of simulation run.

Then, we calculated the inequality of resource distribution in each scenario through a Gini index (see Tab. 4). Results showed that once introduced reciprocity in homogeneous structures, the better scenarios in terms of quality and efficiency of peer review were also the most unequal in terms of resource distribution. When selection was stronger, this trend even exacerbated. This is coherent with our previous findings: in a competitive systems where the “winner takes all” principle is the rule, such as in science, where rewards are only for priorities and new discoveries, well functioning of peer review determines a unequal resource distribution as cumulative advantages for best scientists take place (Squazzoni and Gandelli 2012). This is because best published authors gain more resource and more chances to be re-published by taking advantage of fairness and reliability of referees.

The picture was different in case of heterogeneous community structures, especially when competition was strong. Indeed, in this case, the differences between scenarios disappeared. It is important to note that

while homogeneous structures implied that authors' fairness determined more inequality in resource allocation, as cumulative advantages took place for better scientists, here self-interest or fairness of authors made little difference and generated less inequality than in the previous case.

Table 4: The Gini index in all scenarios in different scientific community structures and in various selective environments (values calculated at the end of the simulation). The index gave 0 when there was complete equality in resource distribution among agents and 1 when a single agent had everything.

Scenario	Gini index	
	Homogeneous community	Heterogeneous community
<i>Weak selection (75% of submissions eventually published)</i>		
No reciprocity	0.55	0.57
Indirect reciprocity	0.34	0.35
Fairness	0.36	0.45
Self-interested authors	0.34	0.41
Fair authors	0.74	0.75
<i>Strong selection (25% of submissions eventually published)</i>		
No reciprocity	0.47	0.30
Indirect reciprocity	0.34	0.30
Fairness	0.45	0.26
Self-interested authors	0.35	0.34
Fair authors	0.88	0.39

5 DISCUSSION AND CONCLUSION

Our simulation results allow us to better understand the importance of reciprocity *per se* to explain the quality and efficiency of peer review, which was suggested by Chang and Lai (2001) and Squazzoni, Bravo and Takács (2012). Indeed, our results indicate that if reciprocity is influenced by self-interest of scientists, its potential positive effect for cooperation in peer review may evaporate. This suggests that the quality of peer review probably depends more on ethical, cultural and competence-based standards of conduct that might reduce the self-interest strategic behaviour of scientists, once provided room for reciprocity motives (e.g., Azar 2008).

The importance of norms and evaluation standards is evident if we compare peer review dynamics in homogeneous and heterogeneous community structures. Our results show that fair behaviour cannot compensate for the lack of shared evaluation standards. By randomly matching authors and referees, it was likely that scientists of a group were evaluated by referees of another group and this increased evaluation bias and inefficiency. This also explained the lower inequality of heterogeneous community structures, as outliers from our groups were likely to be penalised by hostile referees and this kept the system relatively homogeneous and equal in terms of resource allocation. This finding would confirm previous studies where certain differences between 'hard' and 'soft' sciences were outlined that revolve around the respective strength or weakness of shared evaluation standards (Lamont 2009). Our results indicate that without shared evaluation standards, as often happens in social sciences, the positive effect of fairness and pertinent judgment of scientists is less decisive for a better peer review.

Obviously, the picture is different when shared evaluation standards exist that help scientists to pertinently judge colleagues' work. Our results show that this situation ensures less bias and more efficient evaluation. However, in case of strong competition, shared evaluation standards were positive if combined with fairness and pertinent judgment by scientists. Furthermore, it is worth noting that the effect of the random matching between referees and authors did not rule out productivity differences between sci-

entists and cumulative advantages took place for published scientists (see also Squazzoni and Gandelli 2012). This means that good peer review strongly contributes to unequal resource allocation in science.

REFERENCES

- Abbott, A. 2001. *Chaos of Disciplines*. Chicago: University of Chicago Press.
- Alberts, B., B. Hanson, K. L. Kelner. 2008. "Reviewing Peer Review." *Science* 321:15.
- Azar, O. H. 2008. "Evolution of Social Norms with Heterogeneous Preferences: A General Model and an Application to the Academic Review Process." *Journal of Economic Behaviour and Organization*, 65:420–435.
- Björk, B.-C., A. Roos, M. Lauri. 2009. "Scientific Journal Publishing – Yearly Volume and Open Access Availability." *Information Research*, 14, 1: <http://informationr.net/ir/14-1/paper391.html>.
- Bornmann, L. 2011. "Scientific Peer Review." *Annual Review of Information Science and Technology*, 45:199–245.
- Chang, J., C. Lai. 2001. "Is it worthwhile to pay referees?" *Southern Economic Journal* 68:457–463.
- Couzin, J. 2006. "... and how the Problems Eluded Peer Reviewers and Editors." *Science* 311:614–615.
- Crocker, J., M. L. Cooper. 2011. "Addressing Scientific Fraud." *Science* 334(6060):1182.
- Gewin, V. 2012. "Uncovering misconduct." *Nature* 485:137–139.
- Laband, D. N. and J. M. Piette. 1994. "Favoritism Versus Search of Good Papers. Empirical Evidence Regarding the Behavior of Journal Editors." *Journal of Political Economy* 102:194–203.
- Lamont, M. 2009. *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge, MA: Harvard University Press.
- Martins, A. 2010. "Modeling Scientific Agents for a Better Science." *Advances in Complex Systems* 13(4):519–533.
- Northcraft, G. B. and A. E. Tenbrunsel. 2011. "Effective Matrices, Decision Frames, and Cooperation in Volunteer Dilemmas: A Theoretical Perspective on Academic Peer Review." *Organization Science* 22(5):1277–1285.
- Roebber, P. J., D. M. Schultz. 2011. "Peer Review, Program Officers and Science Funding." *PLoS ONE* 6(4):e18680: <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0018680>.
- Seinen, I. And A. Schram . 2004. "Social Status and Group Norms: Indirect Reciprocity in a Repeated Helping Experiment." *European Economic Review* 50(3):581–602.
- Smith, R. 2006. "Peer Review. A Flawed Process at the Heart of Science and Journals." *Journal of the Royal Society of Medicine* 99:759–760.
- Squazzoni, F. 2010. "Peering into Peer Review." *Sociologica* 3: <http://www.sociologica.mulino.it/doi/10.2383/33640>.
- Squazzoni F, G. Bravo and K. Takács. 2012. "Does Incentive Provision Increase the Quality of Peer Review? An Experimental Study." *Research Policy*, In Press, DOI: 10.1016/j.respol.2012.04.014.
- Squazzoni, F. and C. Gandelli. 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 Takács, 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>.
- Turner, S. and R. Hanel. 2011. "Peer Review in a World with Rational Scientists: Toward Selection of the Average." *The European Physical Journal B* 84:707–711.

AUTHOR BIOGRAPHIES

FLAMINIO SQUAZZONI leads the GECS-Research Group on Experimental and Computational Sociology at the University of Brescia, where he teaches "Sociology of Organization" and "Leadership and Complex Organizations". He is review editor of the *Journal of Artificial Societies and Social Simulation*

and member of the management committee of ESSA-The European Social Simulation Association. He is author of “Agent-Based Computational Sociology” (Wiley, 2012). He is mainly interested to understand the impact of social norms and institutions on markets and science by combining experimental and agent-based computational research. His email address is squazzon@eco.unibs.it.

CLAUDIO GANDELLI is research fellow in the “MOSACC-Agent-Based Simulation Models to Support Web Tools for Enterprise Collaborative Knowledge Management” Research Project, co-funded by the Department of Social Sciences, University of Brescia, Project Group and Regione Lombardia. He received a degree in computer sciences and engineering at the University of Brescia and also works as consultant for web and software applications for business. His email address is claudio.gandelli@gmail.com.