



HAL
open science

From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review

David Pontille, Didier Torny

► **To cite this version:**

David Pontille, Didier Torny. From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review. *Human Studies*, 2015, 38 (1), pp.57-79. 10.1007/s10746-014-9335-z . hal-01143310

HAL Id: hal-01143310

<https://hal.science/hal-01143310>

Submitted on 17 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review

David Pontille

Centre de Sociologie de l'Innovation
CNRS (UMR 7185) - Mines-ParisTech
60 boulevard Saint Michel
75006 Paris
phone: +33 (0)1 40 51 91 93
fax: +33 (0)1 43 54 56 28
david.pontille@mines-paristech.fr

&

Didier Torny

Risques, Travail, Marchés, État
RiTME - INRA (UR 1323)
65 Boulevard de Brandebourg
94205 Ivry-sur-Seine
didier.torny@ivry.inra.fr

Human Studies
2015, vol. 38(1), p. 57-79
(online first: September 23, 2014)

Abstract

Born in the seventeenth century, journal peer review is an extremely diverse technology, constantly torn between two often incompatible goals: the validation of manuscripts conceived as a collective industrial-like reproducible process performed to assert scientific statements, and the dissemination of articles considered as a means to spur scientific discussion, raising controversies and civically challenging a state of knowledge. Such a situation is particularly conducive to clarifying the processes of valuation and evaluation in journal peer review. In this article, such processes are considered as specific *tests* in order to emphasize the uncertain properties of pre-tests manuscripts. On the one hand, *evaluation tests* are examined at the core of the validation of manuscripts, such as defining the coordination of judging instances (editor-in-chief, editorial committee, outside reviewers) or controlling the modalities of inter-knowledge between reviewers and authors. They are also studied regarding the dissemination of articles, notably through the contemporary conception of a continuing evaluation test termed “post publication peer review”. On the other hand, *valuation tests* are both part of the validation of manuscripts, such as the weighting of different judgments of the same manuscript and the tensions that these hierarchies cause, and of the dissemination of articles, such as attention metrics recording the uses of articles. The conclusion sketches out how the articulation of these different tests has recently empowered readers as a new key judging instance for dissemination and for validation, potentially transforming the definition of peers, and thus the whole process of journal peer review.

Keywords

Anonymity; Academic journals; Evaluation; Peer Review; Valuation studies

Practices of evaluation have become increasingly salient in many professional circles, be it in the corporate world, in administration and services, in the police or in hospitals (Bruno and Didier 2013). Of course, scientific research does not escape this broad phenomenon. Against a background of international competition, the systematic assessment of universities, the rationalization of research organizations, or improvements of performance and productivity, scientists have been the object of increasing attention from governments and public agencies in numerous countries (Auranen and Nieminen 2010). This infatuation for evaluating academic production has led to an ever-increasing range of tools, such as the classification of journals (Hicks and Wang 2011; Pontille and Torny 2010), bibliometric tools (Bollen *et al* 2009; Cronin and Sugimoto 2014; Pontille and Torny 2013) and adapted forms of peer review (Donovan 2007; Butler and McAllister 2009). Research assessment exercises rely on distinct tools that ascribe values to the entities at stake, give them weighting, ascertain their qualities, and sort them out.

Yet the academic world is not just the object or “target” of these new evaluation practices, often considered to simply be the direct result of new public management. More fundamentally, scientific activity is intimately linked to the regular production of evaluation tools which are to varying extents either renewed or completely new (Wouters and Costas 2012). From this standpoint it constitutes a truly strategic site for analyzing the forms of evaluation which are developed, experimented with and broadly discussed. Of these forms, peer review is the alpha and omega of scientific production, at least as far as the allocation of funding (Lamont 2009) and the publication of articles are concerned (Porter 1964; Zuckerman and Merton 1971). Nevertheless, empirical studies systematically point out not only the broad diversity in concrete practices, but they also engender debates on the nature of the evaluation process, how it unfolds and its limits. And recently, it was even considered to be unnatural or obsolete (Smith 2006).

Such a situation is particularly conducive to clarifying the process of valuation and evaluation in journal peer review. There is no shared conception of valuation and evaluation, as current scholarly works on such processes rely on a very broad and inclusive definition, fostering investigations in different areas (Lamont 2012; Helgesson and Muniesa 2013). In this paper, valuation and evaluation will be considered as specific *tests* — as this notion is defined in the French pragmatic sociology (Latour 1987; Boltanski and Thévenot 2006) to emphasize the uncertain properties of any kind of entity before a test and its potential transformation during it. While we assume evaluation tests are assessment procedures dedicated to arbitrate the quality of an entity, valuation tests are aimed at the commensurability of several assessments, and sometimes at determining their aggregate value.

Who exactly assesses manuscripts? What are the actual conditions under which evaluation is performed? What precisely do the acts of valuation relate to? To answer these questions, we will look at manuscript evaluation as a specific “technology”, following Shapin and Schaffer (1985) who showed that the experimental practice took shape in the 17th century,

based on three technologies which were intimately linked in the production of knowledge¹. But instead of considering manuscript evaluation as a technology which is set in stone, we will examine several specific arrangements which have varied depending on the era, discipline and journal, and which have given rise to the main historical and contemporary criticisms. We will pay particular attention to the dynamic of this technology which is founded on two distinct “orders of worth” (Boltanski and Thévenot 2006) which are constantly put forward by promoters of journal peer review. On the one hand, the validation of manuscripts is described as a collective industrial-like reproducible process performed to assert scientific statements. On the other hand, the dissemination of articles is considered as a means to spur scientific discussion, raising controversies and civically challenging a state of knowledge.

These two worths are often incompatible. For example, the sharing of new results with audiences far removed from the scientific collectives that produced them is considered as sufficiently problematic by Franz J. Ingelfinger (1969), chief editor of the *New England Journal of Medicine*, who systematically chose to refuse to publish articles presenting results previously exposed elsewhere, notably in the general press. Symmetrically, the processing delays resulting from the validation procedures have been often criticized as unacceptable barriers to the dissemination of knowledge, and have led numerous actors to organize the circulation of working papers from the 1990s onwards. We will see that this discordancy is resolved in concrete technologies of journal peer review which depend on particular evaluation and valuation tests.

This paper will focus on the tensions embodied within these technologies and on the conflicting justifications concerning the validation of manuscripts and the dissemination of articles. Firstly, we will show that most of the justifications relating to validation are targeted towards evaluation tests, such as defining the best arrangements to coordinate judging instances and ascribing their role in the evaluation process. We will then focus on the control of modalities of inter-knowledge between reviewers and authors, especially those designed to identify the latter, or, on the contrary, to hide their respective identities. Furthermore, we will deal with other justifications, which targeted valuation tests such as the connections between the different judgments of the same manuscript, how they are weighted and the tensions that these hierarchies cause.

Regarding dissemination, we will return to the old tradition of commentary and dialogue within journals, which has been supported by the conception of a continuing evaluation test even after publication, and examine its contemporary off-spring through “post-publication peer review”. We will then revisit another tradition initially based on citation, which aims to measure the uses of articles, and which has quite recently led to article valuation tests. We will conclude on how the development of these traditions has empowered readers as key

¹ “The establishment of matters of fact in Boyle’s experimental programme utilized three technologies: a material technology embedded in the construction and operation of the air-pump ; a literary technology by means of which the phenomena produced by the pump were made known to those who were not direct witnesses ; and a social technology that incorporated the conventions of experimental philosophers should use in dealing with each other and considering knowledge-claims” (Shapin and Schaffer 1985, p. 25).

figures for dissemination and for validation, potentially transforming the definition of peers, and thus the whole process of journal peer review.

Our analysis is based on three types of document. Firstly, historical, sociological and scientific documents which directly relate to peer review in journals and which examine its multiple dimensions. Using these works of synthesis, we identified the most characteristic (sometimes quantified) aspects, as well as the less well-known or even neglected points. We then selected articles which deal with manuscript evaluation, based on modeling, experiments, studies of practices, normative propositions or case studies based on a given journal's archives. Our selection focused on the most innovative propositions at conceptual (invention of denominations or categories) and operational (the creation of unique arrangements) levels. Finally, we built a corpus of primary sources by performing a systematic search for certain key words ("*peer review*", "*blind refereeing*", "*author identity*", "*editorial procedures*", etc.) on various on-line platforms: JSTOR, Elsevier, Nature Publishing Group, Sage, Springer, Taylor & Francis, Wiley, Google Scholar. We focused most specifically on marked stances, particularly in "*editorials*", "*letters*" and "*comments*". We carried out a certain number of interviews with editors-in-chief and editorial committee members to complete this documentary corpus.

The Gradual Stabilization of Judging Instances

Several historical works have carefully documented the concrete modalities of experimental practice in the 17th century, particularly at The Royal Society in London (Shapin and Schaffer 1985; Shapiro 2000). They focus on the implementation of specific technologies used to perform experiments, to regulate modalities for participation and to ensure validation of the facts. At the same time, Henry Oldenburg, Secretary of The Royal Society, instigated significant changes in the oral and written dissemination of knowledge. He provided a veritable hub for the sorting of abundant correspondence from scientists dispersed throughout Europe, and in 1665 decided to publish certain letters by founding one of the first scientific journals: *Philosophical Transactions of the Royal Society of London*.

This event led to the creation of a new instance in the scientific world, that of the editor-in-chief. Henry Oldenburg wanted to do more than just publish letters as and when they arrived. He selected those which deserved to be published in the journal and determined the order in which they should be presented (Bazerman 1988, chap. 5). The principle objective of this type of publication was to accelerate communication between scholars by regularly informing its members (unable to attend public demonstrations or not belonging to a scientific community) of new experiments. The purpose of publishing short texts was, far more rapidly than was the tradition, to transmit scholarly works and the state of knowledge to the wider audience that was being built. Following criticism that judgment was being made by just one person, responsibility for the publication of *Philosophical Transactions* was extended to another judging instance as from 1750 – an editorial committee collectively responsible for examining and selecting manuscripts. Text validation was thus placed in the hands of a collective instance, made up of internal assessors chosen for their competencies in a variety of fields. The existence of this new judging instance, combined with that of editor-in-chief,

gradually came to light with the publication of the list of editorial committee members in each issue.

The exponential increase in the number of scientific journals in the 18th century in England, France and above all in Germanic countries (Kronick 1962), often under the auspices of new specialist learned societies, led to a diversification in the way that judging instances were organized. Whilst certain journals had chosen to have just an editor-in-chief, others had opted for the combination of an editor-in-chief and an editorial committee. The latter sometimes involved a hierarchy, with associate editors being placed above other members of the committee. These members were only called upon to give their opinions within their own field(s) of expertise, with the editors retaining the vast majority of the evaluation work, or, on the contrary, limiting themselves to the tasks of harmonization and coordination.

In the 19th century, the increase in the number of journals accelerated and was accompanied by gradual specialization; this once again raised the question of the specific competency of the judges. At first on an informal basis, and then more routinely, certain editors-in-chief called upon experts who did not belong to any editorial committee to give their opinions on the specific issues and specific methodological validity of manuscripts (Burnham 1990). An additional judging instance thus appeared within the manuscript evaluation process – outside reviewers. *The British Medical Journal* (BMJ) would appear to have pioneered this innovation; in the 1870s it made systematic use of this instance, calling upon the reviewers' proven expertise in the field in question.

The only system that seems to me adequate to the real needs of professional readers is that in which every unsigned editorial paragraph is written by a specially selected expert. That is the principle on which I have modeled the journal I have the honor to conduct. Every letter received, every paragraph, every cutting editorially dealt with, is referred to an expert having knowledge and being a recognized authority in the matter... It is a laborious and difficult method, involving heavy daily correspondence and constant vigilance to guard against personal eccentricity or prejudice or – that bugbear of journalism – unjustifiable censure².

At the beginning of the 20th century, there were thus three relatively stable judging instances: editors-in-chief, editorial committees and outside reviewers. However, their visibility varied, especially that of outside reviewers. Indeed, it would seem that authors were not fully aware that such reviewers were being increasingly called upon – editors-in-chief were their sole contacts. Unlike internal judges, for a long time outside reviewers remained the less visible entities within the evaluation process. It was only in the 1930s that they began to be acknowledged in published articles, thus showing that the authors were aware of their existence, and it was not until the 1960s that journals began to annually publish a full list of the experts who had reviewed submitted manuscripts. Some thought this new judging instance should not be part of the evaluation test. The most famous example is Albert Einstein: when he was informed that the manuscript he had submitted to *Physical Review* had been the object of a report by an anonymous outside reviewer, he told the editor-in-chief that he had not authorized its dissemination before validation by him; he then withdrew the article and submitted it to another journal (Kennefick 2005).

² Ernest Hart, editor-in-chief of the BMJ speaking to the *American Association of Medical Editors* in 1893 (cited by Burnham 1990, p. 1325).

The increasing number of instances also led to an extension of the time between manuscript submission and acceptance³ and then between acceptance and publication on the other. This evolution was reinforced by the massive increase in the number of researchers and manuscripts during the 20th century. The tension between the initial objective of rapid circulation of knowledge and the duty of competence of evaluating instances came in full light. In fact, although scientific journals had originally been designed to accelerate the dissemination of articles, the process was now being bogged down by the ever-increasing number of manuscript submissions.

In order to resolve this tension, journals developed three solutions that demonstrate the variety of arrangements between the different judging instances. First of all, a small number of journals asserted the omnicompetence of the editor-in-chief and only rarely called upon outside reviewers. A century after the BMJ began to use outside reviewers, a competitor slammed their systematic use:

[Its editor-in-chief] was opposed to peer review and sent few papers for expert opinion, believing that that was conservative, elitist, and even bigoted. *The Lancet* would expect to reject a paper within a week, and publish an acceptable contribution within four or five weeks. Mistakes were made, but authors had the opportunity of taking their work, without delay, to another journal. (Douglas-Wilson 1974, p. 326)

In a context of competition between journals this rapidity was an advantage, but it was also welcomed by authors who wished to mark the anteriority of their works or disseminate their results.

Secondly, numerous journals have maintained this omnicompetence to reject manuscripts by implementing a stage of preselection by the editor-in-chief. Whilst the first version of this preselection was proposed by Glenn (1976) for sociology journals, it was widely adopted in the field of biomedical sciences: for example, *Nature* received 10,000 manuscripts in 2006, 60% of which were immediately rejected. This preselection phase presupposes a homogeneousness of judgment, with the editor-in-chief also acting on behalf of outside reviewers:

I usually receive one manuscript a day, and I answer to the authors in the following two weeks. So I begin my reading with the abstract, then the bibliography. If I think it is not within the scope of our journal, I drop the paper. Then, I read the introduction and the beginning of the manuscript; as soon as I think that it won't pass review, I stop and reject it. If I get to the end of the paper and I judge that referees will like it, I send it to three outside reviewers. Only 30% of manuscripts pass the cut.⁴

This preselection also implies an asymmetry between judgments, as the outside reviewers only assess articles which have been preselected by the editor-in-chief. It also leads to a transformation in the evaluation process: to the need for validation must be added valuation

³ For example, in 1965 this period was extended from 9 to 17 months for journals published by the *American Psychological Association* (1965).

⁴ Interview with one of the seven associate editors of a medical journal (carried out on the 06.12.2012).

criteria. Usually importance, scope and originality are mentioned, especially as the room for publication grows far more slowly than the volume of submitted manuscripts.

Thirdly, companies such as Rubriq and Peerage of Science have recently proposed to modify the order of judging instances. By explicitly positioning themselves as market intermediaries between authors and journals, they promise to ensure expert reviewing by qualified researchers within a very short time frame. The latter are paid by the company, the cost being borne by the authors or journals who subscribe to the service. With such an arrangement, the opinions of outside reviewers precede submission to the journal, and the manuscript is delivered accompanied by their reports. The journal's evaluation activity then focuses exclusively on the manuscript's destination, based on these subcontracted opinions.

Whereas in the 17th century the relationship between authors and judges was founded on the exchange of correspondence between peers who rubbed shoulders on a very regular basis, the modern situation is marked by a dispatching of the editorial decision-making process. So validation of a submitted manuscript now lies in the hands of various actors via multiple evaluation tests. It therefore raises the question of their respective valuation so as to enable commensurability. Before focusing on this question, we will deal in detail with another problem springing from the multiplicity of judging instances: the issue of their inter-knowledge as a crucial component of the evaluation test.

Anonymity: Benefit or Hindrance to Judgment?⁵

The acquaintanceship between authors and reviewers might be valued as the testimony of a mutual recognition: in her transdisciplinary study, Crane (1967) showed that institutional plurality of succeeding authors strongly depends on the institutional diversity of editors-in-chief. But this acquaintanceship is symmetrically criticized as leading to the establishment of publishing "old boys' clubs" on a relational basis rather than on scientific one. So the relationships between authors and reviewers, and their mutual visibility during the validation process, became controversial from the 1950s on.

Blind reviewing

In sociology journals in the middle of the 20th century, manuscripts were almost exclusively evaluated internally. It was within this framework of practices that the issue of author anonymity emerged. In June 1955, the *American Sociological Review* notified its contributors of a modification to its procedures for reviewing manuscripts:

As an experiment in the evaluation of articles, all papers will now be circulated to the assistant editors and judged without name or institutional identification. It will be helpful if contributors to the Review will attach a cover page giving the title, author's name and institutional affiliation.

The first page of the paper should bear the title as a means of identification, but not name and institution. (ASR 1955, p. 341)

In the 1960s, other American sociology journals made the same recommendations to their authors and in 1967 the American Sociological Association applied the principle to all of its

⁵ An extended version of this part is available in Pontille and Torny (2014).

journals. The anonymization of authors spread to other disciplines related to social sciences and the humanities in the mid-1970s, essentially due to the publication of articles examining the issue of reviewer “bias” (Zuckerman and Merton 1971; Cicchetti and Conn 1976; Mahoney 1977). But this extension was not driven solely by editors-in-chief. It was also a demand from groups of women within American learned societies who pointed out the low acceptance rates of articles written by female scholars (Benedek 1976; Weller, 2001: 225-226). The issue was thus less one of ensuring objective judgments than of repairing injustices in the name of equity between peers.

These mobilizations did not lead to researchers criticizing the principle of blind reviews. Only a few isolated researchers let their discontent be known, writing to the journals to stress the importance of a name as an identifying reference.

We frequently forget that a man’s name is important (whether it is widely known or not). It can identify his biases and perspectives (sources of professional training, previous work, occupational, experiences, etc.) and, therefore, can be used as a basis for judging the reliability and relevance of what he says. (Lowry 1967, p. 220)

So the anonymization of authors was not considered to be a mere technical gesture. It served a radically different conception of the evaluation of scientific manuscripts: judgments must be based purely upon the content of a text and any information about its authors will taint said judgments. The opposing vision stated that in order to come to an opinion, one needs to be able to link writings to writers.

When the issue of author anonymity was discussed from the end of the 1960s to the mid-1970s in the experimental sciences (physics, medicine and biology in particular), it was the latter conception of evaluation which was preferred by the vast majority. Several editors-in-chief underlined the inseparability of authors and their texts at three different levels (Ward and Goudsmit 1967). Firstly, they believed that materially speaking, the total anonymization of a text was complicated by the self-quotations that authors use on a regular basis to refer to their own previous works. Secondly, they said that the credit reviewers gave to the described experiments and results was based to a very large extent on past studies and on the use of specific equipment. Anonymization would thus lead to a cognitive confusion which would hinder evaluation tests. Finally, they pointed out that if no names were given, reviewers would attempt to attribute texts to given authors and would base their judgments on this attribution, whether their assumption was correct or not. For these three (at least) reasons, at that time many journals, especially in the field of biomedicine, preferred not to adopt anonymization.

Secretly evaluating

In the field of experimental sciences, although recourse to outside reviewers had become common practice, there were nevertheless debates on whether or not their identity should be revealed. This situation was criticized for the first time by a biochemist in the *New Scientist*.

On occasion, the act of submission of a paper can place the author at the mercy of the malignant jealousy of an anonymous rival. Manifestations of antipathy can take many forms, which range from contemptuous mockery (...) to outright theft. To have suffered such an experience generally induces sympathy from colleagues, but condolences are not enough.

What is needed is justice; and justice can be secured only by altering the system. (Jones 1974, p. 758-759)

Jones did not say that referees were by their very nature deleterious to the editorial process, but he did consider that the anonymization procedure allowed them to act without any fear of sanction. Jones therefore suggested that reviewers should sign their reports as it was the case with other editorial practices, such as book reviews.

Following Jones' article, other researchers criticized reviewer anonymity. Based on their own experiences of the validation process of manuscripts, they underlined the competition between authors and reviewers which gave the latter the opportunity to write defamatory reviews of the former or, on the contrary, which prevented any verification of the referee's competency. These criticisms stigmatized the way in which the principle of anonymity conflicted with the status given to public debate. As a result, they reconnect with a conception of the scientific exchange inherited and adapted from legal practices based on contradictory discussion (Shapiro 2000).

Referring to works which underlined "reviewer bias", critics focused on the repeated divergences between reviewers of the same manuscript, the lack of care given to their reviews, and their lack of awareness of literature which had already been published, thus preventing them from identifying plagiarism or self-plagiarism. Whilst reviewer anonymity was not the only factor to be mentioned, critics stressed the extent to which it contributed towards such bias, the discovery of which had a more wide-ranging consequence: it undermined confidence in the entire evaluation process. The image of an upright reviewer gave way to one of a colleague steeped in self-interest, prejudice and beliefs which formed an integral part of his/her opinion.

For a long time these critical stances nevertheless remained very marginal compared to the assertion of the need for reviewer anonymity, initially supported by journals. In 1974, *Nature* responded to Jones' article in an editorial defending the principle of anonymity for referees (*Nature* 1974). It put forward three arguments. The journal believed that it would be disastrous if referees could be directly contacted by authors who were unhappy that their texts had been refused for publication, thus turning scientific discussion into open disputes. It also held that anonymity allowed room for strong but well-founded criticism. Finally, anonymous evaluation was deemed to be an initial and vital stage in communication with peers, dealing not only with a manuscript's methodological or technical quality, but also with its readability and with its more general utility.

To the arguments put forward by *Nature*, editors-in-chief and researchers gradually added others, also based on their own experiences of the editorial process (DeBakey 1976; Wilson 1978; Morgan 1984). Referees were described as constructive, as making suggestions to improve manuscripts, and even as recusing themselves prior to any evaluation, because they considered themselves to be too close to the authors, or because they felt they lacked the required competency. Moreover, these supporters of anonymity pointed out that referees only offer opinions, with the final decision on manuscripts depending entirely on the editor-in-chief or the editorial board. They also explained that if the names of referees were revealed, it would become necessary to only use tenure academics, younger scholars having the feeling that their careers were under threat, thus causing them to produce evaluations which

were vapid or overly consensual. Finally, some compared reviewers to electors, thus deeming it important that they vote anonymously.

Pros and cons of anonymity for reviewers have been subsequently traded in diverse disciplines: for instance, in geography, a controversy between editors-in-chief happened in the mid-1990s (Berg 2001). But these debates have then occurred in a new structured and polarized space.

Objectivity versus publicity

Within a context of repeated revelations of scientific fraud (Broad and Wade 1982) and of the regular publication of experiments demonstrating the flaws in manuscript evaluation (Peters and Ceci 1982), whether or not to keep the names of authors and reviewers secret was becoming an increasingly pressing issue for the design of evaluation tests. To deal with the problem of anonymity, categories describing distinct solutions stabilized during the mid-1980s. Journals were thus placed in a position of choice within a space of possibilities described in table 1.

Table 1. Anonymity and identification labels in manuscript peer review

	Reviewers	
Authors	Anonymized	Identified
Anonymized	Double blind	Blind review
Identified	Single blind	Open review

It was during this period that the terms “single blind” and “double blind”, borrowed from the methodologies used to control biases in clinical trials, were adopted and constituted the two most common options. “Blind review” did not extend beyond certain social sciences and humanities and was in practice based on evaluations carried out solely by the members of editorial committees. Developed as an alternative to the single blind as from the early 1980s, “open review” was for a long time limited to just a few journals which encouraged or required their reviewers to sign their reports (Knox 1981).

Certain journals positioned themselves within this space, stressing the flexibility of their evaluation tests: some left the issue of anonymity to the authors, allowing them to designate their reviewers, or to allow reviewers to choose whether or not to remain anonymous or to sign their reports. These various denominations lasted through to the 1990s, to such an extent that they gradually came to be considered as alternative models. From a disciplinary point of view, and as a continuation of the stances described earlier on, different surveys showed that the majority of social science journals preferred the double blind system, whereas biomedical journals (especially the most prestigious among them) maintained a single blind procedure (Weller 2001). However, this division should not be taken to mean that real concerted policies were at work beyond the range of a given learned society or academic publisher. Moreover, the trajectory of certain journals showed that depending on the period, there was a see-saw effect between single blind and double blind. The *American Economic Review* is a perfect example, making four procedural changes between 1973 and 2011.

During the 2000s, the single blind and double blind were thus maintained as alternatives, but open review was developed to such an extent that it is systematically included in works which bear upon these different options (Lee et al. 2013). A study on the review of articles among more than 5 000 authors, reviewers and editorial board members, funded by British commercial publishers, recently examined their practices, preferences and values in this domain (Ware and Monkman 2008). It portrays a divided landscape, with half preferring the double blind, a quarter choosing the single blind and an eighth opting for open review.

The scientific and moral values attached to each option are also contrasted. Open review is valorized for the responsibility it places upon reviewers signing their reports and for the transparency and dialogue that it encourages. Supporters of the single blind focus on the anonymity of reviewers, putting them in a position where they are sheltered from the consequences of their reviews and where the fact that they know the names of the authors allows them to more accurately place their manuscripts in relation to their previous works or to detect conflicts of interest. It can also be a default choice, some researchers feeling that the double blind is not achievable in practice. Supporters of the double blind, on the other hand, praise the objective nature of the procedure and the equity that it guarantees, particularly with regard to lesser known authors.

There are thus at least three distinct forms which still endure, supported by different motives relating to evaluation. Far from a linear vision which sees the single blind as a necessarily unfinished version of the double blind, it is in fact the coexistence of a variety of evaluation tests developed by journals or learned societies which prevails, combined with continuous research into the evaluation of the technologies themselves. Discussions about the anonymity of reviewers and authors demonstrate the contrasting conceptions of evaluation tests: the obligation of public debate on the one hand, the secrecy required for serene judgment on the other.

Judgment Valuation

Anonymity or identification may be conditions under which an instance of judgment would be considered fair, accurate or even valid. Like proficiency in the matter of the manuscript or the absence of conflict of interest, it defines the felicity conditions of an evaluation test. Yet as the number of judging instances potentially grows, it does not determine the ways these different validated judgments are designed, produced and weighted. It is only through the shaping of valuation tests that a “commensuration process” (Espeland and Stevens 1998) occurs. As the literature has regularly demonstrated disagreements between reviewers of the same manuscript (Cole et al. 1981; Cicchetti 1991), this commensuration process is crucial and leads to various tensions which traverse contemporary journals (Baruch et al. 2008). It is institutionalized on four different levels: the objectives set for the validation process, frames given to the content of evaluations, places attributed to reviewers for manuscript (re)writing, hierarchies accorded to the different judging instances.

On each of these levels, further oppositions come to light when analyzing the practices declared or observed in journals (Weller 2001, chap. 3). As Hargens (1988) has shown, the relationship between manuscript flow and available space for publication means that the validation process has distinct objectives: a high rejection level does not just mean a phase

of preselection; it also encourages outside reviewers to be highly selective. Journals with a 90% or 95% rejection rate, which is not uncommon in social sciences and humanities, have to share their views with reviewers by insisting that manuscripts be highly innovative, have the broadest scope or the most important findings. Conversely, a higher acceptance rate, a lack of available manuscripts at a given time or the need to fulfill a special issue would lead to more inclusive criteria. Building a consensus on validation objectives can be performed by providing reviewers with explicit guidelines, but also through the frame given to them.

In fact, the level of proceduralization of evaluation test shapes contrasted frames: is it a case of carefully defining all of the tasks which contribute towards judgments, or of relying on a model of expertise, with knowledge and know-how being incorporated by science gatekeepers? This tension between experts and procedural evaluators is especially evident in the documents, which are given to reviewers to help them make a judgment. When journals ask reviewers to give an opinion on a manuscript, their verdict must be accompanied by documents of extremely variable format, be they paper sheets or electronic systems. In certain cases, reviewers are faced with a list of questions to answer; in other cases they have to take a series of criteria into consideration, or are required to tick long lists of items or to give marks. On the other hand, they might receive no instructions at all, or just a simple reminder of moral values (e.g. promptitude, constructive nature of comments, courtesy, legibility) to bear in mind when writing a report in free-text form. The production of judgments thus oscillates between opinions considered to be separable from the reviewers concerned, in the name of reproducibility, and opinions that are intimately linked to the qualities of the reviewers who give them. In the former case, commensurability comes directly from the reviewing frame, often valuing judging instances equally, while in the latter, it is the editor-in-chief or editorial committee which has to produce commensurability at a later stage, by comparing and interpreting the various evaluations.

The way in which reviewers intervene with regard to the actual content of the manuscripts is another divisive stake⁶. Internal or outside reviewers might consider that their role is to help authors express their ideas and results in the manner, which is the most effective for the journal's readership. They therefore behave as catalysts or midwives of ideas (i.e. maieuticians), and do not modify the theses and results put forward by the authors. Symmetrically, they might see reviewing as an active investment in the production of articles, demanding that the authors provide additional references or proof, and deleting passages they feel to be inaccurate or irrelevant. In such situations judges take on the role of contributors who voluntarily take part in writing the finished article. Any intervention mode, except the publication of an unmodified manuscript, implies some rewriting but in the latter case, evaluators can actually become more valued than the authors themselves when it comes to defining manuscript content. This involvement leads authors actually disagreeing with the content of their published articles (Frey 2003), for example 20% of them of being coerced to put some references in it (Wilhite and Fong 2012) or, on a lighter note, to an accepted manuscript being more the work of the reviewers than of the original authors and,

⁶ Bornmann (2011) has shown that the Mertonian tradition never considers reviewers as coauthors of the manuscript, while European constructionist traditions have insisted on their actual role in rewriting manuscripts, whether directly or via authors themselves.

after four revisions, being given the “article of the year” award by the editorial committee (Baruch et al. 2008, p. 136-137).

These rewriting decisions also depend on the ways judgments are ranked, as some reviewers would support certain changes while others stand for different ones. By using a large number of reviewers, one limits the whole series of evaluator “bias” (Zuckerman and Merton 1971), but it is almost certain that there will be conflicting views on possible modifications to the manuscript or on whether it can be published at all. This leads to opposite forms of decision-making: either outside reviewers may be given the responsibility of determining the fate of a manuscript by majority rule, or the editor-in-chief may decide that it is he/she who will ultimately decide whether or not to publish. Reviewers’ reports can themselves be valued in different ways, depending on the quality attributed to them by the editor-in-chief, on the characteristics of the reviewers themselves (specialist in the field concerned, years worked with the journal, etc.) or on their own degree of competence asserted in their review (Hirschauer 2010). When a journal limits the number of reviewers, especially when it only uses members of its editorial committee, it is asserting its editorial preferences in terms of subject matter and theoretical or methodological approach.

The objectives given to the evaluation process, the level of proceduralization, the nature of the evaluators’ work and how diverging opinions are weighted, engender a combination that, in practice, produces highly diverse evaluation tests. Within this combination, certain regularities have been observed (Ware and Monkman 2008), such as the relative preference for the double-blind in social sciences, the role of guarantor of editorial policy for internal evaluators and of methodological and bibliographical adequacy for outside reviewers. However, the weighting of judging instances is more complex than it might appear, because it is not confined to historically rigid roles. To a very large extent it depends on the types of involvement of the editor-in-chief and his/her associates:

Although we have tables with criteria to fulfill, I actually don’t take much into account referees’ opinions. They only read a bunch of submitted manuscripts, whilst I see all of them. So, their judgment is relative, mine is absolute. On the other hand, I carefully analyze the reasons that support their judgment and I determine whether they are valid.⁷

This extract demonstrates that all of the different valuations pass through an obligatory passage point, that of the editor-in-chief, the only prestigious position in a researcher’s career. Legally speaking, he/she is jointly responsible for all published texts and, in practice, four centuries after Henry Oldenburg’s inaugural decision, he/she is the appeal body for dissatisfied authors. Symmetrically, as he/she is only the temporary embodiment of the journal, he/she is expected to subsume all judgments and devices that contribute towards the evaluation.

In order to minimize this cardinal role, certain journals have operated a significant shift in their evaluation criteria. In developing an exclusively electronic journal and considering that it removes all limits to available space, these journals are valorizing dissemination over validation. Created in 2006, *PLoS One* abandoned all references to manuscript scope or innovation, explicitly stigmatized as subjective, in favor of simple rigor of analysis and

⁷ Interview with the editor-in-chief of a journal dedicated to gender studies (carried out on 26.10.2013).

technical soundness as bases for judgment⁸. Evaluation is here reduced to a quality control, which leaves aside all reviewer activity that aims to improve the manuscript, and all “overselection” by editors-in-chief aiming for excellence or the highest level of originality. The devaluing of the validation worth goes hand in hand with a valorization of dissemination through readers, be they the authors’ traditional peers or a far wider readership:

PLoS One redefines what a scientific journal should be – eliminating needless barriers between authors and their audience and transforming the published literature from a static series of articles into a dynamic, interconnected, and constantly evolving resource for scientists and the public.⁹

The diversification of readership is threefold: *PLoS One* is a generalist journal, bringing together numerous specialist publics; it has become the journal which publishes the largest number of articles in the world (23,406 in 2012, 31,500 in 2013) and which does not divide the world between an author elite and a mass of readers; it explicitly targets non-academic audiences and valorizes the use of social networks to disseminate its articles, published daily, as soon as they have been validated and without any subscription being required.

This formula is very successful, to such an extent that leading commercial publishers have copied it to develop their own “mega journal” in the biomedical field (BMJ Open) or for all other scientific milieus, including human and social sciences (Sage Open, Springer Plus). Like other innovations it was developed with a view to accelerating the dissemination of articles, but it also carries with it the legitimization of a fourth evaluating instance – the journal or article reader.

Readers as a Judging Instance

Reading is obviously at the heart of manuscript evaluation: it involves the editor-in-chief, members of the editorial committee and outside reviewers. They all produce judgments based on a preliminary reading, fast or slow, in-depth or not. However, this reading takes place within a very specific framework (Hirschauer 2010). On the one hand, it is confined to restricted interaction, which essentially operates via written correspondence (letters or emails). On the other hand, it aims at authorizing the dissemination of manuscripts-become-articles. But other forms of reading accompany publications and participate in their valuation and evaluation, independently of validation. This is particularly the case through citing, commenting, examining and sharing; four operations that may become tests for manuscripts and articles.

Citing Articles

With the popularization of bibliometric tools, citation counting has become a central element of journal and article valuation. The implementation of these tools nevertheless required a series of operations on articles themselves. The identification of citations meant that one first had to homogenize forms of referencing and isolate the references (Bazerman 1988). From

⁸ <http://www.plosone.org/static/reviewerForm.action> (accessed 21.01.2014).

⁹ http://www.eurekalert.org/pub_releases/2006-06/plos-plo060706.php (accessed 01.04.2013).

among all the texts they have read, readers thus choose those which they believe to be of essential value so as to specifically refer to them in their own manuscripts. Secondly, the tools made it necessary to blur the difference between reference and citation: the act of referencing relates to a given author whereas a citation is a new and perhaps calculable property of the source text. According to Wouters (1999), this reversal radically modified referencing practices and literally created a new “citation culture”. Under this condition, readers become citers who add their voices to the already-published article and to the journal which validated it.

This citing activity produces different valuations which might relate to journals (e.g. impact factor, eigenfactor, etc.), to articles (e.g. article level metrics), to authors (e.g. h-index), or even to entire disciplines (e.g. half-life index) or institutions (e.g. a score for all international classifications). Using citation aggregation tools, it is possible to equitably assess all citers or else to introduce weighting tools relating to time span, to the reputation of the outlet, to their centrality, etc. Because citers are always readers – albeit minimalist – there is no strict opposition between peer review and bibliometric tools in evaluation, just different ways of sequentially interlinking them (Donovan 2007).

These different tools have in common that they cause the use to which the reader puts the original text to disappear. Highly disparate forms of intertextuality are thus rendered commensurable: the measured or radical criticism of a thought or result, integration within a scientific tradition, reliance on a standardized method described elsewhere, existence of data for a literary journal or meta-study, simple recopying of sources referenced elsewhere or self-promotion (Erikson and Erlandson 2014). Citation thus points towards two complementary horizons of reading: science as a system for accumulating knowledge via a referencing operation, and research as a necessary discussion of this same knowledge through criticism and commentary.

Commenting Texts

The second horizon is guided by the desire to organize scientific discussion in a different way: exchanges between the different evaluation stakeholders take place in a different publication space which is explicitly dialogical or polyphonic.

When comments are valorized *before* an article is published, the manuscript evaluation test is no longer mainly directed towards the editor-in-chief or the editorial committee. It directly addresses the authors, particularly in certain types of open review. When signing their evaluation reports, reviewers are not just making themselves publicly identifiable. They can also enter into a dialogue with the authors and thus open up a space for direct confrontation. Before the emergence of electronic spaces for discussion, two journals explicitly made pre-publication commentaries the very principle behind their manuscript evaluation policy: *Current Anthropology (CA)* created in 1960 and *Behavioral and Brain Sciences (B&BS)* founded in 1978. Rather than gathering the opinions of just a few outside reviewers, they systematically contact them in large numbers (15 for *CA*, 8 for *B&BS*) in an attempt to have the greatest possible diversity of points of view. They believe that this allows them to limit reviewer bias, at the expense of profound disagreements on the interest and limits of the manuscript in question. Yet unlike numerous other journals, where disagreements are a

problem, in this case they are considered to be “creative” (Harnad 1979), control dissemination being very much a part of the process of validation through discussion.

The publication of commentaries *alongside* that of the articles themselves has existed for some time – “special issues” or “reports” in which a series of articles are brought together around a given theme to feed off one another after a short presentation. Similarly, the longstanding practice of a commentary followed by the author’s response is a common one. *CA* and *B&BS* employ sophisticated versions of this formula, known as open commentary: once a manuscript has been accepted, they invite dozens of researchers to comment on it, and then give the author(s) the opportunity to provide a short response to the comments as a whole. Even if a manuscript has been validated by the journal, the validation process does not prevent a complementary evaluation, either of the article itself, or in relation to other texts.

Finally, proposals have been made to revamp the traditional field of commentaries *after* publication. For a long time, these commentaries existed in two elementary forms: on the one hand in the form of a letter to the editor offering a point of view, accompanied by reactions and impressions, on an article published in the same journal; on the other hand, as we have just seen, in the form of a reference. As from the 1990s, the emergence of electronic publications was seen as something which would revolutionize “*post-publication peer review*”, by allowing comments and criticisms to be added to the document itself (Lancaster 1995). Among existing journals, *F1000Research* (created in 2012) exemplifies the values of rapidity and open scientific discussion. Electronically submitted manuscripts are published as they arrive, accompanied by their sets of data. Putting them on line triggers an evaluation process which is rapidly performed in accordance with the principles of open review: evaluation reports come with the name and institutional affiliation of the reviewers; the authors can react and produce new versions of their manuscripts; finally, at any time during the process the reports and different versions of the article may be commented upon by readers (open commentary). As the manuscript has already been published, reviewers’ opinions no longer affect its validation; they simply provide additional valuations in the form of notations or comments for readers.

Examining and Sharing Documents

The readers mentioned so far have been peers of the authors of the original manuscript in a very restrictive sense: either their reading leads to a text of an equivalent nature, or it leads to a text published in the same outlet as the article. Until recently, readers other than citers and commentators remained very much in the shadows: library users, students in classes and colleagues in seminars also ascribe value to articles. But two major changes have rendered part of these forms of reading valuable.

The existence of articles in electronic form has made their readers visible. People who access an “html” page or who download a “pdf” file are now taken into account, whereas in the past it was only the distribution of journals and texts which allowed one to assess potential readership. By inventorying and aggregating the audience in this way, it is possible to assign readers the capacity to value articles. Labels such as “highly accessed” or “most downloaded”, frequently used on journal websites, make it possible to distinguish certain

articles. The creation of online academic social networks (ResearchGate, Academia...) has trivialized this figure of the public. Researchers now take part in the dissemination of their own articles and are thus able to grasp the extent and diversity of their audiences.

At the same time, other devices make visible the sharing of articles. First of all it is online bibliographic tools (CiteULike, Mendeley, Zotero...) which objectify the readers and taggers who introduce references and attached documents into their bibliographic databases. Without being citers themselves, these readers select publications by sharing lists of references the pertinence of which is notified by the use of "tags". These reader-taggers are also embedded in the use of "generalist" social networks (FaceBook, Twitter, etc.), by alerting others to interesting articles, or by briefly commenting on them. These different channels for dissemination and sharing have been the object of numerous works (Priem and Costello 2010; Eysenbach 2011) which aimed to determine whether or not they were a means of valuing articles compared to their citations. They have also been reworked by advocates of "article level metrics". The measurements of these different channels are now aggregated and are considered to be a representation of the multiple uses and audiences. For its advocates, the resulting "total impact" is the true valuation of a article's importance. Here the readers, tracked by number and diversity, validate articles in the place of the judging instances historically qualified to do so.

This movement is even more significant in that these tools are applied not only to published articles but also to documents which have not been validated through journal peer review. Indeed, after the arXiv high-energy physics case at the beginning of the 1990s (Bohlin 2004; Gunnarsdóttir 2005), many scientific milieus and institutions acquired repository servers to host *working papers*. Ideally, these manuscripts are preliminary versions submitted for criticism and comments by specialist groups which are notified of the submissions. The resulting exchanges are managed by the system which archives the different versions produced. So readers do not simply exercise their judgment on validated articles, but also produce a collective evaluation of manuscripts. This flow of electronic manuscripts feeds the enthusiasm of the most visionary, who since the 1990s have been announcing the approaching end of validation by journals' traditional judging instances.

With a view to the dissemination, advocates of readers as a judging instance tend to moderate the importance of validation. While the validation process sorts manuscripts in a binary fashion (accepted or rejected), the varying forms of dissemination encourage permanent discussion and argument along a text's entire trajectory. In this perspective, articles remain alive after publication and are therefore always subject not only to various reader valuations, but also to public evaluations, which can be considered as more important than their initial validation.

Conclusion

Throughout the history of scientific manuscript evaluation, four judging instances have gradually emerged: the editor-in-chief, the editorial committee, outside reviewers and readers. Driven by a constant process of specialization, this multiplication of judging instances may appear as a reallocation of expertise, empowering a growing number of people in the name of distributed knowledge. Nevertheless, their respective importance and

the way in which these judging instances are coordinated may be subject to local conventions, at a journal, discipline or historical level, and they are marked by profound divergences due to distinct issues in manuscript evaluation and valuation. As we have shown – particularly with regard to author and reviewer anonymity – the space of possibilities can be a subject for public debate, lead to the invention of labels and the stabilization of categories, and to the elaboration of procedural and moral norms. Merton’s “organized skepticism” (1942) and the agonistic nature of the production of scientific facts described by Latour and Woolgar (1979) are not self-evident. The conditions for operating a cross-evaluation are at the heart of the debates and innovations which punctuate its effective implementation. From one arrangement to another, there are thus significant asymmetries between those who glorify an all-seeing, all-powerful editor-in-chief, those who make open peer review the central pivot of objective judgment, those who combine the editorial committee’s decision-making with the sub-contracting of manuscripts to companies such as *Rubriq* or *Peerage of Science*, and those who, like *PLoS One*, place different audiences at the heart of the valuation process.

Over the trajectory of journal peer review technologies, manuscript evaluation has itself become an object of scientific investigation (Campanario 1998). Authors, manuscripts, reviewers, journals and readers have been scrupulously examined for their qualities and competencies, as well as for their “biases”, faults or even unacceptable behavior. This trend has risen with the pioneering work of Peters and Ceci (1982) who resubmitted to journals articles that they had already published, simply replacing the names of the authors and their institutions with fictitious names and making minor changes to the texts. Much to their surprise, almost all of the manuscripts were rejected, and, three exceptions aside, without any accusation of plagiarism. Thirty years later, hundreds of studies on manuscript evaluation are now available (Chubin and Hackett 1990; Speck 1993; Weller 2001). The diverse arrangements of manuscript evaluation are thus themselves systematically subjected to evaluation procedures. For example, in order to comparatively value single blind and double blind, studies have increasingly used randomized controlled trials (Blank 1991; van Rooyen et al. 1998; Godlee et al. 1998), leading to opposite results and recommendations for journal editors. Contrary to ethical norms for which intensive talks have led to the writing of homogenized standards¹⁰, journal peer review shows a growing diversity within valuation and evaluation technologies. Whilst innovators are still trying to develop the ideal journal¹¹, aggregating the best ways to value and evaluate, they are doing so on their own, or at best in limited partnerships.

This quest for the ideal technology is being conducted in an ongoing context of revelations of scientific fraud (Pontille and Torry 2012), which often implicates editorial processes and journals themselves. Faced with the increasing number of proven cases of image manipulation, plagiarism, invented data sets and dishonest reviewers, manuscript evaluation and article valuation are undergoing a transformation. The dereliction inherent to judging

¹⁰ See “The COPE Ethical Guidelines for Peer Reviewers” (<http://publicationethics.org/files/Peer%20review%20guidelines.pdf>, accessed 07.07.2014); and the ICMJE “Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals” (<http://www.icmje.org/recommendations/>, accessed 07.07.2014).

¹¹ <http://chronicle.com/blogs/percolator/building-the-perfect-journal/30794> (accessed 01.10.2012).

instances prior to publication (Schroter et al. 2008) is extending “post-publication peer review”: the mass of readers is becoming a vital resource for unearthing error and fraud¹². As in other domains where public expertise used to be exclusively held by a few professionals, crowdsourcing has become a collective gatekeeper for science publishing. Thus peerdom shall be reshaped, as lay readers have now full access to a large part of the scientific literature (Archambault et al. 2013) and have become valued audiences as quantified end-users of published articles.

In the scientific world, journals have to a large extent become a point of obligatory passage for the valuation and evaluation of academics, transforming “the cycle of credibility” (Latour and Woolgar 1979) into a circular explanation; as Macdonald and Kam (2007, p. 703) ironically put it: “a journal is a quality journal because it publishes quality papers written by quality authors, known to be quality authors because their papers appear in quality journals”. While this scientific credit given to academics and journals depends on the peer review system, perceived as a coherent whole, we have shown how diverse it is and how many journal peer review technologies embody opposite views on manuscript evaluation. The evaluation and valuation of scholars and journals is thus based on a largely overlooked, multiple and invisible infrastructure. Faced with this complexity in evaluation technologies, we have shown that some actors are currently trying to downsize them in order to emphasize the dissemination of articles over manuscript validation. In so doing, they generate oceans of available literature for which valuation technologies are being developed. Rather than looking to have full control of producers of published science, they trust users to be able to separate the wheat from the chaff. If they succeed, valued readers in different shapes and forms would take over from journals as the ultimate guarantors of the quality of texts in circulation.

Acknowledgments

The authors are grateful to Daniel Céfaï, Bénédicte Zimmermann and three anonymous referees for their thoughtful comments and very helpful suggestions on a previous version of this text. We also thank Chris Hinton for his translating assistance.

References

- Auranen, O., & Nieminen, M. (2010). University research funding and publication performance – An international comparison. *Research Policy*, 39(6), 822–834.
- American Psychological Association (1965). Publications in APA Journals: advices from the editors. *American Psychologist*, 20(9), 711–712.
- American Sociological Review. (1955). Notice to contributors, 20(3), 341.
- Archambault E., Amyot D., Deschamps E, Nicol A., Rebut L., & Roberge G. (2013). *Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011*, Science Metrix for the European Commission DG Research & Innovation.

¹² As a result, the number of articles withdrawn after publication has increased exponentially over the last ten years (Van Noorden 2011).

- Baruch, Y., Konrad, A. M., Aguinis, H., & Starbuck, W. H. (Eds.). (2008). *Opening the black box of editorship*. Palgrave Macmillan.
- Bazerman, C. (1988). *Shaping written knowledge: the genre and activity of the experimental article in science*. Madison, Wisconsin: The University of Wisconsin Press.
- Benedek, E. P. (1976). Editorial practices of psychiatric and related journals: implications for women. *American Journal of Psychiatry*, 133(1), 89–92.
- Douglas-Wilson, I. (1974). Twilight of the medical journal? *British Medical Journal*, 3(5926), 326–327.
- Berg, L.D. (2001). Masculinism, emplacement, and positionality in peer review. *The Professional Geographer*, 53(4), 511–521.
- Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review. *American Economic Review*, 81(5), 1041–1067.
- Bohlin, I. (2004). Communication regimes in competition: The current transition in scholarly communication seen through the lens of the sociology of technology. *Social Studies of Science*, 34(3), 365–391.
- Bollen, J., Van De Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS One*, 4(6), e6022.
- Boltanski, L., & Thévenot, L. (2006). *On Justification: Economies of Worth* (p. 400). New Jersey: Princeton university Press.
- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45(1), 197–245.
- Broad, W., & Wade, N. (1982). *Betrayers of the truth. Fraud and deceit in the hall of science* (p. 256). New York: Simon & Schuster.
- Bruno, I., & Didier, E. (2013). *Benchmarking : l'état sous pression statistique* (p. 208). Paris: La Découverte.
- Burnham, J. C. (1990). The evolution of editorial peer review. *Journal of the American Medical Association*, 263(10), 1323–1329.
- Butler, L., & McAllister, I. (2009). Metrics or peer review? Evaluating the 2001 UK research assessment exercise in political science. *Political Studies Review*, 7(1), 3–17.
- Campanario, J. M. (1998). Peer review for journals as it stands today—Part 1. *Science Communication*, 19(3), 181–211.
- Chubin, D. E., & Hackett, E. J. (1990). *Peerless science: Peer review and U.S. science policy* (p. 267). Albany: State University of New York Press.
- Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, 14(1), 119–135.
- Cicchetti, D. V., & Conn, H. O. (1976). A statistical analysis of reviewer agreement and bias in evaluating medical abstracts. *The Yale Journal of Biology and Medicine*, 49(4), 373–383.

- Cole, S., Cole, J. R., & Simon, G. A. (1981). Chance and consensus in peer review. *Science*, 214(4523), 881–886.
- Crane, D. (1967). The gate-keepers of science: Some factors affecting the selection of articles for scientific journals. *The American Sociologist*, 2(1), 195–201.
- Cronin, B., & Sugimoto, C. R., (2014). *Beyond bibliometrics. Harnessing multidimensional indicators of scholarly impact*, MIT Press.
- DeBakey, L. (1976). *The scientific journal: editorial policies and practices: guidelines for editors, reviewers, and authors*. Saint Louis: CV Mosby Company.
- Donovan, C. (2007). Introduction: Future pathways for science policy and research assessment: metrics vs peer review, quality vs impact. *Science and Public Policy*, 34(8), 538–542.
- Erikson M. G., & Erlandson P. (2014). A taxonomy of motives to cite. *Social Studies of Science*, 44(4), 625-637.
- Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual review of sociology*, 24, 313–343.
- Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, 13(4), e123.
- Frey, B. (2003). Publishing as prostitution? Choosing between one's own ideas and academic success. *Public Choice*, 116(1-2), 205–223.
- Glenn, N. D. (1976). The journal article review process: Some proposals for change. *The American Sociologist*, 11(3), 179–185.
- Godlee, F., Gale, C. R., & Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports. *Journal of the American Medical Association*, 280(3), 237–240.
- Gunnarsdottir, K. (2005). Scientific journal publications: On the role of electronic preprint exchange in the distribution of scientific literature. *Social Studies of Science*, 35(4), 549–579.
- Hargens, L. L. (1988). Scholarly consensus and journal rejection rates. *American Sociological Review*, 53(1), 139–151.
- Harnad, S. (1979). Creative disagreement. *The Sciences*, 19, 18–20.
- Helgesson, C.-F., & Muniesa, F. (2013). For what it's worth: An introduction to valuation studies. *Valuation Studies*, 1(1), 1–10.
- Hicks, D., & Wang, J. (2011). Coverage and Overlap of the New Social Sciences and Humanities Journal Lists. *Journal of the American Society for Information Science and Technology*, 62(2), 284–294.
- Hirschauer, S. (2010). Editorial judgments: A praxeology of 'voting' in peer review. *Social Studies of Science*, 40(1), 71–103.
- Ingelfinger, F. J. (1969). Definition of 'sole contribution'. *New England Journal of Medicine*, 281(12), 676–677.

- Jones R. (1974). Rights, wrongs and referees. *New Scientist*, 61(890), 758–759.
- Kennefick D., “Einstein Versus The Physical Review”, *Physics Today*, 2005, vol. 58 (9), p. 43–48.
- Knox, F. G. (1981). No unanimity about anonymity. *Journal of Laboratory and Clinical Medicine*, 97(1), 1–3.
- Kronick, D. A. (1962). *A history of scientific and technical periodicals: the origins and development of the scientific and technical press, 1665-1790*. Metuchen, N.J.: The Scarecrow Press.
- Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment* (p. 336). Cambridge, Mass. ; London: Harvard University Press.
- Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38(1), 201–221.
- Lancaster, F. W. (1995). Attitudes in academia toward feasibility and desirability of networked scholarly publishing. *Library Trends*, 43(4), 741–752.
- Latour, B., & Woolgar, S. (1979). *Laboratory life: the social construction of scientific facts*. Beverly Hills: Sage.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Milton Keynes: Open University Press.
- Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64(1), 2–17.
- Lowry, R.P.(1967). Communications to the editors, *The American Sociologist*, 2(4), 220.
- Macdonald, S., & Kam, J. (2007). Aardvark et al.: quality journals and gamesmanship in management studies. *Journal of Information Science*, 33(6), 702–717.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2), 161–175.
- Merton, R. K. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology*, 1, 115–126.
- Morgan, P. P. (1984). Anonymity in medical journals. *Canadian Medical Association Journal*, 131(9), 1007–1008.
- Nature. (1974). In defence of the anonymous referee. *Nature*, 249(5458), 601–602.
- Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5(2), 187–195.
- Pontille, D., & Torny, D. (2010). The controversial policies of journal ratings: Evaluating social sciences and humanities. *Research Evaluation*, 19(5), 347–360.
- Pontille, D., & Torny, D. (2012). Behind the scenes of scientific articles: Defining categories of fraud and regulating cases. *Revue d'Épidémiologie et de Santé Publique*, 60(4), 247–253.
- Pontille, D., & Torny, D. (2013). La manufacture de l'évaluation scientifique : algorithmes, jeux de données, outils bibliométriques. *Réseaux*, (177), 25–61.

- Pontille, D., & Torny, D. (2014). The blind shall see! The question of anonymity in journal peer review. *Ada: A Journal of Gender, New Media, and Technology*, 4, doi:10.7264/N3542KVV.
- Porter, J. R. (1964). The Scientific Journal - 300th Anniversary. *Bacteriological Reviews*, 28(3), 211–230.
- Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1–4.
- Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L., & Smith, R. (2008). What errors do peer reviewers detect, and does training improve their ability to detect them? *Journal of the Royal Society of Medicine*, 101(10), 507–514.
- Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton, N.J.: Princeton University Press.
- Shapiro, B. J. (2000). *A culture of fact: England, 1550-1720* (p. 284). Ithaca and London: Cornell University Press.
- Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99(4), 178–182.
- Speck, B. W. (1993). *Publication peer review: An annotated bibliography*. Westport, CT: Greenwood Press.
- Van Noorden, R. (2011). Science publishing: The trouble with retractions. *Nature*, 478(7367), 26–8.
- Van Rooyen, S., Godlee, F., Evans, S., Smith, R., & Black, N. (1998). Effect of blinding and unmasking on the quality of peer review. *Journal of General Internal Medicine*, 14(10), 622-624.
- Ward, W.D., Goudsmit, S.A. (1967). Reviewer and author anonymity. *Physics Today* 20(1), 12.
- Ware, M., & Monkman, M. (2008). *Peer review in scholarly journals: Perspective of the scholarly community - an international study*. Bristol, UK, Mark Ware Consulting.
- Weller, A. (2001). *Editorial peer review: Its strengths and weaknesses* (p. 342). Medford: Information Today, Inc.
- Wilhite, A. W., & Fong, E. A. (2012). Coercive citation in academic publishing. *Science*, 335(6068), 542-543.
- Wilson, J. D. (1978). Peer review and publication. Presidential address before the 70th annual meeting of the American Society for Clinical Investigation, San Francisco, California, 30 April 1978. *Journal of Clinical Investigation*, 61(6), 1697–1701.
- Wouters, P. (1999). *The citation culture*. Amsterdam: University of Amsterdam.
- Wouters, P., & Costas, R. (2012). *Users, narcissism and control – tracking the impact of scholarly publications in the 21st century* (p. 50). Utrecht, SURFfoundation.
- Zuckerman, H. A., & Merton, R. K. (1971). Patterns of evaluation in science: institutionalisation, structure and functions of the referee system. *Minerva*, 9(1), 66–100.