OPENING THE BLACK-BOX OF REFEREE BEHAVIOUR. AN AGENT-BASED MODEL OF PEER REVIEW

Flaminio Squazzoni and Claudio Gandelli
Department of Social Sciences
University of Brescia
Via San Faustino 74/B, Brescia, 25122, Italy
E-mail: squazzon@eco.unibs.it

KEYWORDS
Peer review; referees; referee behaviour; reciprocity; agent-based model.

ABSTRACT

This paper investigates the impact of referee behaviour on the quality and efficiency of peer review. We focused especially on the importance of reciprocity motives to ensure cooperation between everyone involved. We modelled peer review as a process based on knowledge asymmetries and subject to evaluation bias. We built various simulation scenarios where we tested interaction conditions and manipulated author and referee behaviour. We found that reciprocity *per se* can have a negative effect on peer review as it tends to increase evaluation bias. It can have a positive impact only when purged by self-interest motivation and accompanied by disinterestedness and fairness standard.

INTRODUCTION

Peer review is the cornerstone of science. It allows scientists to experimentally pursue new lines of research through a continuous, decentralized and socially shared trial and error process and ensure the quality of knowledge produced. It, directly or indirectly, determines how all the resources of the science system, such as funds, careers, and reputation are allocated. Despite its importance, it is dramatically under-investigated and has no "experimental base" (Smith 2006).

One of the main challenges is to understand referee behaviour and increase commitment and reliability for everyone involved in peer review (Squazzoni 2010; Squazzoni and Takács 2011). Indeed, while journal editors and submission authors have clear reputational rewards, understanding incentives and motives of referees is more difficult. This is not trivial as it has been recently acknowledged that referees are dramatically overexploited and this could undermine their commitment (Björk, Roos and Lauri 2009). A survey estimated that peer evaluation is applied to more than 1 million journal articles per year, not to mention conferences, research proposals, fellowships and university/dept/institute productivity evaluation. Serious doubt has been casted by influential journal editors on the possibility that the peer review system could go on efficiently without reform (Alberts, Hanson and Kelner 2008).

Recent cases of misconduct and fraud called for a reconsideration of the potential traps of peer review. For instance, a group of scientists from South Korea published in 2005 an article on stem cell in *Science* that was based on false data. Certain myopic attitudes of editors influenced by "aggressively seeking firsts" and nine referees dazzled by the novelties of the paper implied that reviewing time was dramatically shortened: the referees took just 58 days to recommend the publication against the average of 81 days typical for this influential journal (Couzin 2006). More recently, the Stapel scandal gained public notoriety in the newspapers and the social media, where data for numerous studies conducted over a period of 15 to 20 years and published in many top journals in psychology were found to be fabricated (Crocker and Crooper 2011). It is worth noting first, that these cases can cause a misallocation of reputational credit in the science system with negative externalities on competition. Secondly, they can determine serious consequences on the credibility of science for external stakeholders.

Unfortunately, only a few studies in economics and social sciences tried to understand the behaviour of the figures involved in peer review and its consequences for the quality and efficiency of the evaluation process. One of the few topics studied has been the reviewing rate. For instance, Engers and Gans (1998) suggested a standard economic analytic model that looked at the interaction between editors and referees. They aimed to understand why referees were willing to ensure good quality of reviewing without any material incentives and whether improving these latter could increase the reviewing rate. They showed that payment could potentially motivate more referees to agree to review a submission, but raising the review rate could bring referees to believe that refusing to review a submission could not impose serious costs to the journal, as other referees could happily accept to review. This could cause a decrease of the reviewing rate and bring journals to increase payment to compensate this effect, generating an escalation of compensation unsustainable for journals.

On the other hand, Chang and Lai (2001) studied reviewing rates and came to different conclusions. They suggested that if reciprocity motives were present, which influenced the relationship between journal

Proceedings 26th European Conference on Modelling and Simulation ©ECMS Klaus G. Troitzsch, Michael Möhring, Ulf Lotzmann (Editors)
ISBN: 978-0-9564944-4-3 / ISBN: 978-0-9564944-5-0 (CD)

editors and referees and provided room for reputation building for referees, a possible snowballing effect could emerge that increased the referee recruitment rate. They showed that if accompanied by material incentives, this effect could significantly improve the review quality. This was also confirmed by Azar (2008), who studied the response time of journals. He suggested that shorter response times of journals in specific communities were due to the strength of social norms, towards which referees were extremely sensitive..

The importance of social norms in peer review has also been confirmed by recent experimental findings, where results showed that indirect reciprocity motives more than material incentives provided reasons to expect commitment and reliability by referees (Squazzoni, Bravo and Takács 2011). By manipulating incentives in a repeated investment game, which was modified to mirror peer review mechanisms, they found that adding material incentives to referees undermined pro-social motivations without generating higher evaluation standards. These results are in line with game theory-oriented experimental behavioural studies, where the importance of reciprocity, direct and indirect, for cooperation in situations of asymmetries of information and potential cheating temptations, which could represent a typical situation of peer review, was widely acknowledged (Bowles and Gintis 2011).

Following this perspective, it is possible to argue that 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 previous publication as a means to protect their own impact. 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. In this respect, looking at referee/author interaction, peer review could be viewed as a "helping game", where referees would act as "donors" who choose whether to give or not a positive benefit to "recipients" (i.e., submission authors) by paying a cost (i.e., time and effort needed to reviewing), which is smaller than the recipients' benefit (i.e., higher impact factor, more citations, higher academic reputation) (Seinen and Schram 2004).

It is worth mentioning that these problems have recently been addressed in *Science*, where Alberts, Hanson and Kelner (2008) suggested the need for seriously reviewing peer review to improve its efficiency and guarantee its sustainability. Unfortunately, all current attempts to reform it, which have insisted especially on the importance of referee reliability and the need for measures to improve it, have followed trial and error approaches not supported by experimental investigation. Although some 'field experiments' on peer review were performed by certain journals or funding agencies, it is widely acknowledged that we lack sound experimental knowledge on essential peer review mechanisms, which can seriously support policy measures (Bornmann 2011).

Our paper is an attempt to contribute on this point by proposing a modeling approach (Martins 2010, Roebber and Schultz 2011; Thurner and Hanel 2011). Indeed, empirical research has serious problems in looking at essential aspects of peer review in general terms and investigating complex mechanisms of interaction such as peer review. We modelled a population of agents interacting as authors and referees in a competitive and selective science system. Our aim was to understand the impact of referee behaviour on the quality and efficiency of peer review and to test the impact of reciprocity strategies between involved agents.

The structure of the paper is as follows. In the second section, we will introduce the model, while in the third we will present various simulation scenarios and the simulation parameters. In the fourth one, we will illustrate our simulation results, while in the concluding section, we will present a summary of results and draw some implications for the current debate on peer review.

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:

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

We assumed that authors varied in terms of quality output depending on their productivity. More specifically, 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 means 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:

$$P = \frac{Q}{v*(1-Q)} \tag{2}$$

where P was the productivity gained and Q the submission quality score assigned by referees. 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 each tick. We assigned a heterogeneous value of M, after various explorations of the parameter space, as this mimicked 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. If not published, following the "winner takes all" rule characterizing science, we assumed that authors lost all the resources invested for submitting. This meant that, at the present stage, we did not consider the presence of a stratified market for publication, where rejected submissions could be submitted elsewhere, as happens in reality (Weller 2001).

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 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 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:

$$S = \frac{1}{2} R_r[1 + (Q_a - \mu_r)] \tag{3}$$

where R_r was agent resources, Q_a was the real quality of the submission and μ_r was the referee's expected quality.

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%). It is worth noting that certain empirical studies showed that these types of errors were more frequent than expected, especially in grant applications (e.g., van den Besselaar and Leydesdorff 2007; Bornmann and Daniel 2007). For example, Bornamm, Mutz and Daniel (2008) examined EMBO selection decisions and found that between 26 and 48 percent of grant decisions showed these type of errors, underrating being more frequent (2/3 of cases).

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: Simulation parameters.

Parameters	Value
Initial resources	0
Fixed productivity gain	1
Publication selection rate	[0.25, 0.50, 0.75]
Publication productivity gain	[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

SIMULATION SCENARIOS

We built various simulation scenarios to test the impact of referee 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 (Casati *et al*. 2009). Obviously, this is a restrictive definition of the various functions that peer review covers in science. Here, we only considered the screening function. Neither the role of peer review in helping authors to add value to their submission by referee feedback (Laband 1990), nor its role in deciding the reputation of journals and their respective position in the market were considered (Bornmann 2011). For efficiency, we meant the capability of peer review of achieving quality by minimizing productivity lost by authors and reviewing expenses by referees.

In the first scenario (called "*no reciprocity*"), we assumed that, when selected as referees, agents always had a random probability of behaving unreliably, which was constant over time and not influenced by any past experiences. When selected as authors, agents always invested all their resources for publication, irrespectively of positive or negative past experience with their submission. In this case, there was no room for reciprocity strategies between authors and referees. In the second scenario (called "*indirect reciprocity*"), we assumed that agents, when selected as referees, were influenced by their past experience as authors. In case of previous publication, they reciprocated by providing reliable evaluations in turn when selected as referees. Note that in this case, authors were self-interested and did not consider the pertinence of the referee evaluation, only their publication success or failure in their previous submission. This meant that they reciprocated negatively in case of rejection and positively in case of publication even if they knew that their submission was of a low quality and wasn't worth a publication.

In the third scenario (called "*fairness*"), when selected as authors, agents formulated a pertinent judgment of the referee evaluation of their submission. They measured the fairness of the referee opinion by comparing the real quality of their submission and the evaluation rate received by the referees. If the referee evaluation approximated the real value of their submission (i.e., ≥ -10 %), they concluded that referees were 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 pure self-interest motivation of agents but on normative standards of conducts.

Finally, the last two scenarios (called "*self-interested authors*" and "*fair authors*") extended the previous two. In the "*self-interested authors*" scenario, 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 of the pertinence of the referee evaluation. In the "*fair authors*" scenario, in case of a pertinent referee evaluation received when authors, they reinforced their confidence on the good quality of the evaluation process and continued to invest everything 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, agents inferred by their experience of authors the overall situation of peer review standards and strategically acted consequently.

To measure the consequences of parameter manipulation under various scenarios, we built the following indexes: (1) evaluation bias, (2) productivity loss, (3) reviewing expenses and (4) career inequality. The first dealt with quality, 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 in order to measure how much resources were 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.

RESULTS

Tab. 2 shows the impact of referee behaviour on the quality and efficiency of peer review in various conditions of the publication rate (weak, medium, and strong selection). Data were averaged on 200 simulation run in any parameter conditions. 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 but at the expenses of referees' resources. Although in general increasing selection implied increasing evaluation bias, as expected, the "*fairness*" scenario implied lower bias and lower productivity loss by authors, although reviewing expenses were generally higher. Furthermore, it ensured higher resilience to changes in competition pressures. On the other hand, indirect reciprocity without fairness by authors implied higher evaluation bias and higher productivity loss when competition increased.

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

Scenario	Evaluation bias	Productivity loss	Reviewing expenses
Weak selection (75% published submissions)			
No reciprocity	14,10	5,69	23,47
Indirect reciprocity	12,58	6,51	44,16
Fairness	13,14	7,48	40,61
Medium-level selection (50% published submissions)			
No reciprocity	26,32	15,65	30,32
Indirect reciprocity	25,32	12,64	39,88
Fairness	15,68	8,60	38,68
Strong selection (25% published submissions)			
No reciprocity	28,00	15,01	29,47
Indirect reciprocity	43,12	16,92	33,39
Fairness	19,52	8,32	38,29

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. When authors reacted to referee evaluation only following their self-interest (i.e., being eventually published), the quality and efficiency of peer review drastically declined. Moreover, in case of authors' fairness, peer review dynamics even improved with the increasing competition.

Then, we calculated the system productivity of all the scenarios, by averaging the resources of all agents at the end of the simulation run. Considered the value of the "*no reciprocity*" scenario as a benchmark, "*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*" scenario determined an exponential growth of resources. Figures 1 and 2 compare system productivity accumulation in weakly and strongly selective environments. Results showed that in the "*fair authors*" scenario, stronger competition determined an exponential growth of resources.

Then, we also calculated the inequality of resource distribution in all the scenarios by measuring a Gini index (see Tab. 4). Results showed that, once introduced reciprocity, the best performing 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 did not radically change and was only exacerbated. This is coherent with findings of Squazzoni and Gandelli (2012): in a competitive systems where the "winner takes all" principle is the rule, such as in science, well functioning of peer review determines a unequal resource distribution as cumulative advantages for best scientists take place. 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

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

	Evaluation bias	Productivity loss	Reviewing expenses
Weak selection (75% published submissions)			
Self-interested	15,30	12,63	46,07
Fair authors	14,85	5,10	29,55
Medium-level selection (50% published submissions)			
Self-interested	30,52	25,63	45,74
Fair authors	14,24	4,00	15,96
Strong selection (25% published submissions)			
Self-interested	45,04	38,31	47,13
Fair authors	14,24	4,00	15,96

.

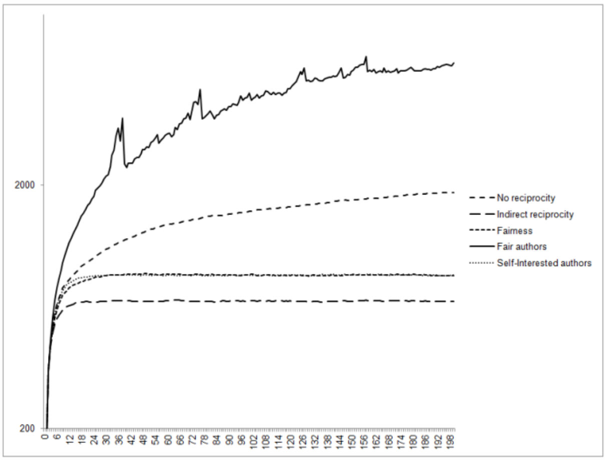

Figure 1: The impact of agent behaviour on system resource accumulation in weakly selective environments. In the x-axis, the number of simulation run.

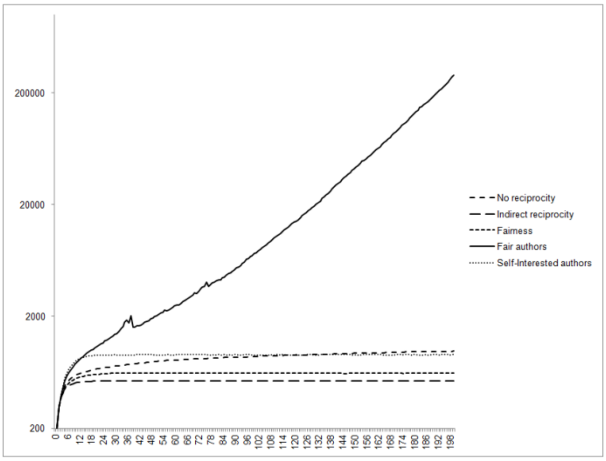

Figure 2: The impact of agent behaviour on system resource accumulation in strongly selective environments. In the x-axis, the number of simulation run.

Table 4: The Gini index in all the scenarios in weakly and strongly 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
Weak selection (75% published submissions)	
No reciprocity	0.55
Indirect reciprocity	0.34
Fairness	0.36
Self-interested authors	0.34
Fair authors	0.74
Strong selection (25% of published submissions)	
No reciprocity	0.47
Indirect reciprocity	0.34
Fairness	0.45
Self-interested authors	0.35
Fair authors	0.88

CONCLUSIONS

One of the most convincing explanations of why referees tend to cooperate with editors and authors in ensuring good quality and efficiency of peer review has been reciprocity (Chang and Lai 2001; Squazzoni, Bravo and Takács 2011). By viewing peer review as a cooperation game, we can argue that referees could rationally bear the cost of reviewing so as to establish good standards of reviewing in prospect of benefiting from cooperation by other referees when they are submitting authors. This is a typical mechanism through which reciprocity can sustain cooperation in repeated interaction as it can transform the cost for referees in an investment for potential future benefit.

Recent experimental findings corroborated this view. By comparing experimental treatments where referees were expected to be selected as future authors with treatments where referees could not be also authors in turn, Squazzoni, Bravo and Takács (2011) found that the quality of peer review significantly increased when roles were changing and that this made the introduction of material incentives for referees even superfluous. By alternating roles of referees and authors, indirect reciprocity motives took place that generalised cooperation.

Although highly abstracted and brutally simplified compared to in-depth empirical or experimental research, our simulation results allow us to cast doubt on the strength of reciprocity *per se* to explain the quality and efficiency of peer review. Indeed, our results indicate that if reciprocity is influenced by self-interest motives, its potential positive effect for cooperation in peer review is neutralised. This suggests that reciprocity motives of everyone involved in peer review should associate with fairness to ensure quality and efficiency in the evaluation process. A possible conclusion is that social norms that reflect ethical, cultural and competence-based standards of conduct for scientists could be more important than strategic behaviour to understand peer review and its important contribution for promoting excellence in science (Lamont 2009).

Obviously, neither theoretical generalisation, nor serious policy implications can be drawn from our simulation study. We did not pretend to convey realism here. Our results were based on a highly abstract model and so every conclusion should be taken cautiously. For instance, in reality, reviewing is not equally distributed over the population and also editors are important to provide room for reputation building and reciprocity motives of referees. These are certainly points for future development of our research.

However, one of the most crucial challenges for any future development of our work is to fill the gap between theory and empirical observation. Although it is difficult to obtain empirical data that point to the behaviour of agents involved in peer review, especially at the scale needed to look at general aspects, a possible development could be to empirically test referee behaviour in highly representative journals.

Certain empirical measures have already been developed that could be used to test our findings. For example, Laband (1990) examined referee reliability by measuring the lines of the report text sent to submission authors, assuming that the longer the text, the higher the quality of the referee comments and more reliable the final score assigned to the submissions. This is a brilliant idea to build an *ex-ante* measure that could complete the most common *ex-post* measures of peer review validity, such as published citations or the fate of rejected submissions (Weller 2001).

Let us suppose that we can select a set of representative journals, possibly of different scientific communities and to have access to the list of referees and authors and to the referee reports. Let us suppose to apply the Laband's measure to assess the *ex-ante* validity of peer review and to measure the *ex-post* validity, e.g., by collecting data on citation of published articles or, even better, by analyzing the fate of rejected submissions, so as to build a statistical measure of reliability of referee evaluation. By measuring the link of referees and authors in these journals, we could test whether room for reciprocity and fairness took place that could influence the quality of the evaluation.

REFERENCES

Alberts, B., B. Hanson, and 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 Behavior and Organization*, 65, 420-435.

Björk, B.-C., Roos, A., M. Lauri. 2009. "Scientific Journal Publishing – Yearly Volume and Open Access Availability". *Information Research*, 14, 1, accessible at: <http://informationr.net/ir/14-1/paper391.html>.

Bornmann, L. 2011 "Scientific Peer Review". *Annual Review of Information Science and Technology*, 45, 199-245.

Bornmann, L. and H.-D Daniel. 2007. "Convergent Validation of Peer Review Decisions Using the *H* Index: Extent of and Reasons for Type I and Type II Errors". *Journal of Informetrics*, 1, 3, 204-213.

Bornamm, L., R. Mut, R. and H.-D. Daniel. 2008. "How to Detect Indications of Potential Sources of Bias in Peer Review: A Generalized Latent Variable Modeling Approach Exemplified by a Gender Study". *Journal of Informetrics*, 2(4), 280-287.

Bowles, S. and H. Gintis. 2011. *A Cooperative Species. Human Reciprocity and Its Evolution*, Princeton, NJ, Princeton University Press.

Campanario, J. M. 1998. "Peer Review for Journals as It Stands Today – Part I". *Science Communication*, 19, 3, 181-211.

Casati, F., M. Marchese, A. Ragone and M. Turrini. 2009. "Is Peer Review Any Good? A Quantitative Analysis of Peer Review". DISI University of Trento, Technical Report # DISI-09-045, accessible at: <http://eprints.biblio.unitn.it/archive/00001654/>.

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.

Engers, M. J. and Gans. 1998. "Why Referees Are Not Paid (Enough)". *American Economic Review*, 88, 1341-1349.

Laband, D. N. 1990 "Is There Value-Added From the Review Process in Economics? Preliminary Evidence from Authors". *The Quarterly Journal of Economics*, 105, 2, 341-252.

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.

Roebber, P. J. and D. M. Schultz, D. M. 2011. "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.

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, doi: 10.2383/33640, accessed at: http://www.sociologica.mulino.it/doi/10.2383/33640

Squazzoni, F., G. Bravo and K. Takács. 2011. "Does Incentive Provision Increase the Quality of Peer Review? An Experimental Study". Proceedings of ICORE 2011-International Conference on Reputation, Montpellier, 19th September 2011, accessible at: http://2011.icore.name/

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 K. Takács. 2011. "Social Simulation that 'Peers into Peer Review'". *Journal of Artificial Societies and Social Simulation*, 14(4) 3, accessible at: <http://jasss.soc.surrey.ac.uk/14/4/3.html>.

Thurner, 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.

van den Besselaar, P. and L. Leydesdorff. 2007. "Past Performance as Predictor of Successful Grant Applications. A Case Study". Den Haag, The Netherlands: Rathenau Institute, accessible at: <http://www.leydesdorff.net/magw/magw.pdf>.

Weller, A. C. 2001. *Editorial Peer Review: Its Strengths and Weaknesses*. Medford, NJ: Information Today, Inc.

AUTHOR BIOGRAPHIES

FLAMINIO SQUAZZONI is assistant professor of economic sociology at the University of Brescia, where he leads the GECS-Research Group on Experimental and Computational Sociology (`www.eco.unibs.it/gecs`). He is review editor of *JASSS* and member of the management committee of *ESSA*-The European Social Simulation Association. He is author of the book *Agent-Based Computational Sociology* (Wiley, Chichester, 2012, see: `www.eco.unibs.it/computationalsociology`). His e-mail is: `squazzon@eco.unibs.it` and his web-page can be found at: `http://unibs.academia.edu/FlaminioSquazzoni`.

CLAUDIO GANDELLI is research fellow in the MOSACC-Agent-Based Simulation Models to Support Web Tools for Entrerprise 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 e-mail address is: `c.gandelli@gmail.com`.