Flaminio Squazzoni

Peering Into Peer Review

(doi: 10.2383/33640)

Sociologica (ISSN 1971-8853)
Fascicolo 3, novembre-dicembre 2010

Ente di afferenza:
()

Symposium / Thinking Academic Evaluation after Michèle Lamont's *How Professors Think*

Peering Into Peer Review

by Flaminio Squazzoni

doi: 10.2383/33640

The evaluation of quality and excellence in science is difficult. We all recognize a good scientific article when we read it, but it is pretty complicate to express what "quality" precisely is. It must be said that, over a few centuries of the evolution of science as an institutional system, the scientific community has discovered practices, standards, and technologies that can help guide intuition which forms an essential part of evaluation, toward sound, robust, and socially shared criteria. However, it is surprising to note that, despite the dramatic increase in figures of science today, in terms of numbers of articles, journals, research grants, and funding agencies, peer review, which is the essential mechanism of evaluation in science, is poorly understood in scientific terms.

That said, unlike Michèle Lamont, who is captivated by the social and cultural aspects that mould decisions whenever experts collaborate face to face, I am personally more interested in understanding the social aspects that make it possible for peers to successfully cooperate in evaluating scientific products in an anonymous and decentralised system. Secret rooms where experts evaluate other peers is a social mechanism that humans discovered many centuries ago, e.g., when political leaders decide who to nominate as president of a crucial and prestigious public authority, or when business managers decide who to select for a position upgrade. Peer review, as we know it in science, is not just this: group deliberation. Rather, it is an anonymous, distributed and decentralised mechanism that makes the evaluation of complex scientific products possible through impersonal cooperation among different figures, by means of specific social norms.

My contribution to this symposium is to examine the social mechanisms behind peer review, by providing a complementary view of Michèle Lamont's book that emphasizes the need for a more general outlook at the basic process of evaluation in science and the added value of modelling to investigate this topic [Lamont 2009]. In particular, I will introduce recent experimental findings that help us to illuminate certain aspects of peer review, in particular the relevance of social norms that regulate the evaluation process, the weakness of social sanctions in peer review, and the counterproductive effect of economic incentives on reviewers. By emphasizing the sociological perspective rather than "behind the scene" of experts' dark rooms, I think sociology can contribute more to the recent debate about peer review reform.

The paper is structured as follows. The first section elaborates on Michèle Lamont's book, by illustrating a complementary viewpoint that looks more generally at peer review and evaluation. The second section summarises the puzzle of peer review, by suggesting why it is so important and so disputed today. The third introduces recent experimental findings that help to understand the core mechanism of peer review, such as the interplay of incentives and moral norms for cooperation between the social figures involved in evaluation. As we will see, the experimental literature provides fundamental insights to understand why social sanctions, pivotal to support cooperation, are poorly exploited in peer review at present. My position is that it is more important to focus on these aspects by keeping a general outlook to evaluation in science, rather than discussing the unavoidable and abiding heterogeneity and difference in evaluation standards between disciplines. The experimental insights have also policy implications suggesting the inadvisability of offering material incentives to guarantee a more stable and productive peer review. As we will see, it is reasonable to expect that monetary incentives can backfire and undermine the moral motives that sustain peer review.

Concerning Lamont's Study on Scientific Evaluation

How Professors Think by Michèle Lamont is a brilliant example of empirical investigation on evaluation in science and the role of cognitive and social factors in making scientific quality and excellence possible. It provides a detailed account on what happens in expert panels, when scientists are called to negotiate evaluations about proposals in a multi-disciplinary context. It puts the rationality of peer review under the spot light, by emphasizing the role of emotion, rhetoric arguments, and other cognitive aspects in group decision making. In doing so, this book pays good service to the sociological approach to these issues, in particular from a constructivist and cultural perspective. It argues that scientific quality and excellence are the result of complex social processes, rather than being "objective" and standardizable outcomes.

This is far from being a trivial issue, at least in sociology. The relevance of peer review and quality evaluation for the advancement of science is without doubt relevant, and evaluation and peer review are obviously influenced by significant social aspects, such as moral norms and reputational incentives, which are by definition sociological topics. Notwithstanding this, there have been few sociological investigations on this topic so far.

At the beginning of June 2010, I looked at 47 sociology journals listed in JSTOR, which included the most influential journals in our discipline, such as *The American Sociological Review* or *The American Journal of Sociology*. It was frustrating to note that articles on peer review in sociology could be counted on the fingers of one hand and that none of them crucially touched the present debate. This is treated better in biomedical sciences, physics or engineering. Therefore, Lamont's book is welcomed since it suggests a careful, detailed and vivid analysis that positions sociology investigations within a core and disputed issue.

That said, Lamont's book has certain limitations, the most important of which are: *a*) it covers only one specific point of the scientific evaluation continuum, being focussed on evaluation of fellowship proposals by students in the humanities and social sciences; *b*) it restricts sociological investigations to the disclosure of "behind the scene" aspects of evaluation, rather than locating sociology at the core of the peer review mechanism. These critical points are of paramount importance also to understand the peculiarity of the social technology illustrated in the book (group deliberation by panelists). I will discuss these two points.

Firstly, my understanding is that evaluating scientific proposals is like evaluating business start-ups in innovative and competitive markets. The challenge for evaluators is to correctly predict a development process that requires a synergy of different aspects, e.g., personal and professional qualities of applicants, support and quality of their research environment, achievement of relevant data that support applicant's investigations, and the prospective impact of their potential scientific achievements. By looking at the evaluation process from the evaluators' angle, information asymmetries and uncertainty make it difficult to estimate in advance the future impact of any proposal. The point is that the same is not true when evaluation is focused on scientific output, such as journal articles [Baccini 2010, 38]. In these cases, the object of evaluation is scientific output that can be subject to peer scrutiny in a less biased and uncertain way.

This makes me suggest that Lamont's evidence is to be framed in a more general model of scientific evaluation that captures the whole continuum of science development. In my view, scientific research is to be viewed as a single process composed of different stages that generate different products. In each stage, peer review is involved through different social technologies that guarantee quality evaluation. Although it can take different forms, the mechanism is the same: quality evaluation is not exercised by agencies external to science, such as political institutions, taxpayers' political representatives or private companies, but by peers, who are mostly called to spend time and energy and provide their opinion not for strictly economic or material interests.

For the sake of simplicity, I suggest we assume three stages and three different scientific products in a continuum between input and output science system functions: funding (input), conference/meeting/workshops/presentations/papers (throughput), journal articles and publication (output). These phases have complex temporal interrelations, such as the influence of output (publications) on input (funding). More importantly, they entail different evaluation process and peer review technologies, the general purpose of peer review to guarantee the best allocation of resources in the science system (e.g., funds, reputation, and careers) being understood. Each stage of this process poses different challenges to evaluation. Let me briefly summarise some of them.

The funding stage can be applied to careers (e.g., PhD bourses, fellowship grants, and scientific leaders' recruitment), research proposals (e.g., experimental, basic, or applied research) or university structures (e.g., labs, and departments). The challenge for evaluation here, in particular in the case of fellowship grants or research proposals for young researchers, is to predict the value of the applicants and proposals in terms of excellence and future achievement, in presence of relevant information asymmetries and difficult research outcome predictability. The rule is "the more you know about the applicant, the better it is" (e.g., CV, past publications, and reference letters), any personal information about applicants being pivotal to understand their potential achievement rather than a possible bias introduced into the evaluation, as occurs when evaluation deals with scientific outputs, e.g., submitted articles. Here, the most suitable evaluation technology is the expert panel (perhaps supported by anonymous peer review), in particular in case of multidisciplinary proposals like those illustrated in the Lamont's book.

This is because: a) funding agencies have little knowledge of the scientific environment and research fields, as well as of the individual skills of the potential experts, and inadequate internal competence, so that it is impossible that they can rely on, and adequately manage decentralised and impersonal peer review; b) the outsourcing of evaluation to a restricted group of experts guarantees an alignment of interests between the "principal" (the funding agency) and the "agent" (the expert) around the funding agencies' objectives, so that a common modus operandi and shared guidelines and priorities are established that can guide the evaluation process; c) the relationship between funding agencies and restricted groups of experts guarantees the legitimacy of the evaluation and sharing responsibility for funding decisions better than other mechanisms, such as unilateral evaluation by funding agencies or anonymous, distributed and decentralized peer review.

This said, evaluation and peer review for research funding are far from being a "perfect" and "efficient" social mechanism. As suggested by the literature, the critical points are as follows: *a*) the panellist group decision can be biased in many respects (e.g., "old boyism", gender, nationality, or language) [e.g., Coates *et al.* 2002; Bornmann and Daniel 2005] and in particular its collective dimension can even reinforce these cognitive biases ([e.g., Kerr and Tindale 2004; Obrecht, Tibelius, and D'Aloisio 2007; Benda and Engels 2010][1]; *b*) if allowed, external reviewers unlike a panel of experts do not always have a sufficient frame to evaluate proposals [e.g., Bornmann and Daniel 2005], so that they generate noise in the evaluation process; *c*) although relatively cheap for funding agencies, if taken as the only mechanism to allocate resources, funding through grants tends to cause the so-called "grant mania," generating excessive costs for the system in terms of decreased productivity [e.g., Goldsworty 2009]. The literature has suggested reforms to solve these critical points, among which the adoption of "the reader system" (i.e., evaluators read all the submissions in their field, not just one) that is expected to reduce these biases [e.g., Jayasinghe, Marsh, and Bond 2006] and the abolition of complex peer review group decision in favour of distributive funding, to maintain productivity caused by the "grant mania" [e.g., Spier 2002b; Gordon and Poulin 2009]. The debate is still open on this point.

The case of publications and scientific outputs in general is different. Here, the evaluation aims to guarantee that only excellent/high quality research work gets published. The challenge for reviewers is to evaluate research work in presence of

[1] The literature cited in the text provides a perspective on evaluation biases different from Lamont's. Lamont [2009, 158] emphasizes that "the fact that panelists must convince one another of the value of a proposal certainly contributes to their belief in the legitimacy of the process. In contrast, evaluations of journal submissions are conducted in the privacy of a reviewer's office or home and are not 'defended publicly'. This may leave more room for greater personal arbitrariness." The contributions cited in the text point out that committees and group panels tend to develop particularistic cultures that can even exacerbate shared cognitive biases.

relevant information asymmetries, in particular on data sources. Unlike evaluation for funding, here the “impersonality” and the absolute invisibility of authors are of paramount importance: the rule, in this case, is “the less you know about the author, the better it is” (e.g., anonymous article submission, avoiding cross-references). Given the unequal distribution of knowledge in the community, the increasing sophistication of research contents, the specialisation of the knowledge and the research technologies, as well as the increasing amount of research products in the science market (e.g., journals and books, see below), the only social technology that can guarantee a suitable evaluation is anonymous, distributed and decentralized peer review.

Also here, several problems arise. Let us assume that the goal of journal editors is to maximise the number of high-level articles published in each issue, to increase the reputation of their journals. On the other hand authors want to maximise their publications and in particular those in top-level journals, given time and resources constraints (the same assumptions could be made for the interaction between conference chairs and submission authors). Given that editors cannot manage the evaluation either themselves or by restricted groups of specialists, due to the increased complexity of research technologies and specialisation, all evaluation rests upon the capacity of the editors to have as complete a knowledge as possible of the scientific domains of reviewers. This is difficult and so editors suffer from imperfect knowledge of potential reviewers and information asymmetries on the quality of the reviews reported. Coordination problems arise as it is unlikely that each article submission is paired with the best expert in the field, so that evaluation can also suffer from an editor’s knowledge biases. Discussions about how to reform the reviewers’ selection process exploiting information and communication technologies, by reducing the gate keeping role of editors and promoting more open and bottom-up self-organisation of matching submissions and reviewers are under way. However, at present the problems are squarely on the editors’ shoulders and take place behind closed doors [e.g., Gura 2002].

Secondly, it must be said that the biases referred to panellists’ group decision mentioned above can also influence anonymous, distributed and decentralized peer review. Moreover, unlike the former case, reviewers sometimes do have poor contextual knowledge of the journal and its typical journal readership, so that they tend to overemphasize irrelevant specific aspects or under-emphasize important ones. Moreover, reviewers (in particular those with good reputation) are often overloaded and time and resource devoted to reviewing conflict with time and resource dedicated to publishing [Henderson 2001]. A survey conducted in 2007 on a sample of 3.000 scientists showed that most active reviewers (presumably the better ones) were overloaded and covered about 80% of all reviews, with an average of 14 reviews per year [Ware 2007].

Since reputational incentives for reviewing are low compared with publishing or with other scientific activities (e.g., grant competition, or presidency of important scientific institutions), scientists probably prefer to allocate time to other competing activities. It must be said that scientists' pay-offs, as presently settled, do not put reviewing top of the list. On the other hand, since reviewing is a private affaire between editors and reviewers with low probabilities of social sanctioning from others, cheating editors (e.g., returning back reports of a few lines that provide neither precise justification for the evaluation, nor insights for authors) is not a risky strategy for reviewers. As a matter of fact, if you cheat an editor, nobody knows it apart from the editor![2]

These are aspects on which we will return in the third section when we focus on certain experiments. The point here is that, in my view, it is possible to understand many general implications of evaluation and peer review in the sciences, as well as the heterogeneity of peer review technologies, just if a general perspective on evaluation is adopted. And so, it does not matter whether evaluation is done in sociology, physics or biology.

A second limitation of Lamont's book lies in its cultural perspective. My understanding is that, in case Lamont's perspective is viewed as the real and sole way of sociologically investigating evaluation in science, the risk is of a retreat of sociology into the disclosure of "behind the scene" aspects, leaving generalisation and policy implications to other scientists and so marginalising our discipline. This is the reason why I would like to suggest a more general approach to understand the relevance of incentives, moral norms and social practices that guide evaluation. I believe this point is crucial to place sociology in the current debate on peer review and its reform, which is so urgent today.

The Tangled Puzzle of Peer Review

Peer review is the cornerstone of science, the origin of which dates back to 1.752 when the Royal Society of London obtained responsibility for the "Philosophical

[2] It is likely that in small and cohesive academic communities social control and sanctions are stronger and therefore reviewers' cheating is risky. On the other hand, it is reasonable to suppose that in those communities the anonymous, distributed and decentralized peer review loses its strength in favour to other evaluation mechanisms. However, the presence of large and dispersed communities that compete at the global level is the present picture of scientific research, where standard peer review is fundamental and the problems mentioned in the text are important.

Transactions" [Spier 2002a]. It guarantees the quality of science by institutionalising the Mertonian value of "organised scepticism" [Merton 1973]. It is fundamental for institutional agencies to discriminate research proposals and decide which ones are worth funding, as well as for universities and research institutes to recruit the highest quality research staff. It is fundamental for scientists to increase the quality of their work, e.g., collecting referee reports, or peers' opinion during a conference. It is fundamental for conference chairs and journal editors to guarantee the best quality of presentations and articles and the reputation of their conferences or journals. As such, it is the pillar on which all the resources of the science market, such as funds, reputation and careers, are based [Smith 2006]. Therefore, it is also fundamental for policy makers to guarantee that tax payers or research fund investors are putting money into a credible system.

Peer review allows us to manage an increasingly sophisticated and complex scientific environment. In recent decades, the tremendous expansion of specific topics, interdisciplinary research and the increasing technical sophistication of research methods and tools have been reflected in the growing numbers of journals, conferences, and funding agencies for science, as well as in the continuous stratification of the scientific community into a mosaic of specialties. This is why individuals who are in charge of evaluating scientific products, such as journal editors, conference chairs and those responsible for research funds, strongly rely on peer review to guarantee time and quality of the evaluation, as they are "held hostage by the limited knowledge and relative ignorance of a single mind in this complex scientific system" [Grainger 2007, 5200].

Let us take *Nature*, one of the leading scientific journals in general, as an example. *Nature* formally established peer reviewing in 1953 and now receives about 10.000 papers every year. Its peer reviewing allows to end with about 7% of submissions eventually published and it is of paramount importance to maintain the high reputation of the journal. For conferences, a good example is AAMAS, a particularly successful multi-disciplinary annual conference on multi-agent systems, with no more than 10 years of history. It receives about 1.000 submissions annually from a varied set of authors, ranging from computer to cognitive scientists, from anthropologists to sociologists. Through an efficient web-based peer review system, where each submission has a responsible chair, reviewers are allowed to negotiate evaluations in case of conflict and authors to respond to their anonymous reviewers, AAMAS is capable of assigning conference slots to about 30% of submissions, guaranteeing the highest quality. For institutional agencies, one of the most influential examples is the National Institutes of Health of the US Government, which invest over USD 31 billion annually in medical research, 80% of it awarded through competitive grants based on peer review. The maximum quality is guaranteed by a stringent selection, with 20-25% of the proposal eventually granted.

Notwithstanding all this, peer review has attracted its share of criticism. Since the late 1980s, a survey of members of the Scientific Research Society revealed that only 8% of interviewees agreed that peer review worked well as it was [Chubin and Hackett 1990]. In recent years, the criticism has been fuelled by some scandals that also gained public notoriety. In 1996, *Social Text* published a paper where the author, Alan Sokal, a physicist, intentionally introduced false sentences. Sokal's intention was to parody the seriousness of post-modernism as a new paradigm for social sciences. Reviewers took this parody as a serious piece of science and the journal published it. In 1997 the editors of the *British Medical Journal* intentionally inserted eight errors into a short paper and asked reviewers to identify the mistakes. Out of 221 reviewers, the median number spotted was two [Couzin 2006]. In recent years, a Norwegian study published in *The Lancet* was found to be based on imaginary patients [Marris 2006]. But, the biggest impact was reached by the stem cell scandal caused by a group of scientists from South Korea who published in 2005 a paper in *Science* based on false data. Under the principle of "aggressively seeking firsts" that often influences leading journals and dazzled by the supposed novelties of the paper, the nine reviewers selected by *Science* took just 58 days to positively evaluate the paper, slightly more swiftly than the average of 81 days typical for this influential journal [Couzin 2006, 23-24].

These events triggered several disputes on peer review reforms, such as: whether and how to detect fraud and misconducts through peer reviewing [Nature 2006]; how to more intensively use new information technologies to improve communication during the reviewing process [Mandviwalla, Patnayakuni, and Schuff 2008]; how to combine peer review and bibliometric methodologies to provide public agencies with more detailed methods to evaluate research proposals [Abramo, D'Angelo, and Caprasecca 2009]; whether to reduce the fragmentation of publications, the proliferation of journals and the side-effect of the "publish or perish" rule [de Carvalho 2006; Dost 2008]; whether and how to pass from authorship to contributorship rule to clarify and increase of accountability of an individual scientist's contribution [Cho, McGee, and Magnus 2006]; how to promote training initiatives on the ethical issues of reviewing at the doctoral and post-doctoral level [Bosetti and Toscano 2008].

More radical analysts argued that these scandals confirmed that peer review was nothing but a black box [Smith 1997, 2006]. A survey conducted at the end of the 1990s in clinical neurosciences indicated that agreement between reviewers as to whether manuscripts should be accepted, revised or rejected was not significantly greater than expected by chance alone [Rothwell and Martyn 2000]. A survey in a cardiovascular journal conducted on data between 1997 and 2002 indicated a strong correlation between the ratings of the reviewers and the impact factor of the published papers [Opthof, Coronel, and Janse 2002], so prospective peer review errors can largely amplify their impact in the community. Quoting evidence from the National Science Board on the quality of the US research between the 1980s and the 1990s, Henderson [2001] argued that taxpayer-sponsored peer review science benefited from growing investment but the quality of output showed a clear degradation. In his view, this could be explained by the conflict of interest of reviewers who, being forced to invest time and energy for their publications, did not spend more than three hours to return their evaluation report to the journals.

By relying on the case of peer reviewing in funding agencies like the National Science Foundation, Spier [2002b] argued that peer review could even prevent funding for really innovative projects and ended up suggesting allocating funds by random or by following social priorities determined by lay panels. Innovation can be prevented because of cognitive biases of reviewers reported in an empirical study by Travis and Collins [1991] called "old boyism" ("old boys" network of reviewers tend to promote conservative research) and "cronyism" or "particularism" (reviewers affected by cognitive similarity tend to promote similar researches). These biases were particularly evident when peer review took place for funding agencies with the consequence that innovative proposals were often penalized [Mayo et al. 2006]. Horrobin [2001] argued that "a process that is central to the scientific endeavour as peer review has no validated experimental base" and that it refuses open scrutiny. We will come back to this point later, because it is of paramount importance. Spier's position follows certain arguments pointed out by Lamont [2009] as well. It questions the superiority of anonymous, distributed and decentralized peer review as a social mechanism to endogenously determine the science pay-offs and spontaneously guarantee research coordination.

To sum up, recently, peer review and scientific research evaluation have become a hot issue. This is for two reasons, which seem paradoxically in contrast with each other: 1) recent scandals about peer review failures and inefficiencies within the scientific community, and 2) the increasing emphasis on peer review and research evaluation in the society and the economy in general. On the one hand, it seems that within science, peer review has been increasingly criticised and, according to many also influential observers, a reform of it is needed [e.g., Alberts, Hanson, and Kelner 2008]. This is a reaction against the over-confidence that peer review was a perfect social mechanism that automatically prevented individual frauds and errors and allocated resources (funds, reputation, academic carriers) within science perfectly competitively. On the other hand, public and private institutional agencies that invest in scientific research and society in general, have great expectations in the thaumaturgical features of peer review to guarantee perfect allocation of scientific resource, as well as possibly other fields.

On this point, two aspects are of particular importance. Firstly, apart from idealistic representations, science is a largely "imperfect," extremely complex, and constantly increasing system, that rests upon moral norms and social practices, the self-reproduction, stability and good functioning of which are delicate and difficult to maintain. Secondly, the present complexity of the science landscape and the increasing competitiveness among scientists sees the moral norms and social practices that sustain peer review and quality evaluation ever more under attack.

To elaborate these two points, it is important to look at the science market and to what science has become today. In my view, most of the problems of peer review are in fact due to the increasing decentralized competition that regulates science in many respects; from research grants to tenures, from university appeal to journal publications. Although consubstantial to science progress, this increasing decentralized competition poses relevant challenges to peer review being an inescapable aspect of it, especially if some dramatic imperfections in the market mechanism are taken into account. I will focus on two aspects of this competition: the journal publishing market and the research grant market.

The figures of the market for scientific publication are impressive. The core scientific, technical and medicine publishing market was estimated between USD 7 and 11 billion, with journal prices increasing 200%-300% beyond inflation in the period 1975-1995, before the large diffusion of the Internet and the digital libraries. In their answer to a UK House of Commons' committee in 2004, Elsevier estimated that the 2.000 publishers in science, technology and medicine published 1.2 million peer-review articles annually. A recent survey has estimated that peer-reviewed scientific journal articles published world-wide in 2006 were approximately 1.346.000, with 70% covered by the ISI [Björk, Roos and Lauri 2009]. An estimate of the annual growth of this old and mature market was 3.5% yearly. From the 1970s to the 1990s, the growth in journals and publications has basically followed the growth in scientists and the public funding of science in general [e.g., Mabe and Amin 2001].

The point is that the publishing market is far from being a "perfect market" [European Commission 2006]. There are particular reasons for this. The market has been historically supported by public funds in different parts of its chain value, such as author and referee time and journal purchases. This negatively affected the strength of market selection. Since producers (authors) and consumers (readers) are the same subjects (researchers), the private and social values of the exchanged good may differ, when researchers behave respectively as authors or readers. The market chain value has been controlled by intermediaries so far, such as journal publishers and libraries, which have relevant information asymmetries toward producers and consumers but can influence both as well. The electronic era has dramatically exacerbated this situation, starting a race for size through dramatic hierarchical concentration or mergers on the publishers' side [McCabe 2002].

This has weakened the price sensitivity of consumers, introducing relevant biases in the market. Since authors want to publish in high impact journals, which attract other good potential authors and high interest from readers, the readers want good journals and easily identify the best ones, to avoid excessive selection costs (e.g., search for journals). In turn, journals can acquire reputational signals that might influence both authors and readers. Through the certification process, the market can present relevant barriers to entry and fall under the "Matthew effect," that is, a high inflation rates for best titles. Moreover, the market has been populated by different types of subjects so far, such as learned societies that manage a small number of highly successful long-standing journals, sold for low prices, as well as for profit publishers that have given new impetus to scientific publishing, by introducing new journals on average of lower quality and sold at much higher prices [European Commission 2006, 32]. The features of this market affect what is published and read and can deform the reliability of evaluation.

If we look at the research grant market, we find a similar competitive environment that can also have potential negative consequences on science. The journal publishing market guarantees at least the presence of exploitative niches for non-standard research, e.g., the possibility to publish innovative articles in less established or more specialised journals. The research grant market, based on centralized funding agencies and competitive bidding, is much more selective and takes up a relevant part of researchers' everyday work. For instance, a report on the "grant making mania" in biomedical sciences in US stated that, at the University of Pennsylvania, School of Medicine, faculty members spent more than 50 percent of their time working on grants [Schaffer 2009]. This is common to most faculty members in many experimental/applied sciences, but also increasingly diffuse in social sciences and humanities [Link, Swann, and Bozeman 2008].

Many reports allow us to reasonably suppose that competing bidding on research proposals causes a ratio of selection similar to the case of the National Institutes of Health of the US Government, i.e., about 25-30%, with impressive waste of resources, in terms of time and energy dedicated to writing failed proposals. An incredible example of wasted resources caused by this "grant mania" was the FIRB 2008 grant addressed to young PhD students and researchers (below 39 years) in all scientific fields by the Minister of University and Research in 2009 in Italy. For

EUR 45 million allocated through competitive bidding, 105 proposals were selected against 3.792 proposals, i.e., a ratio of less than 3% of successful proposals!

This evidence puts a spotlight on the economic efficiency of peer reviewed competitive research grant markets for taxpayers and investors and supports reform proposals. For instance, using statistics from the Natural Science and Engineering Research Council Canada, Gordon and Poulin [2009] calculated that the cost of preparing grant applications in 2007 exceeded that of giving every qualified investigators a conspicuous direct basic grant, guaranteeing at the same time the quality of applicants by already available reputational markers, such as university hiring, promotion and tenure proceedings, published articles and patents, whose collective scrutiny far exceeded the cost of grant peer review.

However this may be, the point is that these markets, as well as the entire science market, are not just "imperfect" from an economic point of view. Of course, this would not be a surprise, since we know very well the imperative of imperfect markets in many sectors of the economy and society. The point is that the reliability of these markets as science resource allocation mechanisms is based on certain social practices that do not fully respond to economic incentives and are not economic at all, therefore being largely unpredictable. Peer review has primarily a normative basis that is often taken for granted, as if it were self-reproducible, robust against external perturbations, and largely predictable social mechanism. This is not so. At this point, the question is: do we have scientific evidence of social mechanism that matters to understand evaluation and peer review? The next section aims to answer this question.

Peer Review under the Microscope

In my view, modelling and formalisation are pivotal to investigate specific and concrete social mechanisms that regulate social interaction, in that they allow us to capture essential aspects of it and to analyze its macro implications for large and more complex social systems. My suggestion is to view peer review in general as a triadic interaction between three figures, e.g., editors (or funding agencies or conference chairs), authors (or proposal applicants), and reviewers. Their cooperation is essential to guarantee good and effective evaluation, as well as positive outcomes to society in terms of knowledge advance, effective technologies, and new products for a very diverse range of users (e.g., medical treatment for patients, new recipes for policy makers etc.). The fact that this peer interaction is regulated by anonymous, distributed and decentralized mechanisms guarantees that science pay-offs are endogenously determined. Research guidelines, priorities and methods are experimentally defined, in that they follow trial and error learning mechanisms involving the whole scientific community and not centralized exogenous planning decision. In this, there is a close analogy between the logics of market and that of science.

Together with some colleagues, I recently conducted some laboratory experiments designed to frame this triadic interaction as a cooperation problem under uncertainty and trust situations, where conflicting interests, cheating and moral hazard are possible. We started from the so-called "investment game," a common experimental framework used by psychologists, economists, and sociologists, particularly suitable to frame certain important social mechanisms that explain how cooperation emerges among rational agents under uncertainty and trust situations [e.g. Berg, Dickhaut, and McCabe 1995; Ortmann, Fitzgerald, and Boeing 2000].

Figure 1 shows a typical investment game [Barrera 2008]. This game models a typical social interaction where *ego* was called to trust *alter* and *alter* to honour trust. The rules of the game are simple. Each round, participants are coupled and randomly assigned to two different roles, called player A (the investor) and player B (the trustee). Both players receive an initial endowment, E_1 and E_2 expressed in ECU (experimental currency unit) with a fixed exchange rate in real money. First, player A (the investor) has to decide whether to send all, some, or none of his/her endowment to player B (the trustee), keeping for him/herself the rest, if any. The amount sent by A, denoted M_1, is multiplied by a factor m by the experimenter (in most cases, m = 3) and sent to the trustee, in addition to the endowment. The parameter m should be interpreted as the returns the player B made due to the investment of player A. Then, player B decides whether to return to A all, some, or none of the amount received. The amount returned by player B (the trustee), denoted K_2, is not multiplied. Each round ends with the payoffs respectively communicated to the players. Each round, the payoff earned by player A (V_1) is:

$$V_1 = E_1 M_1 + K_2$$

whereas the payoff earned by player B (V_2) is:

$$V_2 = E_1 + mM_1 K_2.$$

indicates the extent to which the investor trusted the trustee and is a kind of measure of the degree of trust. K_2 represents the extent to which the trustee is trustworthy and is a kind of measure of the degree of trustworthiness [*ibidem*, 10-11]. At the end of the game, the final payoff of each player is the sum of the payoffs of the N rounds.

The game is called the "investment game" since the rule of multiplying the amount sent by the investor implies that *a*) the investor deals with the uncertainty of paying a cost at the beginning of the interaction to possibly gain higher revenues at the end, and *b*) the trustee has a return from the investor's decision.

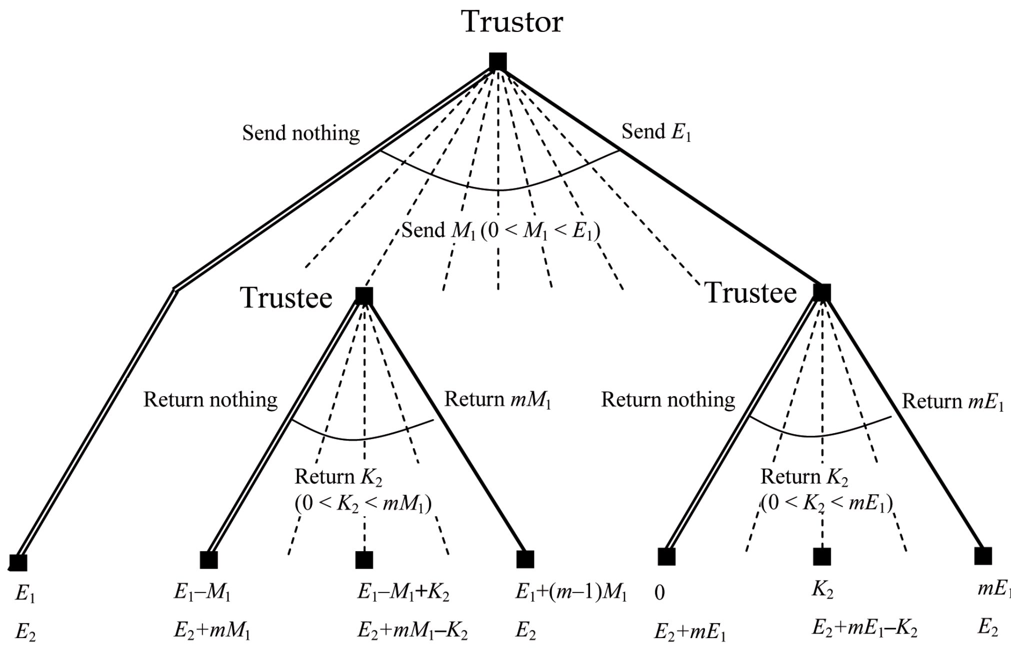

FIGURE 1. The Investment Game.
Source: Barrera [2008, 9].

Following Keser's example [2003], we modified this famous game to include evaluation by third-parties (i.e., reviewers), by modelling a situation similar to the process of submissions' evaluation in scientific journals. We called it "Third-Party Investment Game" and we tested it with a population of 126 students recruited through public announcement in various faculties of the University of Brescia [Boero *et al.* 2009b]. All game interactions took place through a computer network and the subjects were unable to identify their counterparts. The subjects played 25 periods in all and were informed in advance of the duration of the game and the exchange rate of the experimental currency units (ECU) , with 1 ECU = 1.5 Euro cents. The game took nearly one hour (including instructions) and the average earning was 15 Euros, which was paid immediately after the experiment.

The rules of the game were as follows: 1) subjects were randomly assigned as *A*, *B,* and *C* players each period (they played on average the same number of period in each role at the end of the game); *A* was the investor, *B* the trustee, and *C* the evaluator (third party); 2) player *A* (the investor) received an endowment of 10 experimental currency units (ECU); 3) *A* decided how much to send to *B* (the trustee) between 0 and 10; 4) *B* received the amount sent by *A* tripled plus an endowment of 10 experimental currency units (ECU); 5) *B* decided how much to return to *A*; 6) the sums earned by both players in the current period were displayed to both subjects; 7) *C* observed the amount exchanged between *A* and *B* and assigned a reputation score to *B* ("neutral", "positive", or "negative"). This score was made available between step 2 and 3 to the next *A* player who interacted with the *B* player who had been rated. In case of no rating available, e.g., when players had not yet played in a *B*'s role, as in the initial periods, the player appeared to the counterpart as "unknown".

It will not pass unnoticed that this interaction is similar to what happens between editors, authors, and reviewers in science. Editors, as investors, should deal with information asymmetries in respect to authors' submissions (and to reviewers reliability and competence) and with the uncertainty of providing room for unworthy articles that might negatively influence the reputation of their journals, whereas loosing good ones. Authors should evaluate whether to cooperate with editors by guaranteeing that their submissions contain true data, have been released after serious research investment and so on, while they are tempted to cheat by reducing the unit cost per submission, e.g., by adopting economies of scales on research findings. Evaluators should guarantee editors' investment by evaluating authors' submissions, in absence of monetary or reputational incentives and competing at the same time for submissions with authors against editors. Although our experimental game mirrors an abstract interaction that misses out many behavioural motives and social constraints that can have an influence in the empirical reality, these analogies convinced me that our "third-party investment game" embodies at least certain essential aspects of peer review as a social institution[3].

The result of our experiment was that by introducing third-party evaluation, cooperation dramatically increased [*ibidem*]. This confirmed the original findings of Keser [2003]. Her results showed that the effect was stronger for the trustees (+41.5% of returns in respect to the baseline treatment, without any evaluation) than for the investors (+31.5%). Our results show that third-party evaluation provides even more room for reputational investment strategies by trustees. In addition to this, we introduced evaluation on both sides, both on investors and trustees and both before and after the investment decision [Boero *et al.* 2009a]. Even when cheating did not create any reputational sanction, the evaluation of trustees being known to

[3] Of course, my position starts from the assumption that modelling and abstraction are quintessential for scientific investigation. For a plea for the experimental method in sociology, see Boero *et al.* [2009a].

investors after the investment decision, the fact of being under evaluation by others tended to increase the reliability of trustees and thereby the investors' risk attitude. These results indicate that a significant part of trustees' behaviour is not explicable simply as rational investment in reputation. More specifically, besides the obviously strong effect of rational reputation seeking, we found that third-party evaluation also matters when the observed behaviour can only be based on cognitive mechanisms akin to the ones that become apparent in the famous "eyespot" experiments.

Indeed, Haley and Fessler [2005] found that the presence of stylized eyespots on the computer desktop used for experimental sessions significantly increases the generosity of players in a dictator game despite no differences in actual anonymity. In another work, conducted in a real-world setting, Bateson, Nettle, and Roberts [2006] found a similar effect of apparently unimportant cues of being watched. Their results show that people put nearly three times as much money in a "honesty box," used to collect money for drinks in a university coffee room, when the cost of the drinks was displayed on a board along with a picture of eyes staring at the consumer than when the notice included a flower control picture. Of course, the effects of those mechanisms are less uniformly distributed among subjects, but are still noticeable at the aggregate level.

The point is that these experiments show that we are sensitive to other people's judgment, by believing that other opinions can have dramatic consequences in the future even when these are largely unpredictable. This is the reason why peer evaluation, even when anonymous as in our experiment and in science, motivates cooperation beyond rational expectations. But, this is particularly true when the results of the evaluation and thereby personal reputation, circulate within the social system. The condition is that interaction triggers some information about subjects' past behaviour. If this is true, the point is that evaluation in science, as presently performed, triggers only good reputation of contributors (e.g., publications, price announcements), whereas reviewers and their evaluation are covered by secrecy. The fact that reviews are not published, nor are they shared among journals, not only increases the cost of reviewing at the system level (e.g., multiple reviews of the same submission), but makes sanctions weaker against unfair reviewers and does not make reviewing a reputational credit for good reviewers [e.g., Alberts, Hanson, and Kelner 2008].

If this is true, the next step is to tackle the question: could monetary incentives reinforce the strength of the evaluation and thereby cooperation between the three figures involved? To answer this question, we designed a new experiment similar to the former that introduced monetary incentives to the evaluators, tested on a population of 108 students.

We introduced three new treatments to the previous game, where monetary incentives were added to evaluators: *a*) fixed monetary incentives; *b*) incentives aligned to investors' interests, and *c*) incentives aligned to trustees' interest [for details, see Boero, Bravo, and Squazzoni 2010]. It is worth noting that, in treatment *b*), our "third-party investment game" created a typical principal-agent model. In this case, the rational choice literature dictates that monetary incentives are pivotal to motivating trustees to act on investors' behalf, by guaranteeing that the self-interest of the former coincides with the objectives of the latter. This is the situation that models peer review supported by material incentives to the reviewers, as suggested by Hauser and Fehr [2007] among others.

Contrary to any rational choice prediction, monetary incentives caused less cooperation than before (i.e., when there were no incentives for evaluators), no matter whether fixed or aligned and no matter who they were aligned to. The fact that material interests can undermine moral motives has been found in a number of experiments [e.g., Gintis *et al.* 2005; Bowles 2008]. A recent theory called "motivation crowding theory" has been elaborated that accounts for a broad range of empirical phenomena where external interventions, such as monetary incentives or fines, might undermine intrinsic pro-social motivation and lower cooperation [e.g., Frey and Jegen 2001]. In our case, the reason why monetary incentives backfired is that the credibility of the evaluation by third-parties was damaged in the authors' eyes, since evaluation was perceived as subjected to material interests. This might warn those who suggest offering material incentives for reviewers.

If we take the findings of experimental literature on cooperation seriously, it is evident that the present situation of evaluation in science does not exploit all the strengths of social sanctions as in other social situations where cooperation and collective action are pivotal, e.g., economic markets, neighbourhood or community life and friendship networks. The fact that science is a collective endeavour and that scientists know it very well, should not prevent us from trying to strengthen the exploitation of social sanctions that we know reinforce cooperation.

To sum up, these experimental results allow us to understand important aspects of the peer review puzzle. Firstly, the fact of being under evaluation by others dramatically increases cooperation among investors and trustees. This helps us to confirm that peer review is fundamental to guarantee cooperation in an interaction under uncertainty and trust situations like a journal submissions' or research proposals' evaluation. Secondly, evaluation is effective when reputational sanction for trustees is at its most. It is the fact that information about trustees'/authors' cheating strategies can circulate in the system that makes it costly for them to cheat investors/editors. It is reasonable to argue that the same is true for evaluators/reviewers. This is a crucial point for peer review in science, since, social sanctions are poorly established: reviewers or authors cheating editors encounter insignificant sanctions, because of the anonymity and relatively secrecy of reviewers' reports or authors' submissions. For instance, at present, unfair authors can rove from journals to journals, exploiting reviewers' and editors' time, thereby causing a significant cost to the system, since they will probably succeed in being published somewhere.

In my view, any proposal to reform these aspects, e.g., by publishing reviewers' reports together with publications, periodically reporting the list of rejected submissions, giving the chance to authors to rate reviewers and publishing results and so on, favoured after all by electronic journal version platforms, is a necessary step toward the improvement of peer review. This would make reviewing a more appealing activity as well, thereby strengthening reputational incentives around it.

Thirdly, of course evaluation guarantees cooperation not only for reputational investment rational strategies of trustees, but also because third-party evaluation naturally triggers a close reference to moral norms, such as "reciprocating trust" and "avoiding others' disapproval." But, and this is extremely important for peer review, the experimental literature on cooperation clearly shows that this is true because social sanctions are at their most stringent level [e.g., Gintis *et al.* 2005; Bowles 2008]. As I have said before, this is the weak point of peer review at present.

The point is that peer review is a fragile social mechanism undermined by increasing social pressures and expectations. This understanding has been further confirmed by a recent and noticeable agent-based simulation model [Thurner and Hanel 2010]. These authors have developed a model that clearly shows that peer review is a social mechanism which is extremely sensitive to perturbations caused by behaviour heterogeneity. In their model, it is assumed a population of *N* scientists playing two roles, i.e., reviewers and authors, against *N* editors. They produce papers that vary in quality according to a Gaussian distribution and can be selected as reviewers by editors to recommend high quality submissions, with given requirements. Each reviewer assigns a binary recommendation (accept or reject) according to five types of behaviour: the correct (accepting good and rejecting bad submissions), the stupid (random recommendations), the rational (not accepting submissions with a quality higher than his/her own), the altruist (accepts all submissions), and the misanthropist (rejects all submissions). The simulation results show that the simple presence of 10% of rational agents in the reviewers' population drastically lowers the quality of publications, whereas the number of altruists and misanthropists only affect the total number of accepted papers but not the quality of the evaluation process.

If mutually favouring friendship networks are assumed, i.e., a nepotistic networks of scientists who tend to accept papers of co-authors, as a selection mechanism (i.e., scientists are likely to support similar types of research), with rational reviewers, the increase of quality requirements by the editors lead to an adverse effect at the system level. While correct reviewers become more selective, outperforming good quality papers, the rational ones gain more influence by increasing the chance that bad submissions get published. The authors' provocative conclusion is that "under these circumstances – which are not totally unrealistic for certain communities – a purely random refereeing system would perform equally well, and would at the same time save millions of man-hours spent on refereeing every year" [ibidem, 4]. However, this is true simply because social sanctions for unfair behaviour are not present.

Concluding Remarks

Although anonymous, distributed and decentralized peer review is an expensive mechanism as it is a large, continuous and imperfect "social experiment" that involves the whole scientific community. The undisputable evidence of scientific progress over the last decades in many scientific fields, particularly those where peer review is involved at most, should make us conclude that science essentially works. My understanding is that this is true exactly because everyone involved in the evaluation process do not behave merely as rational self-interest agents, but largely as moral norm abiders. Unlike other social situations, where collective action is difficult to achieve, scientists know very well that they are taking part in a collective endeavour.

This said, the point is that the present and future landscape of science, with its increasing figures and pressures on peer review's exploitation, makes it difficult that moral foundations of this important social mechanism can be preserved, if some reform is not undertaken [e.g., Alberts, Hanson, and Kelner 2008].

It is not a case that recently the most interesting attempts to reform peer review in journals have revolved exactly around this point. For instance, the Neuroscience Peer Review Consortium has been established in 2008 between 37 journals that agreed to share reviewers' reports under authors' request. This was to speed the publication and reduce the burden on reviewers, but the sharing of reviews was also expected to motivate reviewers to take their task more seriously.[4] *The British Medical Journal* asked reviewers to sign their reports, so as to increase their sense of responsibility. Following the idea of making peer review more transparent, *The European Molecular Biology Organization* journal decided to publish all the anonymous reviewer reports of the published articles [Pulverer 2010].

[4] See: http://nprc.incf.org/.

These attempts to regulate journal submissions' evaluation in a more precise way and to make it more transparent is expected to stimulate reviewers to work at best. But, as Bernd Pulverer, the editor of The European Molecular Biology Organization recently noted, these examples do not nurture and fortified peer review as much as needed:

> Most successful scientists spend a good fraction of their time reviewing papers. Yet, there is little tangible individual credit derived from the anonymous and voluntary contribution to this cornerstone of the research system. Thankfully, the remarkable culture of willingness to help colleagues and journals through peer review remains healthy, despite ever-increasing publication rates. Nevertheless, we are keenly pursuing means to allow funding agencies and tenure committees to take this essential activity into account [*ibidem*, 31].

My purpose with this contribution was to suggest that reforms should follow the abundant scientific knowledge that already exists on cooperation problems and that behavioural and social sciences are definitively called to take part in this. For this reason, it is sad to note that this topic is more investigated and disputed in the natural rather than the social sciences. The dominant use of experimental knowledge in natural sciences makes evidence replicability by peers a pillar of scientific progress, so robust social evaluation are found and reinforced in the community. However, this cannot prevent sociologists from seeing interesting social aspects in the process of peer review and science evaluation, that could be investigated not only through qualitative, field and "behind the scene" observations, as well as through experimental approaches and more abstract theoretical models. In my view, this is a field where sociological investigations could made a real difference, also by integrating contributions from diverse disciplines, such as behavioural sciences, psychology and economics.

To return to Lamont's book, one of the most interesting parts is where she discriminates between the evaluation modality and technologies in the US and Europe. Her argument is that anonymous, distributed and decentralized peer review was consubstantial to the history of the United States, since a large, decentralised and differentiated social system implied the emergence of socially shared norms that regulated impersonal interactions. She suggests that "peer review works in a largely anonymous system where trust and tight social control through impersonal contacts cannot be assumed [...]. These aspects do not apply to most European countries" [Lamont 2009, 244].

Of course this was especially true until some decades ago, before an EU market for science was created and the globalisation of research entered the picture also at the EU research level. But now, we can say that this difference is less important. Also in Europe, the debate about evaluation and research productivity revolves around the same concepts as the US system and the emergence of an EU market weakened the strength of national academic communities, in particular for research fund competition.

Therefore, anonymous, distributed and decentralized peer review is expected to play an increasingly crucial role everywhere as: *a*) research is becoming more and more internationalized and competition among national and local scientific institutions (to attract human capital and funds) is expected to increase; *b*) the market for research and academic positions is becoming global, particularly in many EU countries where national budget restrictions for science urge scientists to play at a global level; and *c*) funding agencies and academic institutions are increasingly attracted toward research evaluation, metrics and productivity, so that these last in the future will be more important than the in-group local seniors' opinion for careers. This changing landscape implies that the influence of particularistic local cliques and the national self-reference of academic communities in Europe is expected to dramatically diminish in the future, luckily so in my view. On this point, a study that systematically compares the US system of evaluation and Europe, as well as the features of their academic communities, would be really welcomed.

I would like to thank Anna Carola Freschi for remarks on a draft version of the paper and Robert Coates for the linguistic revision of the text.

References

Abramo, G., D'Angelo, C. A., and Caprasecca, A.

2009 "Allocative Efficiency in Public Research Funding: Can Bibliometrics Help?." *Research Policy* 38: 206-215.

Alberts, B., Hanson, B., and Kelner, K. L.

2008 "Reviewing Peer Review." *Science* 321: 15.

Baccini, A.

2010 *Valutare la ricerca scientifica. Uso e abuso degli indicatori bibliometrici*. Bologna: Il Mulino.

Barrera, D.

2008 "The Social Mechanisms of Trust." *Sociologica* 2, doi:10.2383/27728

Barrera, D., and Buskens, V.

2009 "Third-Party Effects." Pp. 37-72 in *eTrust: forming relationships in the online world*, edited by K. Cook, C. Snijders, C., V. Buskens, and C., Cheshire. New York: Russell Sage Foundation.

Bateson, M., Nettle, D., and Roberts, G.

2006 "Cues of Being Watched Enhance Cooperation in a Real-World Setting." *Biology Letters* 2: 412-414.

Benda, W.G.G., and Engels, T.C.E.

2010 "The Predictive Validity of Peer Review: A Selective Review of the Judgmental Forecasting Qualities of Peers, and Implications for Innovation in Science." *International Journal of Forecasting* 27, 1: 166-182.

Berg, J., Dickhaut, J., and McCabe, K.

1995 "Trust, Reciprocity and Social History." *Games and Economic Behavior* 10: 122-142.

Björk, B.-C., Roos, A., and Lauri, M.

2009 "Scientific Journal Publishing: Yearly Volume and Open Access Availability." *Information Research* 14, 1. http://informationr.net/ir/14-1/paper391.html

Boero, R., Bravo, G., Castellani, M., Laganà, F., and Squazzoni, F.

2009a "Pillars of Trust: An Experimental Study on Reputation and Its Effects." *Sociological Research Online* 14, 5. doi:10.5153/sro.2042.

Boero, R., Bravo, G., Castellani, M., and Squazzoni, F.

2009b "Reputational Cues in Repeated Trust Games." *Journal of Socio-Economics* 38, 8: 871-877.

Boero, R., Bravo, G., and Squazzoni, F.

2010 "Can Material Incentives Backfire? An Experimental Study on Reputation." *GECS Working Paper 2010-02*. http://www.eco.unibs.it/gecs/GECS2010IncentivesGB.pdf

Bornmann, L., and Daniel, H.-D.

2005 "Selection of Research Fellowship Recipients by Committee Peer Review. Reliability, Fairness and Predictive Validity of Board of Trustees' Decisions." *Scientometrics* 63, 2: 297-320.

Bosetti, F., and Toscano, C.D.

2008 "Is It Time to Standardize Ethics Guiding the Peer Review Process?." *Lipids* 43: 107-108.

Bowles, S.

2008 "Policies Designed for Self-Interest Citizens May Undermine 'The Moral Sentiments': Evidence from Economic Experiments." *Science* 320: 1605-1609.

Cho, M. K., McGee, G., and Magnus, D.

2006 "Lessons from the Stem Cell Scandals." *Science* 311: 614-615.

Chubin, D.R., and Hackett, E.J.

1990 *Peerless Science. Peer Review and US Science Policy*. New York: State University of New York Press.

Coates, R., Sturgeon, B., Bohannan, J., and Pasini, E.

2002 "Language and Publication in Cardiovascular Research Articles." *Cardiovascular Research* 53: 279-285.

Couzin, J.

2006 "... And How the Problems Eluded Peer Reviewers and Editors." *Science* 311: 614-615.

de Carvalho, L.B.

2006 "Pressure Also Leeds to Worthless Publications." *Nature* 439, 16: 784.

Dost, F.N.
2008 "Peer Review at the Crossroads – A Case Study." *Environmental Science and Pollution Research* 15: 443-447.

European Commission
2006 *Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe*. Brussels: European Commission. http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf

Frey, B., and Jegen, R.
2001 "Motivation Crowding Theory." *Journal of Economic Survey* 15: 589-611.

Gintis, H., Bowles, S., Boyd, R., and Fehr, E. (eds.)
2005 *Moral Sentiments and Material Interests. The Foundations of Cooperation in Economic Life*. Cambridge, MA: The MIT Press.

Goldsworthy, J.
2009 "Research Grant Mania." *Australian Universities Review* 50, 2: 17-24.

Gordon, R., and Poulin, B.J.
2009 "Cost of the NSERC Science Grant Peer Review System Exceeds the Cost of Giving every Qualified Researcher a Baseline Grant." *Accountability in Research: Policies and Quality Assurance* 16, 1: 13-40.

Grainger, D.W.
2007 "Peer Review as Professional Responsibility: A Quality Control System Only as Good as the Participants." *Biomaterials* 28: 5199-5203.

Gura, T.
2002 "Scientific Publishing: Peer Review Unmasked." *Nature* 416: 258-260.

Haley, K.J., and Fessler, D.M.
2005 "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game." *Evolution and Human Behavior* 26: 245-256.

Hauser, M., and Fehr, E.
2007 "An Incentive Solution to the Peer Review Problem." *PLOS Biology* 5: e107. doi_10.1371/journal.pbio.0050107

Henderson, A.
2001 "Undermining Peer Review?." *Society* January/February: 47-54.

Horrobin, D.F.
2001 "Something Rotten at the Core of Science?." *Trends in Pharmacological Sciences* 22, 2: 51-52.

Kerr, N.L., and Tindale, R.S.
2004 "Group Performance and Decision Making." *Annual Review of Psychology* 55: 623-655.

Keser, C.
2003 "Experimental Games for the Design of Reputational Management Systems." *IBM Systems Journal* 42: 498-506.

Jayasinghe, U.W., Marsh, H.W., and Bond, N.
2006 "A New Reader Trial Approach to Peer Review in Funding Research Grants. An Australian Experiment." *Scientometrics* 69, 3: 591-606.

Lamont, M.

2009 *How Professors Think: Inside the Curious World of Academic Judgement*. Cambridge, MA: Harvard University Press.

Link, A., Swann, C., and Bozeman, B.

2008 "A Time Allocation Study of University Faculty." *Economics of Education Review* 46, 4: 363-374.

Mabe, M., and Amin, M.

2001 "Growth Dynamics of Scholarly and Scientific Journals." *Scientometrics* 51, 1: 147-162.

Mandviwalla, M., Patnayakuni, R., and Schuff, D.

2008 "Improving the Peer Review Process with Information Technology." *Decision Support Systems* 46: 29-40.

McCabe, M.J.

2002 "Journal Pricing and Mergers: A Portfolio Approach." *The American Economic Review* 92, 1: 259-269.

Marris, E.

2006 "Should Journals Police Scientific Fraud." *Nature* 439, 2: 520-521.

Mayo, N.E., Brophy, J., Goldberg, M.S., Klein, M.B., Miller, S., Platt, R.W., and Ritchie, J.

2006 "Peering at Peer Review Revealed High Degree of Chance Associated with Funding of Grant Application." *Journal of Clinical Epidemiology* 59: 842-848.

Merton, R.K.

1973 *The Sociology of Science. Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.

Nature

2006 "Peer Review and Fraud." *Nature* 444: 971-972.

Obrecht, M., Tibelius, K., and D'Aloiso, G.

2007 "Examining the Value Added by Committee Discussion in the Review of Applications for Research Award." *Research Evaluation* 16: 79-91.

Opthof, T., Coronel, R., and Janse, M.J.

2002 "The Significance of the Peer Review Process Against the Background of Bias: Priority Ratings of Reviewers and Editors and the Prediction of Citation, the Role of Geographical Bias." *Cardiovascular Research* 56, 3: 339-346.

Ortmann, A., Fitzgerald, J., and Boeing, C.

2000 "Trust, Reciprocity, and Social History: A Re-examination." *Experimental Economics* 3: 81-100.

Pulverer, B.

2010 "Transparency Showcases Strength of Peer Review." *Nature* 468: 29-31.

Rothwell, P.M., and Martyn, C.N.

2000 "Reproducibility of Peer Review in Clinical Neuroscience. Is Agreement Between Reviewers Any Greater than Would Be Expected by Change Alone?." *Brain* 123: 1964-1969.

Schaffer, A.

2009 "America's Got Science Talent: The Biomedical Research Community Goes Bananas for $200 Million in Stimulus Funding." *Slate*, April 29. http://www.slate.com/id/2217229

Smith, R.

1997 "Peer Review: Reform or Revolution? Time to Open the Black Box of Peer Review." *British Medicine Journal* 315: 759-760.

2006 "Peer Review. A Flawed Process at the Heart of Science and Journals." *Journal of the Royal Society of Medicine* 99: 759-760.

Spier, R.E.

2002a "The History of Peer Review." *Trends in Biotechnology* 20, 8: 357-358.

2002b "Peer Review and Innovation." *Science and Engineering Ethics* 8: 99-108.

Thurner, S., and Hanel, R.

2010 "Peer-Review in a World with Rational Scientists: Toward Selection of the Average." *arXiv:1008.4324v1*. http://arxiv.org/PS_cache/arxiv/pdf/1008/1008.4324v1.pdf

Travis, G.D.L., and Collins, H.M.

1991 "New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System." *Science, Technology and Human Values* 16, 3: 322-341.

Ware M.

2007 *Peer Review in Scholarly Journals: Perspective of the Scholarly Community – An International Study*. Bristol: Mark Ware Consulting. http://www.publishingresearch.net/documents/PeerReviewFullPRCReport-final.pdf

Peering Into Peer Review

Abstract: This article examines the social mechanisms behind peer review and provides a complementary view to Michèle Lamont's book, *How Professors Think*. It emphasizes the need for outlook review of the entire process of evaluation in science in more general terms and suggests the added value of modelling to investigate it. It introduces experimental findings on the relevance of social sanctions and the counter-productive effect of economic incentives on peer review that can support the recent debate about its reform. It illustrates the relevance of reputational incentives to guarantee cooperation between the different figures involved in the evaluation process.

Keywords: Peer review, third-party evaluation, laboratory experiment, investment game, Michèle Lamont.

Flaminio Squazzoni is Assistant Professor of Economic Sociology at the Department of Social Sciences, University of Brescia, where he heads the GECS-Research Group on Experimental and Computational Sociology. He is review editor of *JASSS-Journal of Artificial Societies and Social Simulation* and has been chair of the Fifth European Social Simulation Association Conference (Brescia, 2008) and director of the First ESSA Summer School in Social Simulation (Brescia, 2010). His research interest is social simulation. In particular, he is doing research on the modelling and understanding of social norms and institutions in economic systems, by combining experimental and simulation methods. Currently, he is working on a monograph on "Agent-Based Computational Sociology" (Wiley Blackwell, UK), forthcoming in 2011. Email: squazzon@eco.unibs.it.