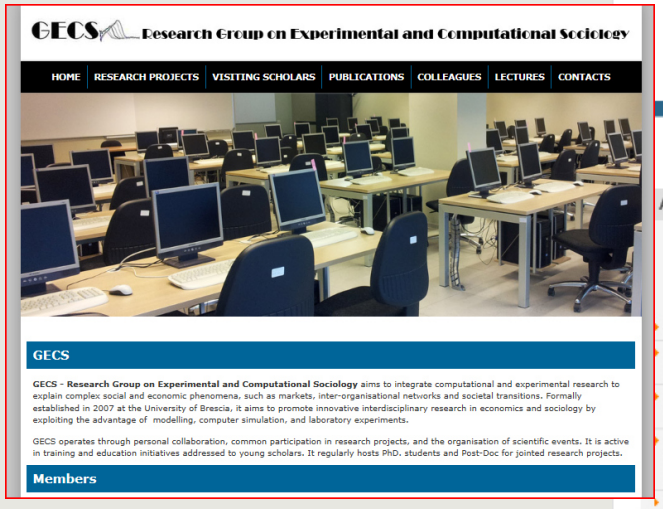
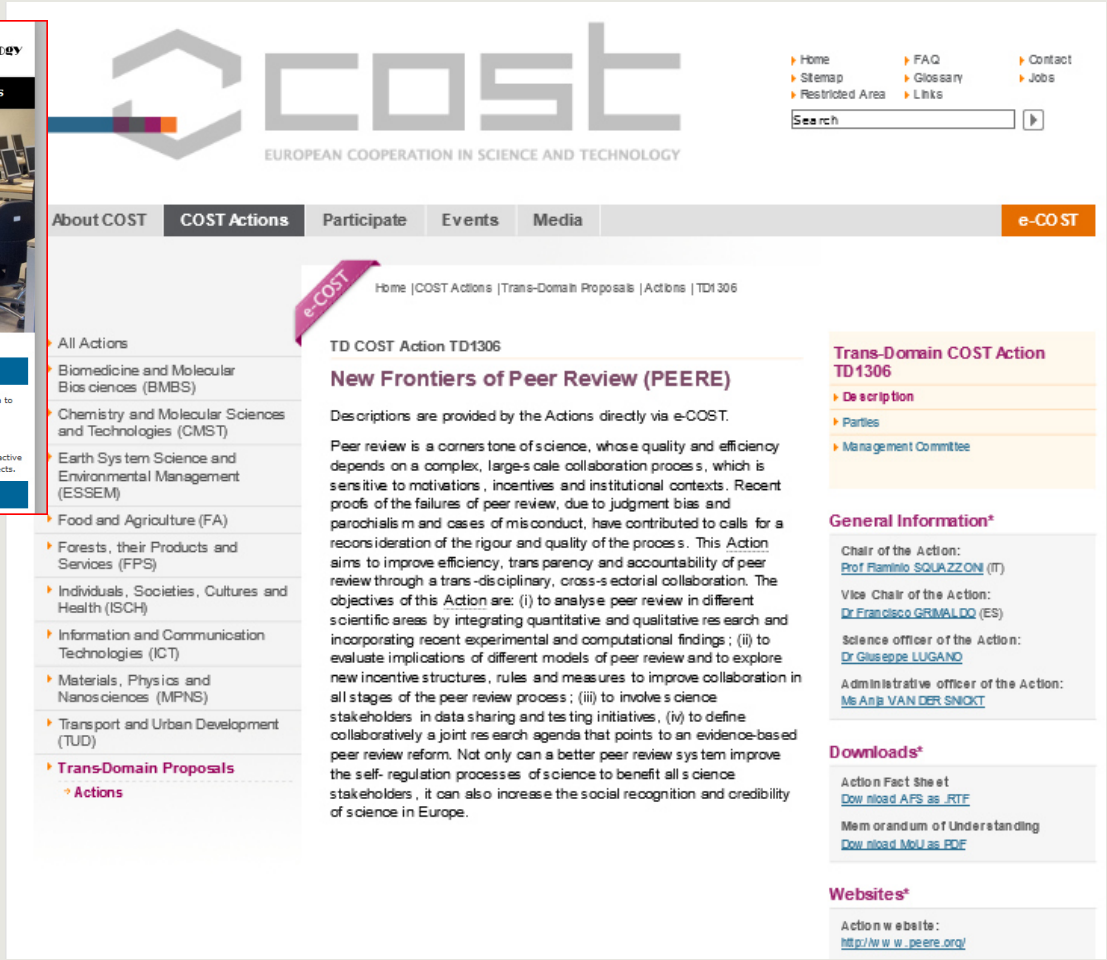
Competition, serious “gamification” and scientist misbehaviour: Can quantitative indicators and rankings be neutral and context-free measures of the quality of science?

Flaminio Squazzoni
GECS Research Group
University of Brescia, Italy
flaminio.squazzoni@unibs.it
www.eco.unibs.it/gecs/squazzoni.html

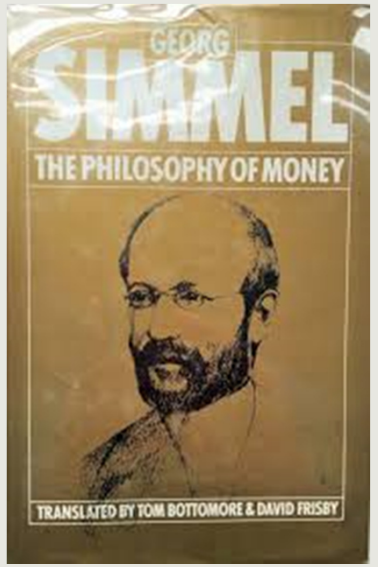
GECS: www.eco.unibs.it/gecs
PEERE: www.peere.org





Reputation as money, citations as prices

- Reputation/citations = money/prices as a means of exchange that regulates the science system
- Money and prices as quantification devices triggered calculative rationality of individuals, including allocation and discretization of time, strategization of effort/output measures, value recognition among people, status and power

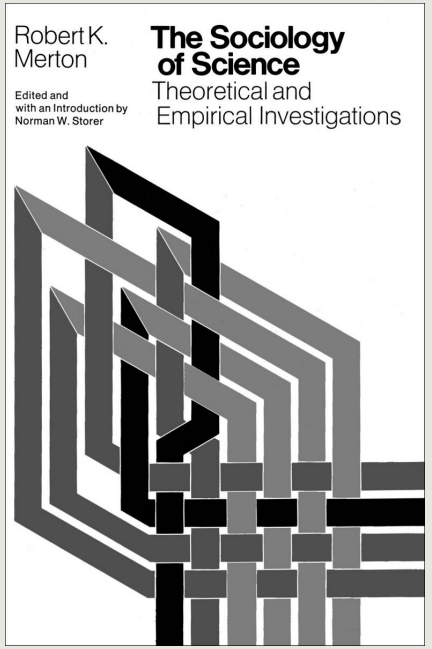


Reputation

- Reputation is a complex artefact
- Scientists built disciplines, institutions and associations to self-regulate and manage reputational credit allocation
- Reputation is productive if competitive spirits (i.e., the "priority for reward" game) are constrained by strong social norms
- Attention and signalling devices:

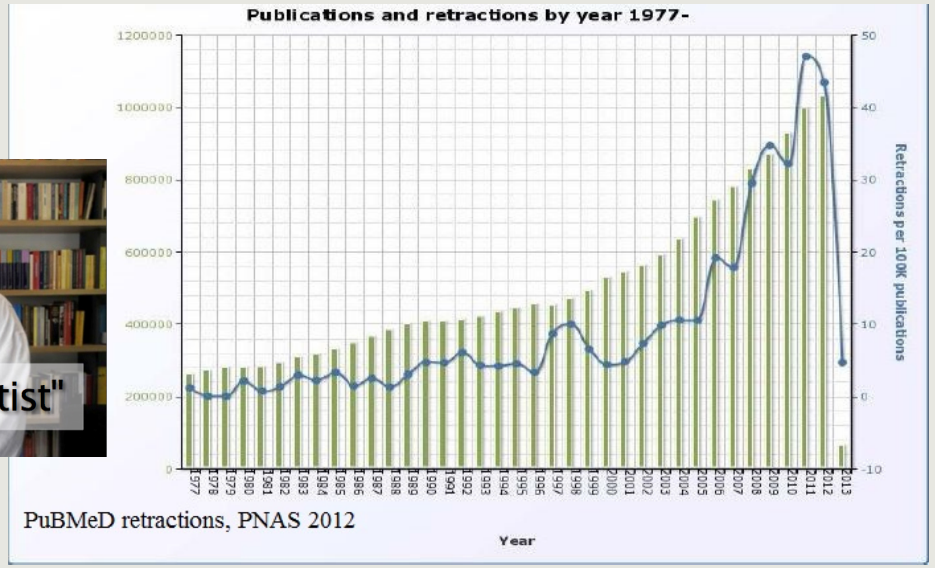
  metrics could help scientists to deal with

  coordination problems



Our “now”

- Growing competition at all levels
- Increasing role of scientists from emerging countries, exposed to strong competitive rewards
- Growing fragmentation and knowledge specialisation
- Serious concerns on allocation problems



“Rankitude”

- Rankings are natural social artefacts
- They are built-in competition devices
- They are a reference with an objective allure that are used to allocate power/reputation resources
- They tend to transform relations in “serious games”
- In times of scarce attention the “rankitude” could bring people to easy, broad-tent view conclusions about value of people independently of contexts and situations

Rankings & preferences

- Salganik, Dodds and Watts (2006, Science)
- 14000 participants who were shown a list of 48 unknown songs in two experimental conditions (independent and social influence)
- Exp 1: previous downloads in a grid; Exp 2: list

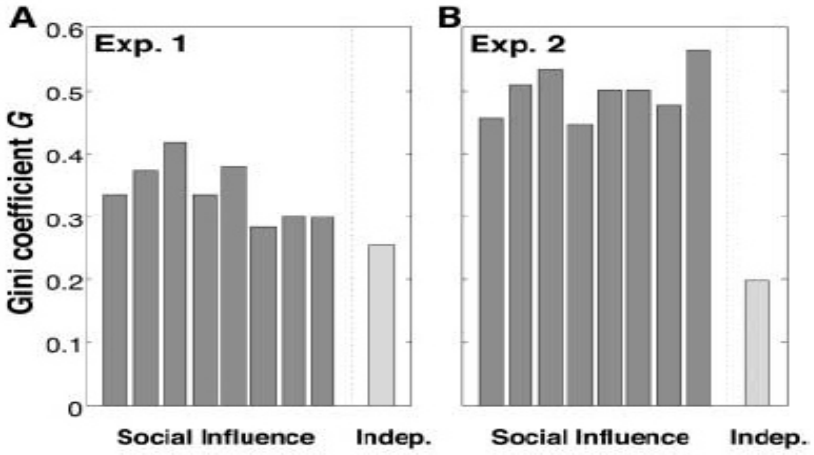


Fig. 1. Inequality of success for social influence (dark bars) and independent (light bars) worlds for (**A**) experiment 1 and (**B**) experiment 2. The success of a song is defined by m_i, its market share of downloads ($m_i = d_i / \sum_{k=1}^{S} d_k$, where d_i is song i's download count and S is the number of songs). Success inequality is defined by the Gini coefficient $G = \sum_{i=1}^{S} \sum_{j=1}^{S} |m_i - m_j| / 2S \sum_{k=1}^{S} m_k$, which represents the average difference in market share for two songs normalized to fall between 0 (complete equality) and 1 (maximum inequality). Differences between independent and social influence conditions are significant ($P < 0.01$) (*18*).

Examples

- MIUR and ANVUR in Italy and Norwegian Association for Higher Education Institutions (2004-2010): a performance indicator used to allocate a percentage of the total funds
- Pros: it mapped productivity differences, stimulated low-quality institutions, paid-off more active excellent centres
- Cons: no “neutrality” across disciplines and so penalised certain domains; rankings were used internally as a political means to allocate resources and compare individuals; it frustrated specialists in certain fields by exposing them to conflicting incentive schemes
- PAQ Research quality assessment of University of Brescia

NAHEI report: conclusions

- "Simplicity and transparency of the point system mean that, even for departments where local efforts to prevent that the indicator is used in undesirable ways, it is difficult to prevent if from playing a role at the individual level. Experience with bibliometric measures shows that when these types of indicators first exist and are readily available, they will often be used in both intended and unintended ways"
- The deep "reflexivity" of science system

Competitors' detection!

Eur. Phys. J. B 84, 707–711 (2011)
DOI: 10.1140/epjb/e2011-20545-7

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

Peer-review in a world with rational scientists: Toward selection of the average

S. Thurner[1,2,3,a] and R. Hanel[1]

[1] Section for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria
[2] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA
[3] IIASA, Schlossplatz 1, 2361 Laxenburg, Austria

Received 6 July 2011 / Received in final form 27 October 2011
Published online 23 November 2011 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2011

PDF (2,799 KB) | PDF Plus (586 KB)

FRANCISCO GRIMALDO and MARIO PAOLUCCI, *Advs. Complex Syst.* DOI: 10.1142/S0219525913500045

A SIMULATION OF DISAGREEMENT FOR CONTROL OF RATIONAL CHEATING IN PEER REVIEW

FRANCISCO GRIMALDO
Departament d'Informàtica, Universitat de València, Av. de la Universitat, s/n, Burjassot, 46100, Spain

MARIO PAOLUCCI
Institute of Cognitive Sciences and Technologies, Italian National Research Council, Via Palestro 32, Roma, 00185, Italy

Received: 27 August 2012
Revised: 28 November 2012
Accepted: 26 December 2012
Published: 14 March 2013

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Saint Matthew strikes again: An agent-based model of peer review and the scientific community structure

Flaminio Squazzoni*, Claudio Gandelli

Department of Social Sciences, University of Brescia, Via San Faustino 74/B, 25122 Brescia, Italy

ARTICLE INFO

Article history:
Received 24 October 2011
Received in revised form
22 November 2011
Accepted 19 December 2011

Keywords:
Peer review
Referees
Referee reliability
Matthew effect
Agent-based model

ABSTRACT

This paper investigates the impact of referee reliability on the quality and efficiency of peer review. We modeled peer review as a process based on knowledge asymmetries and subject to evaluation bias. We tested various levels of referee reliability and different mechanisms of reviewing effort distribution among agents. We also tested different scientific community structures (cohesive vs. parochial) and competitive science environments (high vs. low competition). We found that referee behavior drastically affects peer review and an equal distribution of the reviewing effort is beneficial only if the scientific community is homogeneous and referee reliability is the rule. We also found that the Matthew effect in the allocation of resources and credit is inherent to a 'winner takes all' well functioning science system, more than a consequence of evaluation bias.

The model

- 200 agents, authors and referees
- Endowment and resources
- Quality (as authors and referees) [E. 1]
- Publication investment and reviewing cost [Eq. 2]
- Resources multiplier (depending on publication) [*m*]
- Evaluation (randomly matching of authors and referees, noise)
- Publication selection rate (25, 50%, 75%)
- Referee behaviour: random, fair or “rational”

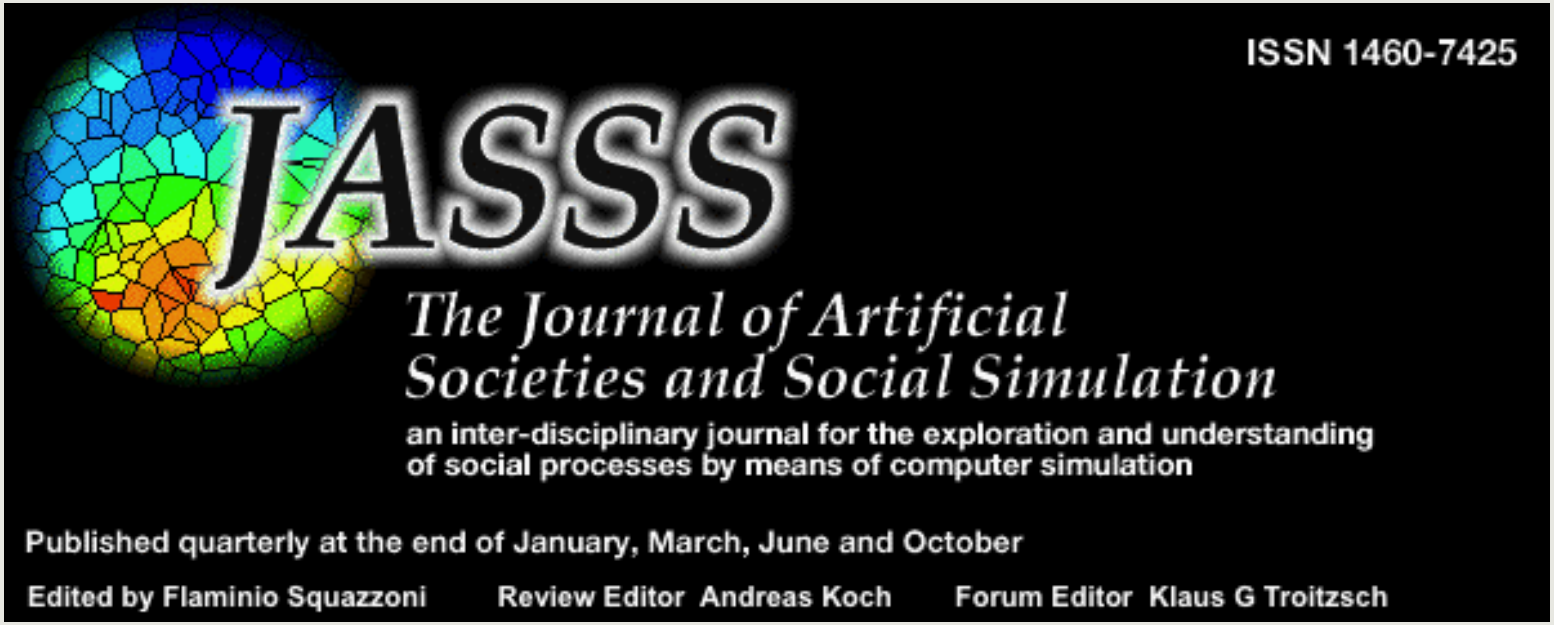
Implications

Scenario	Evaluation bias	Productivity loss	Reviewing expenses
Weak selection (75% published submissions)			
Random behaviour	16.51	7.68	25.98
Cheating	20.07	4,91	21.34
Medium-level selection (50% published submissions)			
Random behaviour	25.27	14.98	30.77
Cheating	56.63	28.02	32.21
Strong selection (25% published submissions)			
Random behaviour	29.42	15.00	29.42
Cheating	70.86	34.72	35.24

Table 2: Percentage of cheaters among the referees in the "cheating scenario" in various selection rate environments (values of cheaters in percentage on the total number of referees).

Selection rage	Cheaters
Strong selection	0.27
Medium selection	0.28
Weak selection	0.35

A recent personal experience



Lessons I believe to have learnt

- The mantra of rankings can erode variety in the science system by distorting resource allocation, reducing species niches and promoting homophily pressures (e.g., “top-five focal point” effect)
- Excessively simplified, although “big picture” valid quantitative indicators tend be politically used in power relations locally (e.g., “reflexivity”)
- Rankings must be improved, e.g., “real” productivity measures and be completed by more qualitative principles and evaluation criteria when they are used for resource allocation
- Scientific reputation cannot be fully captured by quantitative indicators and so indicators must be used intelligently to set up priorities and allocate resource

Thank you!