Disagreement for control of rational cheating in peer review: a simulation

(Extended Abstract)

Mario Paolucci
Institute of Cognitive Sciences and Technologies
Italian National Research Council
Via Palestro 32, Roma, Italy, 00185
mario.paolucci@istc.cnr.it

Francisco Grimaldo
Departament d'Informàtica
Universitat de València
Av. Universitat, s/n, Burjassot, Spain, 46100
francisco.grimaldo@uv.es

ABSTRACT

Understanding the peer review process could help research and shed light on mechanisms that underlie crowdsourcing. We present an agent-based model of peer review built on three entities - the paper, the scientist and the conference. The model allows us to define a rich model of scoring, evaluating and selecting papers for conferences. Some of the reviewers apply a strategy (called "rational cheating") aimed to prevent papers better than their own to be accepted. We show how programme committee update, based on disagreement control, can remove rational cheaters.

Categories and Subject Descriptors

I.2.0 [**ARTIFICIAL INTELLIGENCE**]: General—*Cognitive simulation*

General Terms

Algorithms, Human Factors, Design

Keywords

Artificial social systems, Peer Review, Agent-based simulation, Trust, Reliability, Reputation, Cognitive Modeling, Rational Cheating

1. INTRODUCTION

Peer review, the process that scrutinizes scientific contributions before they are made available to the community, lies at the core of the social organization of science. Curiously, while the measurement of scientific production, that is, the process that concerns the citation of papers - scientometrics - has been an extremely hot research issue in the last years, we can't say the same for what concerns the process of selection of papers, although some attention has been focused on its shortcomings [3, 2].

Although being extremely important, the actual effectiveness of peer review in ensuring quality has yet to be fully investigated. While the heterogeneous review approach to a decision between two options is supported by Condorcet's jury theorem, if we move beyond simple accept/reject decisions by considering scoring and ranking, we find several kinds of potential failures that are not waived by the theorem. To understand and possibly to apply policies to peer review, we need more evidence coming from both the analysis and review of the process as it is, as well as from the creation of numerical, agent-based models, that could be validated both on the micro and the macro level.

The literature of simulation models about peer review is scarce. In [5], the authors focus on an optimizing view of the reviewer for his/her own advantage. To this purpose, they define a submission/review process that can be exploited by a *rational cheater* [1] strategy in which the cheaters, acting as reviewers, reject papers whose quality would be better than their own. They find out that a small number of rational cheaters reduces rather quickly the process to random selection. In this paper, we propose an more complete agent-based model of peer review and we test how a simple mechanism based on disagreement control could help controlling this kind of cheating.

2. THE PEER REVIEW MODEL

The key entities we identify within the peer review process are: the *paper*, the *scientist* and the *conference*. Thus, the proposed model represents the peer review problem by a tuple $\langle S, P, C\rangle$, where S is the set of *scientists* playing both the role of authors that write *papers* and the role of reviewers that participate in the PC of a set of *conferences* C. Papers produced by scientists have an associated value representing their intrinsic value, and receive a review value from each reviewer. These values are expressed as integers in an N-values ordered scale, ranging from strong reject (i.e. value 1) to strong accept scores (i.e. value N).

Every scientist $s \in S$ is represented by a tuple of the form $s = \langle ap, aq, as, cd, rs, rt\rangle$. Regarding paper production, each scientist has an associated author productivity ap, meaning the number of papers uniformly written per year. The intrinsic value of each paper is calculated considering the author quality $aq \in [1, N]$ and the author skill value $as \in [0, 1]$. The latter represents production reliability, so that scientists write papers of value aq with probability as, and of random value with probability $(1 - as)$. Each scientist also has an associated reviewer skill value $rs \in [0, 1]$ and a reviewing strategy $rt \in \{\texttt{normal}, \texttt{rational}\}$. We model a noisy evaluation of papers, where the result of reviewing is accurate with probability rs, and completely random with probabil-

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

ity $(1 - rs)$. Furthermore, *rational* cheaters punish those papers whose intrinsic value is greater than his own author quality (by scoring them with the lowest reviewing value), thus trying to clear the way for their own papers.

Finally, conferences $c \in C$ are represented by a tuple of the form $c = \langle PC, av, I, pu \rangle$. Conferences employ a subset of scientists $PC \subseteq S$ as their programme committee, who accept those papers whose average review value is greater than the acceptance value av. Additionally, conferences maintain an image of each scientist that has ever been a PC member (I), accounting for the disagreements with the other reviewers. Disagreements are calculated on a paper basis as the difference between the review value given by the reviewer and the average review value for that paper. Thus, reviewer images are used to update the PC by discarding the pu% of reviewers with a higher ratio of disagreement. As a response to each call for papers, scientists decide to submit papers provided the distance between the estimated paper value (authors perform on their paper the same noisy evaluation seen before) and the conference acceptance value av is less than or equal to the cautiousness degree of the author, expressed by the integer value cd.

3. RESULTS

Here, we present the results of a set of simulations of the proposed model involving 1000 scientists and 10 conferences across 50 years. Each scientist writes 2 papers uniformly distributed over the year $(ap = 2)$. Paper intrinsic values and review values are expressed in a 10-values ordered scale from 1 to 10 $(N = 10)$. Authors' qualities (aq) follow a discretized bell shaped curve with mean 5.5 and symmetrically distributed between 1 and 10, in the hypothesis that average papers are more common than either excellent or bogus papers. Authors' skills (as) and reviewers' skills (rs) follow a Uniform distribution in [0.5,1], that we consider a moderate level of noise. With respect to the reviewing type (rt), we only show results with rational cheaters up to 30% since greater ratios reverse the system, ending up with no papers accepted at all. For the conferences, parameters have been set in order to reproduce two different experimental scenarios that we call *homogeneous condition* (i.e. all av are equal to 5.5) and *heterogenous condition* (i.e. av range from 1 to 10). The percentage of PC update is $pu = 10\%$.

Our research hypothesis is that the PC update mechanism proposed will effectively find out and expel the rational cheater scientists. The argument that rational cheaters will find themselves in disagreement with others every time they act strategically makes sense and, in fact, in figure 1 we can observe how rationals cheaters decrease substantially in the conditions where they are more abundant. The PC update mechanism results more effective in the homogeneous condition than in the heterogeneous one (two-sided t test with p-value of 0.036, comparing MR-30 and SR-30), since the PC update mechanism fails in moving rational cheaters away from the PC when the quality of the conference is low.

A deep analysis of the results has been conducted in order to elucidate the effects of reducing the presence of rational cheaters. Though, due to space reasons, we can barely mention some of its main conclusions. The simulations show a significant decrease of the number of papers that should be accepted, but end up being rejected. In turn, as rational cheaters are expelled, the number of accepted papers grows to approach that of conditions without rational cheaters.

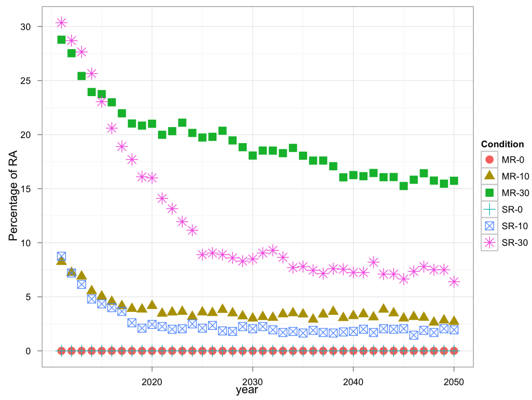

Figure 1: Percentage of Rational Scientists (rational cheaters) under different conditions: homogeneous (SR) and heterogeneous (MR) conditions with initial percentages from none to 30%.

4. CONCLUSIONS AND FUTURE WORK

Results from our simulations show how the mechanism we introduced to control disagreement in the PCs is also effective in removing most of the rational cheaters from the process. The benefit for the system can be measured in terms of the growing number of accepted papers and the decrease in the number of mistakes (good papers rejected).

A next step in this research would be to ground our model against data extracted from one of the several automated conference review systems. However, this data has proven surprising difficult to obtain. Not only the authors' queries to the owners of those systems went unanswered, but we have come to learn that other researchers had the same situation (none of [4, 5] manages to ground their assumption either). The difference between the immediate availability of publication and citation data is especially striking.

5. ACKNOWLEDGMENTS

Work supported by the Spanish MICINN, Consolider Programme and Plan E funds, European Commission FEDER funds and Universitat de València, under Grants CSD2006-00046, TIN2009-14475-C04-04, UV-INV-AE11-40990.62, as well as under the FuturICT coordination action.

6. REFERENCES

[1] D. Callahan. Rational Cheating: Everyone's Doing It. *Journal of Forensic Accounting*, pages 575+, Oct. 2004.

[2] F. Casati, M. Marchese, A. Ragone, and M. Turrini. Is peer review any good? A quantitative analysis of peer review. Technical report, Ingegneria e Scienza dell'Informazione, University of Trento, 2009.

[3] B. D. Neff and J. D. Olden. Is Peer Review a Game of Chance? *BioScience*, 56(4):333–340, Apr. 2006.

[4] F. Squazzoni and C. Gandelli. San Matthew Is Always There. A Study on Peer Review and the Scientific Community Structure. 2011.

[5] S. Thurner and R. Hanel. Peer-review in a world with rational scientists: Toward selection of the average. Aug. 2010.