Quality Control in Scholarly Publishing: A New Proposal*

Stefano Mizzaro
Department of Mathematics and Computer Science
University of Udine
Via delle Scienze, 206
I33100 Udine, Italy
mizzaro@dimi.uniud.it
http://www.dimi.uniud.it/~mizzaro

February 27, 2003

Abstract

The Internet has fostered a faster, more interactive and effective model of scholarly publishing. However, as the quantity of information available is constantly increasing, its quality is threatened, since the traditional quality control mechanism of peer review is often not used (*e.g.,* in online repositories of preprints, and by people publishing on their Web pages whatever they want).

This paper describes a new kind of electronic scholarly journal, in which the standard submission-review-publication process is replaced by a more sophisticated approach, based on judgments expressed by the readers: in this way, each reader is, potentially, a peer reviewer. New ingredients, not found in similar approaches, are that each reader's judgment is weighted on the basis of the reader's skills as a reviewer, and that readers are encouraged to express correct judgments by a feedback mechanism that estimates their own quality. The new electronic scholarly journal is described in both intuitive and formal ways. Its effectiveness is tested by several laboratory experiments that simulate what might happen if the system were deployed and used.

Keywords: Scholarly publishing, electronic publishing, quality control, peer review.

Introduction

This paper proposes a new model for quality control in electronic scholarly communication. In this model, the standard submission-review-publication process is replaced by a more sophisticated approach, based on judgments expressed by the readers: in this way, each reader is, potentially, a peer reviewer. New ingredients, not found in similar approaches, are that each reader's judgment is weighted on the basis of the reader's skills as a reviewer, and that readers are encouraged to express correct judgments by a feedback mechanism that estimates their own quality.

In this Introduction, the background scenario for the model is presented: scholarly publishing, both electronic and traditional, and peer review are briefly surveyed, and the outline of the rest of the paper is described.

*To appear in Journal of the American Society for Information Science & Technology (JASIS&T). The printed version might be slightly different from this one.

Scholarly publishing and peer review

The communication mechanism still used within modern science arose in 1665, when the first scholarly journals were published in order to report, in paper form, the ideas, discoveries, and inventions of researchers. The control of quality of the published papers was, of course, an issue; at first, since the number of submitted papers was low, the editor of the journal could review them. Then, because of the increasing number and specialization of submitted papers, their quality assessment was given to the *peers*, *i.e.,* experts working in the same field of the author. Still nowadays, the dissemination of scholarly information is therefore based on *peer review*: the researcher who wants to disseminate her work writes a paper and submits it to a scholarly journal; the paper is not immediately published, rather it is sent by the editor to some (usually two or three) referees that judge it; the editor, on the basis of referee's reports, decides if the paper is to be published. The peer review mechanism ensures a reasonable *quality* of the published papers, and it is usually considered an adequate solution, though not the ideal one.

Even though far from being perfect, the mechanism of scholarly publishing based on peer review has been working in a satisfying way for centuries. Nevertheless, the situation is changing and scholarly publishing is undergoing a (mainly economical) crisis (Borgman, 2000; Kling and Callahan, 2003; Apt, 2001a,b): many libraries have discontinued several journal subscriptions because of the increasing cost of scholarly journals; at the same time, the Internet opens new challenges and discloses new possibilities, both for scholarly publishing and for peer review, as discussed in the following.

Electronic scholarly publishing

The Internet has caused a stir in the publishing world (Lesk, 1997; Bailey, 2002; Wells, 1999). Nowadays, digital libraries and electronic publishing are a reality: what could previously be published and circulated only by paper and ink, can now be stored and distributed by electronic means. Also scholarly publishing and communication are being affected by the Internet: today a significant portion of communication among scholars takes place electronically. When analyzing the new scenario, one of two positions is usually taken (Borgman, 2000):

- *Supporters.* Scholarly publishing, as it is today, is dead. The new technologies make it possible to discard paper and ink, to speed up the scholarly communication process and to make it more effective. This is probably the position taken by the majority of scholars (at least in the hard sciences).
- *Detractors.* Be careful: first of all, when a new technology is introduced, people need some time (usually a long time) to get acquainted with it and, secondly, history shows that "new technologies have supplemented, rather than supplanted, old way of doing things" (Borgman, 2000, p. x). Also differences among scholarly fields may lead to different reactions to the new technologies.

Let us analyze more in detail these two positions. On the one side, supporters claim that a peer reviewed scholarly journal can be distributed by electronic means, leading to a more effective scholarly communication: electronic scholarly communication is faster, cheaper, and discloses new possibilities for alternative models (Ginsparg, 1994, 1996; Harnad, 1990, 1995, 1996, 2000; Odlyzko, 1995, 2002; Okerson, 1991; Resh, 1998). The refereeing process can take place completely electronically, drastically reducing time and money: see, *e.g.,* JAIR (http://www.jair.org), JHEP (http://jhep.sissa.it), and Earth Interactions (http://EarthInteractions.org). Multimedia information can lead to a more effective communication (Holoviak and

Seitter, 1997). It seems unfair that authors, editors, and peer reviewers work for free, whereas publishers earn money and libraries receive funds (Harnad, 2001a). Therefore, publishers and libraries can be removed from the scholarly communication process, thus reducing the costs. To reach this aim, authors and their institutions should publish on their own, as in the Guild Model (Kling et al., 2002) or as suggested by LaPorte et al. (1995) and Harnad (2001b). The Open Archives Initiative (http://www.openarchives.org) and the Eprint free software (http://www.eprints.org) are already available tools that can be used to implement this model.

The supporters' viewpoint is strengthened by several examples of full-featured electronic communication systems already working and widely used (Kling and McKim, 2000; Peek et al., 1998). Alternative models of scholarly communication have been already deployed. The most known is probably the ArXiv repository for mainly physics, mathematics, and computer science (http://arXiv.org), but this model has met with some success also in other fields (see, *e.g.,* http://cogprints.soton.ac.uk). Several software tools for journal manuscript management and peer review are already available (McKiernan, 2002). Supporters note that, of course, there are also some drawbacks of electronic journals (copyright problems, legal validity, accessibility, and so on), and they had a slow start (Harter, 1998), but this is a temporary situation, the trend is constantly increasing (Brown, 2001; Odlyzko, 2002; Zhang, 2001), and we just have to wait some years for overcoming these temporary problems.

On the other side, detractors highlight some pitfalls in which supporters often incur, and some problems with scholarly electronic communication as well. One problem is that the term "electronic journal" is used in an ambiguous manner, often implicitly. Rather, it should be emphasized that there are various kinds of electronic journals, and that other communication means like personal, institutional, or world-wide e-print repositories, commentaries, email, etc. are not electronic journals (Kling and McKim, 1997; Kling and Callahan, 2003). Also, bibliometric studies on scholarly communication (Borgman and Furner, 2002) emphasize how electronic scholarly communication means are perceived differently from standard paper journals (Harter and Ford, 2000). The results about the growth of electronic scholarly publishing usually do not take these issues into account.

Detractors remark that supporters usually adopt a "Standard Model" of scholarly communication, that disregards important social issues, whereas a more comprehensive "Socio-Technical Network Model" can be devised and should be used (Kling et al., 2001; Kling and Callahan, 2003). By relying on the more general model, several of the supporters' claims are seen under a different light. Scholars often feel that the quality of electronic scholarly journals is lower than standard paper journals (Kling and Covi, 1995; Harter, 1998; Zhang, 2001), and this may prevent or delay the coming of the electronic scholarly communication age. Also, there are differences among fields (Kling and McKim, 2000; Kling and Callahan, 2003) that cannot suddenly disappear. Thus, Ginsparg's (1994; 1996) ArXiv, with its great success in high-energy physics, might be an inadequate model in other fields with different features and work practices. For instance, timeliness and rapidity in publication of results are important in physics, and are of almost no importance in classical literature; moreover, the success of electronic communication in physics is also based on a previously established "preprint practice" (in which researchers in physics were used to circulate paper preprints before submission and publication, with the aim of receiving feedback). Also, by using the "Socio-Technical Network Model", the claim that electronic communication is faster and cheaper than paper communication turns out to be exaggerated, since most of the time in the peer review process is taken by the referee (and this is not changed in the electronic environment), and it is difficult to estimate correctly the costs of the equipment used in the referee process (computers, network connection, and so on). Finally, libraries and publishers have important roles (see, *e.g.,* Borgman, 2000): libraries preserve and select, and publishers disseminate. Removing them from the publishing cycle might be extremely dangerous.

Peer review in electronic scholarly publishing

As mentioned above, peer review is usually deemed a sensible solution for quality control in scholarly publishing. Indeed, the peer review mechanism has been criticized, even before the Internet coming, and independently from it. Peer review critics remark that, sometimes, the reviewing process can last more than one year, and this might lead, in highly dynamic fields, to published papers describing something old. Sometimes, the reviewers do not do a good job, accepting a bad paper or not accepting a good one. Sometimes these two problems are combined, and after one year a good but rejected paper cannot be resubmitted because it is too obsolete. In certain cases, the referee cannot correctly judge the paper, *e.g.,* when the paper reports data from an experiment that the referee cannot re-perform (Arms, 2002), or when the proof of some mathematical proposition is obtained by running a simulation program for a time longer than that available for the referee process (Odlyzko, 1995). Peer review can introduce some bias in published papers: for instance, in biomedical field, papers describing negative results seem more difficult to publish than papers describing positive ones. Also, the so called "Schön affair" (Service, 2002) is a recent and unpleasant example that peer review might not be able to spot fraudulent behavior (though the scientific community seems capable to do so).

In the Internet age, different models of publishing are possible, and the peer review practice is even more discussed, criticized and, in some respect, changing (Weller, 2000). For instance, a peer reviewed journal can be distributed by electronic means and the refereeing process too can take place completely electronically, thus reducing, accordingly to supporters, time and money: see, *e.g.,* JAIR (http://www.jair.org), JHEP (http://jhep.sissa.it), and Earth Interactions (Holoviak and Seitter, 1997) (http://EarthInteractions.org). Another opportunity is to replace peer review with *peer commentary*: readers will write in a public commentary their judgments on the read papers, and the commentaries will be used to judge the quality of the papers. This seems a viable solution, but Harnad (1997, 2000), after some practical experience, says that "peer commentary is a superb supplement to peer review, but it is certainly no substitute for it".

Several researchers suggest, using Nadasdy's words (1997), to substitute peer review with democracy: each submitted paper is published, possibly before or without a peer review, and readers will judge it, selecting what they deem useful. Variants of this approach are proposed by Nadasdy (1997), Rogers and Hurt (1990), Stodolsky (1990), and Varian (1998): in these proposals, papers are published sometimes after a standard peer review, sometimes after a somewhat reduced peer review, and sometimes without any peer review at all. All these proposals have the common feature that the readers of a published paper can vote for it, thus assessing the quality of the papers. Of course, the problem with this approach is that the readers may not be capable of correctly judging the paper: whereas the referees are chosen among the experts in the field, everybody can read and judge a paper published on the Internet.

It is important to understand that all these proposals are not only abstract models. In a few years, a model of publishing similar to one of these might become a *de facto* practice, as witnessed by the examples, already existing today, of the "do-it-yourself publishing" (authors publishing in a web site their ideas), and the public (institutional or world-wide) repositories of scholarly papers. Money is also an important issue: as mentioned above, journals are often too expensive, whereas the repositories are, at least accordingly to their supporters, cheaper. Supporters and detractors may or may not agree, but, in sum, the threat (or hope) that these (without-quality-control) mechanisms will eventually replace the (with-quality-control) peer review journals is a real one.

In my opinion, it is not important to take side in the debate pro or against peer review, or between supporters and detractors. It is more important to understand that, perhaps, peer review might not be adequate in some cases, and that some alternatives to it might exist, in particular due to the appearance of new technologies. These alternatives might be used both to complement peer review and to replace it, either when the alternatives are more adequate or in those cases in which peer review is not viable. Since it is difficult to forecast the future, and since the above sketched "democratic" alternatives are rather naïve and do not assure a good quality of the published papers, it is interesting to study different, and potentially more adequate, models. This is the aim of this paper.

Quality control in scholarly publishing: A new proposal

The above sketched scenario is the background for the research presented in this paper. I describe a new kind of electronic scholarly journal, that exploits a new model for the submission-review-publication process.[1] The aim of this model is to make scholarly communication more automatic, while at the same time keeping the quality of scientific papers at a high level, and providing a way of measuring in an automatic and objective way the quality of *researchers*. I try to make a step further in the road suggested by the not-refereed journals mentioned above, and to present a more sophisticate mechanism that avoids some of the previously described problems. As it will be discussed in the last section, this mechanism can be seen both as an alternative to peer review and as a supplement to it; it is also an extension and an improvement of the well known *impact factor* mechanism (Garfield, 1972).

This paper, which revises, extends, and formalizes previous work (Mizzaro, 1999, 2000; Mizzaro and Zandegiacomo Riziò, 2000), is structured in the following way. The next two sections describe the mechanism. First I present a general and intuitive description; then the behavior of the whole system is formally defined by means of some formulæ, divided into two groups. The formulæ in the first group (section "Invariant properties") completely specify the behavior of the system by providing a static description: they define some invariant properties, that must hold at each time instant. The formulæ in the second group (section "Updating formulæ") provide a dynamic description, by defining how to update some numeric figures as time evolves. These two different formalizations allow a seamless presentation of the system: the invariant properties are simpler than the updating formulæ, and thus it is straightforward to understand and justify them from an intuitive point of view; the correctness of the updating formulæ is more difficult to grasp at an intuitive level, but they lead to a more efficient system, and the equivalence of the two approaches is formally proved. In the subsequent section ("Examples and discussion") several examples are presented, with a twofold aim: firstly, to better understand how the system works and secondly, to demonstrate the feasibility of this approach in real life, by discussing some potential problems and their solutions. Finally, the last section discusses some open problems and future developments.

[1] This model can be applied not only to scholarly journals, but also to other means of scholarly communication like, for instance, e-prints repositories. For the sake of simplicity, in the following I will restrict the discussion to scholarly journals only, but the proposal is described at an abstraction level that allows its application to other scholarly communication means without any difficulties.

General description

This section intuitively describes the basic idea on which the proposed system is based. More details are provided in the following sections. Let us imagine a scholarly journal in which each paper is immediately published after its submission, without a refereeing process. Each paper has some scores, measuring its quality (accuracy, comprehensibility, novelty, and so on). For the sake of simplicity, in the following I will use a single score, measuring overall quality, but the generalization to multi-dimensional quality measures is straightforward. This score is initially zero, or some predetermined value, and it is later dynamically updated on the basis of the readers' judgments. A subscriber to the journal is an author or a reader (or both). Each subscriber has a score too, initially set to zero (or some predetermined value) and later updated on the basis of the activity of the subscriber (if the subscriber is both an author and a reader, she has two different scores, one as an author and one as a reader). Therefore, the scores of subscribers are dynamic, and change accordingly to subscribers' behavior: if an author with a low score publishes a very good paper, *i.e.,* a paper judged very positively by the readers, her score increases; if a reader expresses an inadequate judgment on a paper, her score decreases accordingly, and so on.

Every object with a score (author, reader, paper) also has a *steadiness* value, representing how much steady the score is: for instance, old papers, *i.e.,* papers that have been read and judged by many readers, will have a high steadiness; new readers and authors will have a low steadiness. Steadiness affects the score update: a low (high) steadiness allows quicker (slower) changes of the corresponding score. While a score changes, the corresponding steadiness value increases.

As time goes on, readers read the papers, judgments are expressed, and the corresponding scores and steadinesses vary accordingly. The score of a paper can be used for deciding to read or not to read that paper; the scores of authors and readers are a measure of their research productivity, then they will try to do their best for keeping their score at a high level, hopefully leading to a virtuous circle (publishing good papers and giving correct judgments to the read papers). A steadiness value is an estimate of how stable and, therefore, reliable the corresponding score is.

For understanding the details of the automatically refereed journal proposed here, let us follow the events that happen when a paper is read and judged by a reader:

1. **Paper.** First of all, the paper score is updated: if the judgment is lower (higher) than the actual paper score, the paper score decreases (increases). The score of the reader determines the weight of the judgment: judgments given by higher-rated readers will be more important, and will lead to higher changes, than judgments given by lower-rated readers.

   The steadiness of the paper increases, since the score of the paper is now computed on the basis of one more judgment, and is therefore statistically more reliable.

2. **Author.** Then, the author's score is updated: when the score of a paper written by an author decreases (increases), the score of the author decreases (increases). Thus, authors' scores are linked to the scores of their papers.

   The steadiness of the author, similarly to the steadiness of the paper, increases, since the score of the author is now obtained with one more judgment and is therefore statistically more reliable.

3. **Reader.** Then the reader's score is updated: if one reader's judgment about a document is "wrong" (*i.e.,* as we will see in the next section, too far from

the average), the reader's score has to decrease. Therefore, the reader's score is updated depending on the *goodness* of her judgment, *i.e.,* how much adequate her judgment is, or how much it agrees with the current score of the paper.

Again, the steadiness of the reader increases, since her score, computed on the basis of the goodness of her judgments, is obtained on the basis of one more judgment.

4. **Previous readers.** Finally, the scores of the readers that previously read the same paper are updated: if a judgment causes a change in a paper score, all the goodnesses of the previously expressed judgments on that paper have to be re-estimated. Therefore, a judgment on a certain paper leads to an updating of the scores of all the previous readers of that paper.

   As before, the steadinesses of the previous readers increase since the goodnesses of the readers, that lead to their scores, are obtained on the basis of one more judgment.

The updating of the scores of the previous readers deserve further explanation. After the paper score has changed, it is possible to revise the goodness of the old readers' judgments, and to update the old readers' score consequently: for instance, if an old reader r expressed a judgment j that was "bad" (distant from the paper score) at that time, but after that the paper score changes and becomes more similar to j, then the score of r (s_r) has to increase. Let us take into account a simple concrete example (Figures 1, 2, and 3 show the temporal evolution):

- At time t_0, we have a paper p with score $s_p(t_0)$, three readers r_1, r_2, and r_3 with their scores $s_{r_1}(t_0)$, $s_{r_2}(t_0)$, and $s_{r_3}(t_0)$.

- At a following time instant $t_1 > t_0$ (Figure 1), reader r_1 reads paper p expressing the judgment $j_{r_1,p}(t_1)$ (continuous double arrow line in figure). This causes the updating of the scores of p (dashed line in figure labelled with 1.1) and r_1 (dashed line labelled with 1.2), obtaining $s_p(t_1)$ and $s_{r_1}(t_1)$.

- At time $t_2 > t_1$ (Figure 2), reader r_2 reads p expressing $j_{r_2,p}(t_2)$. The scores of p and r_2 are updated consequently, leading to $s_p(t_2)$ and $s_{r_2}(t_2)$ (dashed lines labelled with 2.1 and 2.2). But also the score of r_1 has to be updated (dotted line labelled with 2.3), since the goodness estimated at time t_1 for $j_{r_1,p}(t_1)$ with respect to $s_p(t_1)$ has to be re-estimated now that the score of p is $s_p(t_2)$.

- At time $t_3 > t_2$ (Figure 3), r_3 reads p expressing $j_{r_3,p}(t_3)$. This changes the score of p ($s_p(t_3)$, dashed line labelled with 3.1), the score of r_3 ($s_{r_3}(t_3)$, dashed line labelled with 3.2), and the scores of the previous two readers ($s_{r_2}(t_3)$ and $s_{r_1}(t_3)$, dotted lines labelled with 3.3 and 3.4).

In other words, the goodness of a reader's judgment is an approximation of the *ideal goodness*, defined as the difference between the reader's judgment and the final score of the paper (*i.e.,* the score obtained when the last judgment on that paper has been expressed). Since the final score is obviously not available when the judgment is expressed, it has to be estimated (updating of the reader), but this estimate is revised and refined as time evolves and tends to $+\infty$ (updating of previous readers).

Formal description

In this section, some formulæ are presented, in order to formally specify the system. It is shown how to compute, on the basis of the expressed judgments, the values

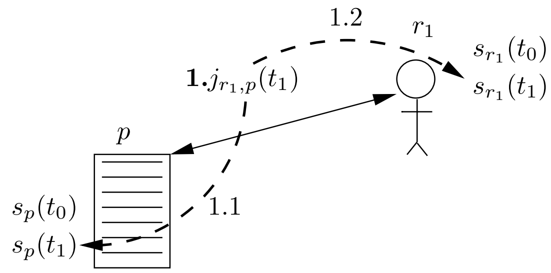

Figure 1: The updating of previous readers' scores: t_1.

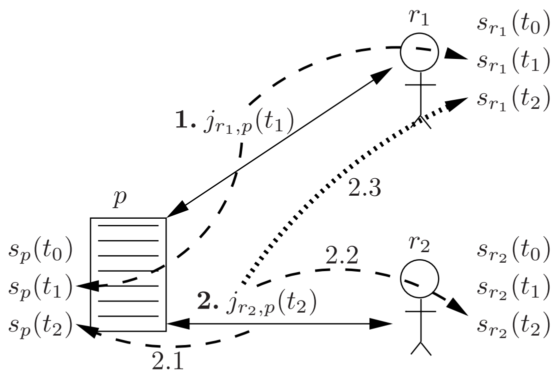

Figure 2: The updating of previous readers' scores: t_2.

of scores and steadiness of papers, authors, readers, and previous readers. In the subsection "Invariant properties" the formulæ describe some invariant properties that score and steadiness values must fulfil. In the subsection "Updating formulæ" the formulæ specify how to *update* the values of the scores and steadiness as new judgments are expressed. Let us start with some notation.

Notation

I will denote with:

- t and t_i the discrete time instants. I assume that t_{i+1} immediately follows t_i, and that between t_i and t_{i+1} only the explicitly specified events will happen.
- $s_p(t), s_a(t), s_r(t)$ the score of a paper, an author, and a reader, respectively, at time t. The time indication will sometimes be omitted, when this does not rise ambiguity. The values for $s_p(t)$ and $s_a(t)$ are in the range $[0, 1]$ (0 is the minimum and 1 the maximum), whereas the values for $s_r(t)$ are in the range $]0, 1]$. This difference will be explained in the following.
- $\sigma_p(t), \sigma_a(t), \sigma_r(t)$ the steadiness of a paper, an author, and a reader, respectively, at time t. All the steadiness values are in the range $[0, +\infty[$.
- $j_{r,p}(t)$ the judgment expressed at time t by reader r on paper p. The time indication will sometimes be omitted, when this does not rise ambiguity.
- $t_{r,p}$ the time instant of the judgment expressed by r on p (I am implicitly assuming that each reader can judge each paper only once). I will write $j_{r,p}$ instead of $j_{r,p}(t_{r,p})$.

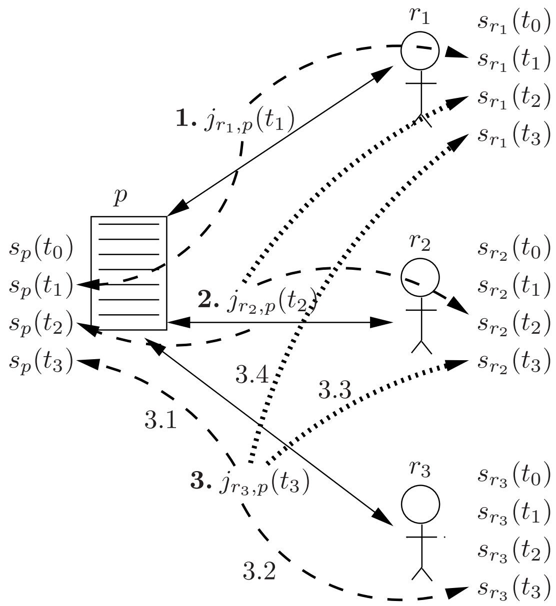

Figure 3: The updating of previous readers' scores: t_3.

Invariant properties

Paper

Given a paper p, its score is the weighted mean of the judgments previously expressed by readers on p. The weight of each judgment is the score that the reader has when she expresses the judgment, to give more importance to the judgments given by best readers.

Definition 1 (Paper score s_p) Given a paper p and the set $R_p(t)$ of readers that expressed a judgment on p before time t, we have $\forall r \in R_p(t)$ the judgment $j_{r,p}$ expressed by r on paper p at time $t_{r,p}$, and the score $s_r(t_{r,p})$ that the reader r had when she expressed the judgment. We then define the score of paper p at time t as:

$$s_p(t) = \frac{\sum_{r \in R_p(t)} s_r(t_{r,p}) \cdot j_{r,p}}{\sum_{r \in R_p(t)} s_r(t_{r,p})}. \tag{1}$$

■

Remark 1 Consistently with Formula 1, the score of the paper is zero before any judgment is expressed on it. It is also possible to use other values, *e.g.,* the score of the author publishing the paper, as well as 0.5, *i.e.,* an intermediate score. However, this choice is not critical, since the initial score has a very limited importance: it does not affect s_p as soon as the first judgment is expressed.

Remark 2 The score $s_p(t)$ of a paper is modified only when a judgment on p is expressed.

Remark 3 In Formula 1, each judgment is weighted on the basis of the score that the reader had when she expressed the judgment ($s_r(t_{r,p})$). The alternative of using the score that the reader has "now", *i.e.,* when the mean is calculated ($s_r(t)$), seems less preferable since the reader's competence has probably changed in this period.

Let us now see how to measure the steadiness of a paper. The steadiness of a paper has to measure how stable its score is. A first tentative definition might be the number of judgments expressed on that paper. However, it seems reasonable that a judgment expressed by a good reader should be more important, and give more steadiness to the paper, than a judgment expressed by a reader with a low score. Therefore, we define the steadiness of paper p at time t as the summation of the scores that readers have when they express their judgments on p.

Definition 2 (Paper steadiness σ_p) Given a paper p, the set $R_p(t)$ of readers that expressed a judgment on p before time t, and $\forall r \in R_p(t)$ the time instants of judgment expression $t_{r,p}$, with $t_{r,p} < t$, the steadiness of p at time t is:

$$\sigma_p(t) = \sum_{r \in R_p(t)} s_r(t_{r,p}). \tag{2}$$

■

Remark 4 Consistently with Formula 2, the steadiness value of a just published, and not yet judged, paper is zero. The alternatives of using the steadiness of the author, or the average steadiness of the published papers, would be also feasible, but more complex. Anyway, the importance of the initial steadiness decreases as judgments are expressed.

Remark 5 The expression in the denominator in Formula 1 is exactly the steadiness of the paper (Formula 2). Therefore, we can rewrite Formula 1 as:

$$s_p(t) = \frac{\sum_{r \in R_p(t)} s_r(t_{r,p}) \cdot j_{r,p}}{\sigma_p(t)}. \tag{3}$$

Author

Given an author a, her score at time t can be defined in two equivalent ways:

- As the weighted mean of the scores of the papers previously published by a. The weight of each paper p is the steadiness of p, a value that sums up all the scores of the readers that expressed a judgment on p (see Formula 2).
- As the weighted mean of the judgments previously expressed by readers on the papers published by a. The weight of each judgment is the score that the reader had when she expressed the judgment.

The first alternative, that uses the steadiness of a paper to weight the papers scores, is formally defined as follows.

Definition 3 (Author score s_a) Given an author a and the set $P_a(t)$ of papers published by a before time t, we have $\forall p \in P_a(t)$ the score $s_p(t)$ of p at time t and the steadiness $\sigma_p(t)$ of p at time t. We now define the score of author a at time t as:

$$s_a(t) = \frac{\sum_{p \in P_a(t)} \sigma_p(t) \cdot s_p(t)}{\sum_{p \in P_a(t)} \sigma_p(t)}. \tag{4}$$

■

Following the second alternative, we can define:

$$s_a(t) = \frac{\sum_{p \in P_a(t)} \left(\sum_{r \in R_p(t)} s_r(t_{r,p}) \cdot j_{r,p} \right)}{\sum_{p \in P_a(t)} \left(\sum_{r \in R_p(t)} s_r(t_{r,p}) \right)}, \tag{5}$$

where $P_a(t)$ is the set of papers published by a before time t, $R_p(t)$ is the set of readers that judged paper p before t, $s_r(t_{r,p})$ is the score of r at time $t_{r,p}$, $t_{r,p} < t$ are time instants of judgment expression, and $j_{r,p}$ is the judgment expressed by r on paper p.

Remark 6 Formulæ 4 and 5 give the same result, *i.e.,* they compute the score of an author in equivalent ways. This is simply proved by using Formula 3, rewritten as

$$\sum_{r \in R_p(t)} s_r(t_{r,p}) \cdot j_{r,p} = s_p(t) \cdot \sigma_p(t),$$

and Formula 2 to substitute the summations in parentheses in Formula 5: Formulæ 4 is obtained.

Remark 7 Consistently with Formulæ 4 and 5, the score of an author before any judgment is expressed on her papers is zero.

Remark 8 $s_a(t)$ is modified only when the score of one of the papers published by a changes, *i.e.,* when a judgment on a paper published by a is expressed.

Remark 9 As discussed in Remark 3, in Formula 5 each judgment is weighted on the basis of the score that the reader had when expressed the judgment ($s_r(t_{r,p})$).

The steadiness of an author has to measure how stable her score is. We can define it in two equivalent ways: either as the summation of the steadiness of her papers, or as the summation of the scores that the readers had when they expressed a judgment on a paper of the author. The first alternative leads to the following definition.

Definition 4 (Author steadiness σ_a) Given an author a, the set of papers $P_a(t)$ published by a, and the steadiness $\sigma_p(t)$ of each paper $p \in P_a(t)$ at time t, the steadiness of author a at time t is:

$$\sigma_a(t) = \sum_{p \in P_a(t)} \sigma_p(t). \tag{6}$$

■

Following the second alternative, we can define (with the usual notation):

$$\sigma_a(t) = \sum_{p \in P_a(t)} \sum_{r \in R_p(t)} s_r(t_{r,p}). \tag{7}$$

The equivalence of Formulæ 6 and 7 follows immediately from Formula 2.

Remark 10 Since the denominator in Formula 4 is exactly the steadiness of the author (Formula 6), we can rewrite Formula 4 as:

$$s_a(t) = \frac{\sum_{p \in P_a(t)} \sigma_p(t) \cdot s_p(t)}{\sigma_a(t)}. \tag{8}$$

Reader

First of all, we need to define a measure of how good a judgment is, on the basis of the distance (difference) between the judgment and the score of the paper.

Definition 5 (Goodness) Given the score $s_p(t)$ of a paper p at time t and the judgment $j_{r,p}$ by reader r on p, the goodness, calculated at time t, of the judgment $j_{r,p}$ is defined as:

$$g_{j_{r,p}}(t) = 1 - \sqrt{|j_{r,p} - s_p(t)|}. \tag{9}$$

■

Remark 11 The square root is used to have more uniformly distributed judgments. As a matter of fact, the values of $|j_{r,p} - s_p(t)|$ are in the range $[0, 1]$. However, if we assume $j_{r,p}$ and $s_p(t)$ uniformly distributed, we do not get a uniform distribution for $|j_{r,p} - s_p(t)|$ (and, consequently, for g). This can be easily seen by observing that the expected value of $j_{r,p}$ and $s_p(t)$ is 0.5, whereas the expected value of $|j_{r,p} - s_p(t)|$ turns out to be 0.25. With the square root correction, the expected value for g is about 0.5.

Remark 12 The goodness depends on the time instant at which it is computed (t in Formula 9), since s_p can change also after the judgment expression (because of other judgments).

The score of a reader r is the weighted mean of the goodness of the judgments she has previously expressed. If we gave the same weight to all the goodnesses, we might define:

$$s_r(t) = \frac{\sum_{p \in P_r(t)} g_{j_{r,p}}(t)}{|P_r(t)|} \tag{10}$$

where $s_r(t)$ is the score of reader r at time t, $P_r(t)$ is the set of papers judged by r before t, $|P_r(t)|$ is the cardinality of $P_r(t)$, $j_{r,p}$ is the judgment expressed by r on paper p, and $g_{j_{r,p}}(t)$ is the goodness, calculated at time t, of the judgment $j_{r,p}$.

However, the weight of each goodness should be different on the basis of the steadiness of the paper being judged. For instance, a wrong judgment on a paper that has been previously judged by many readers should weight more than a wrong judgment on a paper whose score is calculated on the basis of very few judgments. The steadiness of the paper being judged should be taken into account, and the previous Formula 10 for $s_r(t)$ has to be rewritten as follows.

Definition 6 (Reader score s_r) Given a reader r and the set $P_r(t)$ of papers judged by r before time t, we have $\forall p \in P_r(t)$ the steadiness $\sigma_p(t)$ of p at time t, the judgment $j_{r,p}$ expressed by r on p, and the goodness $g_{j_{r,p}}(t)$ of the judgment $j_{r,p}$, calculated at time t. We define the score of reader r at time t as:

$$s_r(t) = \frac{\sum_{p \in P_r(t)} \sigma_p(t) \cdot g_{j_{r,p}}(t)}{\sum_{p \in P_r(t)} \sigma_p(t)}. \tag{11}$$

■

Remark 13 Accordingly to Formula 11, the initial value of s_r should be zero. However, since this would lead to a division by zero in Formula 1 (if we compute the score of a paper that has not been judged yet), and since the initial score is not important (Remark 1), I define this initial value as $s_r = \varepsilon$, where ε is a small value that will be neglected after some judgments are expressed. I will discuss further on the necessity of such an ε in the first example presented in the next section.

Remark 14 The score of a reader r changes when:

- r judges a paper. This adds one element to the set $P_r(t)$ and one new addendum to the summations in Formula 11.
- A reader $r' \neq r$ judges a paper that has previously been judged by r. This causes the score of the paper to change, and this leads to changing the goodness of the judgment previously expressed by r.

Remark 15 All the goodnesses are calculated at time t, to have the best estimate: in Formula 11, we have $g_{j_{r,p}}(t)$, not $g_{j_{r,p}}(t_{r,p})$. For the same reason, the steadinesses are calculated at time t too: $\sigma_p(t)$, not $\sigma_p(t_{r,p})$. Therefore, it is implicitly assumed that the estimate improves with the number of judgments.

The steadiness of a reader has to measure how stable the score of a reader is. A good measure might be the number of judgments expressed by the reader. However, since a judgment expressed on a paper with a high steadiness value should give more steadiness to the reader, I define the steadiness of a reader as the summation of the steadinesses of the papers judged by her.

Definition 7 (Reader steadiness σ_r) If $P_r(t)$ is the set of papers judged by a reader r before time t, the steadiness of r at time t is

$$\sigma_r(t) = \sum_{p \in P_r(t)} \sigma_p(t). \tag{12}$$

■

Remark 16 The denominator in Formula 11 is the steadiness of the reader (Formula 12). Therefore, Formula 11 can be rewritten as:

$$s_r(t) = \frac{\sum_{p \in P_r(t)} \sigma_p(t) \cdot g_{j_{r,p}}(t)}{\sigma_r(t)}. \tag{13}$$

Remark 17 Although the steadiness of a reader is not used in the other formulæ, it is however useful to judge the reliability of a reader. It will also be useful in the next subsection and in one example (subsection "Lazy readers") below.

Remark 18 When the above formulæ are used to compute the score and steadiness values as new judgments are expressed, we have to pay attention to the order of computation. For instance, the score of an author at time t depends on the score of her papers at time t (see Formula 4). Therefore, $\forall p \in P_a(t)$, $s_p(t)$ must be computed before $s_a(t)$. Analogous restrictions are: $\forall p \in P_a(t)$, $\sigma_p(t)$ before $\sigma_a(t)$ (Formula 6); $\forall p \in P_r(t)$, $s_p(t)$ and $\sigma_p(t)$ before $s_r(t)$ (Formulæ 11 and 9); and $\forall p \in P_r(t)$, $\sigma_p(t)$ before $\sigma_r(t)$ (Formula 12).

Updating formulæ

The above presented formulæ are a formal specification of the values that must hold, at a specified time, for scores and steadinesses of papers, authors, and readers (*e.g.,* Formula 1 says that the score of a paper is the weighted mean of the judgments expressed so far by the readers, with the scores of the readers as weights). However, these formulæ might become too complex and long to compute if the number of papers, authors, and readers is high enough, because of the summations and weighted means. To show the feasibility of the mechanism proposed here, it is important to understand if and how score and steadiness updating can take place in a more efficient way. In this section, the above formulæ are rewritten in a way allowing fast computations.

In the following proposition, the values of score and steadiness for papers, authors, and readers (both the reader that is expressing the judgment and the readers that previously judged the paper being judged) at time t_{i+1} are defined in terms of the values at time t_i, avoiding the long summations. In other words, these formulæ show how to *update* the values of score and steadiness at time t_{i+1} on the basis of the values at time t_i and of a judgment expressed at time t_i. Since the formulæ in the previous subsection and those in this section are equivalent, and since the former are simpler than the latter, in the following sections I will not use anymore the formulæ in Proposition 1.

Proposition 1 Given, at time t_i, a paper p with score $s_p(t_i)$ and steadiness $\sigma_p(t_i)$, written by an author a with score $s_a(t_i)$ and steadiness $\sigma_a(t_i)$, a reader r with score $s_r(t_i)$ and steadiness $\sigma_r(t_i)$, another reader r' with score $s_{r'}(t_i)$ and steadiness $\sigma_{r'}(t_i)$, a judgment $j_{r,p}$ expressed on p by r at time t_i, and another judgment $j_{r',p}$ expressed on p by r' at time $t' < t_i$, we have the following values at time t_{i+1}:

1. The steadiness of p at time t_{i+1}:

$$\sigma_p(t_{i+1}) = \sigma_p(t_i) + s_r(t_i). \tag{14}$$

2. The score of p at time t_{i+1}:

$$s_p(t_{i+1}) = \frac{\sigma_p(t_i) \cdot s_p(t_i) + s_r(t_i) \cdot j_{r,p}}{\sigma_p(t_{i+1})}. \tag{15}$$

3. The steadiness of a at time t_{i+1}:

$$\sigma_a(t_{i+1}) = \sigma_a(t_i) + s_r(t_i). \tag{16}$$

4. The score of a at time t_{i+1}:

$$s_a(t_{i+1}) = \frac{\sigma_a(t_i) \cdot s_a(t_i) + s_r(t_i) \cdot j_{r,p}}{\sigma_a(t_{i+1})}. \tag{17}$$

5. The steadiness of r at time t_{i+1}:

$$\sigma_r(t_{i+1}) \quad = \quad \sigma_r(t_i) + \sigma_p(t_{i+1}). \tag{18}$$

6. The score of r at time t_{i+1}:

$$s_r(t_{i+1}) \quad = \quad \frac{\sigma_r(t_i) \cdot s_r(t_i) + \sigma_p(t_{i+1}) \cdot g_{j_{r,p}}(t_{i+1})}{\sigma_r(t_{i+1})}. \tag{19}$$

7. The steadiness of r' at time t_{i+1}:

$$\sigma_{r'}(t_{i+1}) \quad = \quad \sigma_{r'}(t_i) + s_r(t_i). \tag{20}$$

8. The score of r' at time t_{i+1}:

$$s_{r'}(t_{i+1}) \quad = \quad \frac{\sigma_{r'}(t_i) \cdot s_{r'}(t_i) - \sigma_p(t_i) \cdot g_{j_{r',p}}(t_i) + \sigma_p(t_{i+1}) \cdot g_{j_{r',p}}(t_{i+1})}{\sigma_{r'}(t_{i+1})}. \tag{21}$$

Proof. See Appendix. ■

Remark 19 The formulæ in the above proposition can be understood by taking into account the formulæ in the previous subsection. For instance, Formula 14 shows what happens to the steadiness of a paper when it is judged by a reader: σ_p is increased by an amount equal to the score of the reader at time t_i ($s_r(t_i)$). Formula 15 shows how the score of the judged paper is obtained as the weighted mean between the previous score $s_p(t_i)$ and the expressed judgment $j_{r,p}$, where the weights are the steadiness of the paper $\sigma_p(t_i)$ (*i.e.*, the sum of the scores of the readers that read the paper so far) and the score of the reader expressing the judgment $s_r(t_i)$. The following formulæ can be understood in similar ways; the case of the score of previous readers (last formula) is the most complex and perhaps deserves some further explanations. The idea is that $s_{r'}(t_{i+1})$ is obtained by first subtracting what was (wrongly) added the last time that $s_{r'}$ was modified (let us call this time instant t^*) and then adding the correct value. The amount to be subtracted is easily obtained from score and steadiness of the paper at time t_i, since they have not changed during the time interval from the last modification time t^* to t_i.

Remark 20 The formulæ in the above proposition have been presented in a specific order, different from the order in the previous subsection (see Remark 18). This is necessary since some of the values at time t_{i+1} depend on other values at t_{i+1}, and these latter values have to be computed beforehand: $\sigma_p(t_{i+1})$ is used to compute $s_p(t_{i+1})$, $\sigma_r(t_{i+1})$, $s_r(t_{i+1})$, and $s_{r'}(t_{i+1})$; $s_p(t_{i+1})$ is used to compute $g_{j_{r,p}}(t_{i+1})$, that in turn is used to compute $s_r(t_{i+1})$ and $s_{r'}(t_{i+1})$; $\sigma_a(t_{i+1})$ is used to compute $s_a(t_{i+1})$; $\sigma_r(t_{i+1})$ is used to compute $s_r(t_{i+1})$; and $\sigma_{r'}(t_{i+1})$ is used to compute $s_{r'}(t_{i+1})$. Since there are no circularities in these dependencies, the formulæ could be rewritten to define the values at time t_{i+1} without using other values at time t_{i+1} (*e.g.*, $\sigma_r(t_{i+1}) = \sigma_r(t_i) + \sigma_p(t_i) + s_r(t_i)$), though this would lead to more complex formulæ.

A system implementing this mechanism must respect these constraints on the order of computation. The order (1.–8.) defined by the eight formulæ in the proposition is a correct one.

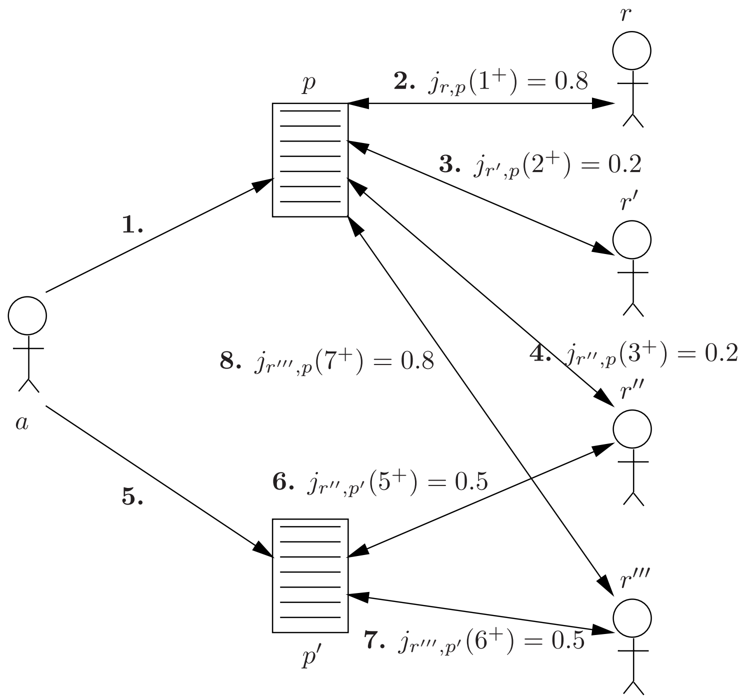

Figure 4: The bootstrap example in the text. Bold numbers correspond to the eight steps of the example. Double arrow lines are the expressed judgments, single arrow lines represent paper publications.

Examples and discussion

This section presents some examples of typical cases, discusses potential problems, and delineates the corresponding solutions. The aim of this section is twofold: the examples should help the reader to understand in more detail how the system works, and to better grasp at an intuitive level the meaning of the above presented formulæ; the problems and their suggested solutions should convince the reader that a system based on the idea presented in this paper may work in a satisfying way in the real world.

Bootstrap

Let us start with the empty system (no authors, papers, and readers) and let us assume that the following steps take place immediately after the corresponding time instants (see also Figure 4):[2]

1. At time $t = 0^+$, author a publishes p, the first paper in the system.

2. At $t = 1^+$, reader r reads p, expressing a 0.8 judgment ($j_{r,p}(1^+) = 0.8$).

3. At $t = 2^+$, reader r' reads p, expressing a 0.2 judgment ($j_{r',p}(2^+) = 0.2$).

4. At $t = 3^+$, reader r'' reads p, expressing a 0.2 judgment ($j_{r'',p}(3^+) = 0.2$).

5. At $t = 4$, author a publishes p', a second paper.

[2] t^- and t^+ denote the time instants immediately before and after t, respectively. This will help to avoid ambiguities.

	t	a		p		p'		r		r'		r''		r'''	
		s_a	σ_a	s_p	σ_p	$s_{p'}$	$\sigma_{p'}$	s_r	σ_r	$s_{r'}$	$\sigma_{r'}$	$s_{r''}$	$\sigma_{r''}$	$s_{r'''}$	$\sigma_{r'''}$
	0^+	0	0	0	0										
	1	0	0	0	0			ε	0						
$j_{r,p}(1^+) = 0.8 \Rightarrow$	2	0.8	ε	0.8	ε			1	ε						
	2^+	0.8	ε	0.8	ε			1	ε	ε	0				
$j_{r',p}(2^+) = 0.2 \Rightarrow$	3	0.5	2ε	0.5	2ε			0.45	2ε	0.45	2ε				
	3^+	0.5	2ε	0.5	2ε			0.45	2ε	0.45	2ε	ε	0		
$j_{r'',p}(3^+) = 0.2 \Rightarrow$	4	0.4	3ε	0.4	3ε			0.37	3ε	0.55	3ε	0.55	3ε		
	5	0.4	3ε	0.4	3ε	0	0	0.37	3ε	0.55	3ε	0.55	3ε		
$j_{r'',p'}(5^+) = 0.5 \Rightarrow$	6	0.5	0.55	0.4	3ε	0.5	0.55	0.37	3ε	0.55	3ε	1	0.55		
	6^+	0.5	0.55	0.4	3ε	0.5	0.55	0.37	3ε	0.55	3ε	1	0.55	ε	0
$j_{r''',p'}(6^+) = 0.5 \Rightarrow$	7	0.5	0.55	0.4	3ε	0.5	0.55	0.37	3ε	0.55	3ε	1	0.55	1	0.55
$j_{r''',p}(7^+) = 0.8 \Rightarrow$	8	0.69	1.55	0.8	1	0.5	0.55	1	1	0.23	1	0.50	1.55	1	1.55

Table 1: Evolution of scores and steadinesses for the bootstrap of the system.

6. At $t = 5^+$, reader r'' reads p', expressing a 0.5 judgment ($j_{r'',p'}(5^+) = 0.5$).

7. At $t = 6^+$, reader r''' reads p', expressing a 0.5 judgment ($j_{r''',p'}(6^+) = 0.5$).

8. At $t = 7^+$, reader r''' reads p, expressing a 0.8 judgment ($j_{r''',p}(7^+) = 0.8$).

Table 1 represents the evolution for all the eight steps. Let us analyze what happens. Paper p is judged by four readers. The first three (r, r', r'') have an ε score at the time of judgment, whereas the fourth (r''') has a score of 1. Therefore, the resulting score for p depends only on the judgment by r''', as well as the steadiness. Paper p' is judged by two readers. Both of them give a 0.5 judgment, but the resulting score and steadiness for p' depend on the judgment by r'' only, since it weights 0.55 against the ε weight of the judgment by r'''. Author a publishes two papers, p and p'. The score of a depends on both of them (p weighting a bit more, 1 *vs.* 0.55). The steadiness of a is the sum of the steadinesses of the two papers. Reader r expresses one judgment, with goodness (at time $t = 8$) 1. Notice the evolution of the score and steadiness of r in Table 1. Reader r' expresses one judgment, with goodness (at time $t = 8$) 0.23. Reader r'' expresses two judgments, a "bad" one, with goodness (at time $t = 8$) 0.23, and a "good" one with goodness (at time $t = 8$) 1. The final score respects this. Reader r''' expresses two judgments, both of them with goodness 1 at time $t = 8$.

Evolution of score and steadiness

Table 2 shows the evolution of score and steadiness of three papers p_1, p_2, and p_3 with a low initial score (0.1) and initial steadiness 1, 10, and 100 respectively, when n high (0.9) judgments are expressed by readers having a 0.5 score. The horizontal lines in the table visually emphasize the gaps in the values of n. As foreseen, high steadiness values imply slower score changes.

Bad author, good paper

Let us assume that an author a with score $s_a = 0.1$ and steadiness $\sigma_a = 2$ publishes a quite good paper p, and that all the readers recognize that the paper is good, so that readers $r_1, r_2, \ldots, r_n$, all of them with score 0.5, give a judgment of 0.9 (Figure 3 can be seen as representing this situation with $n = 3$). The first five

n	s_{p_1}	σ_{p_1}	s_{p_2}	σ_{p_2}	s_{p_3}	σ_{p_3}
0	0.1	1.0	0.1	10.0	0.1	100.0
1	0.367	1.5	0.138	10.5	0.104	100.5
2	0.5	2.0	0.173	11.0	0.108	101.0
3	0.58	2.5	0.204	11.5	0.112	101.5
4	0.633	3.0	0.233	12.0	0.116	102.0
5	0.671	3.5	0.26	12.5	0.12	102.5
6	0.7	4.0	0.285	13.0	0.123	103.0
7	0.722	4.5	0.307	13.5	0.127	103.5
8	0.74	5.0	0.329	14.0	0.131	104.0
9	0.755	5.5	0.348	14.5	0.134	104.5
10	0.767	6.0	0.367	15.0	0.138	105.0
20	0.827	11.0	0.5	20.0	0.173	110.0
30	0.85	16.0	0.58	25.0	0.204	115.0
40	0.862	21.0	0.633	30.0	0.233	120.0
50	0.869	26.0	0.671	35.0	0.26	125.0
60	0.874	31.0	0.7	40.0	0.285	130.0
70	0.878	36.0	0.722	45.0	0.307	135.0
80	0.88	41.0	0.74	50.0	0.329	140.0
90	0.883	46.0	0.755	55.0	0.348	145.0
100	0.884	51.0	0.767	60.0	0.367	150.0

Table 2: Evolution of score and steadiness of three papers having different initial steadiness.

columns in Table 3 represent what happens to score and steadiness values of both paper and author, depending on n: the score of the paper increases, as well as the score of the author (though more slowly), and the steadiness values.

If the initial steadiness of the author were $\sigma_a = 10$, or $\sigma_a = 100$ (in place of $\sigma_a = 2$), the increasing of the score of a would be lower. This is shown in the last four columns in Table 3.

Bad author, bad paper, late recognized

Let us assume that an author a publishes a paper p, which is not a good paper, but the first 10 readers $r_1 \dots r_{10}$ judge it as a good one, giving a judgment of 0.9, like in the previous example. After that, the readers $r_{11} \dots r_n$ judge p correctly, giving a 0.1 judgment. The first three columns of Table 4 represent what happens to score values of both paper and author, depending on n: for $n \in [0..10]$ the score values are equal to the ones in Table 3, but the following lines of the table show how paper and author's score decrease. The following two columns show the evolution of the score of the author if it had an initial steadiness of 10 and 100, respectively. Steadiness values are not shown since they increase, with the same values of Table 3. The meaning of last two columns is explained in the subsection "Lazy readers" below.

The temporal evolution the scores in Table 4 is shown in Figure 5: the first chart shows s_p (thickest line) and s_a, with the three different initial steadiness values (the higher the steadiness, the lower the s_a values); the second chart shows r_{10} (that initially increases and later decreases) and r_{11}.

n	s_p	σ_p	$s_a(2)$	$\sigma_a(2)$	$s_a(10)$	$\sigma_a(10)$	$s_a(100)$	$\sigma_a(100)$
0	0.0	0.0	0.1	2.0	0.1	10.0	0.1	100.0
1	0.9	0.5	0.26	2.5	0.14	10.5	0.1	100.5
2	0.9	1.0	0.37	3.0	0.17	11.0	0.11	101.0
3	0.9	1.5	0.44	3.5	0.2	11.5	0.11	101.5
4	0.9	2.0	0.5	4.0	0.23	12.0	0.12	102.0
5	0.9	2.5	0.54	4.5	0.26	12.5	0.12	102.5
6	0.9	3.0	0.58	5.0	0.28	13.0	0.12	103.0
7	0.9	3.5	0.61	5.5	0.31	13.5	0.13	103.5
8	0.9	4.0	0.63	6.0	0.33	14.0	0.13	104.0
9	0.9	4.5	0.65	6.5	0.35	14.5	0.13	104.5
10	0.9	5.0	0.67	7.0	0.37	15.0	0.14	105.0
20	0.9	10.0	0.77	12.0	0.5	20.0	0.17	110.0
30	0.9	15.0	0.81	17.0	0.58	25.0	0.2	115.0
40	0.9	20.0	0.83	22.0	0.63	30.0	0.23	120.0
50	0.9	25.0	0.84	27.0	0.67	35.0	0.26	125.0
60	0.9	30.0	0.85	32.0	0.7	40.0	0.28	130.0
70	0.9	35.0	0.86	37.0	0.72	45.0	0.31	135.0
80	0.9	40.0	0.86	42.0	0.74	50.0	0.33	140.0
90	0.9	45.0	0.87	47.0	0.75	55.0	0.35	145.0
100	0.9	50.0	0.87	52.0	0.77	60.0	0.37	150.0

Table 3: Authors with score 0.1 and steadinesses 2, 10, and 100 publishing a good paper.

n	s_p	$s_a(2)$	$s_a(10)$	$s_a(100)$	$s_{r_{10}}$	$s_{r_{11}}$
0	0.0	0.1	0.1	0.1	0.5	0.5
1	0.9	0.26	0.14	0.1	0.5	0.5
10	0.9	0.67	0.37	0.14	0.86	0.5
11	0.83	0.63	0.36	0.14	0.67	0.24
12	0.77	0.6	0.35	0.14	0.6	0.26
13	0.72	0.57	0.34	0.14	0.55	0.28
14	0.67	0.54	0.34	0.14	0.52	0.3
15	0.63	0.52	0.33	0.14	0.49	0.32
16	0.6	0.5	0.32	0.14	0.46	0.33
17	0.57	0.48	0.32	0.14	0.44	0.35
18	0.54	0.46	0.31	0.14	0.42	0.36
19	0.52	0.45	0.31	0.14	0.4	0.38
20	0.5	0.43	0.3	0.14	0.39	0.39
30	0.37	0.34	0.26	0.13	0.3	0.49
40	0.3	0.28	0.23	0.13	0.25	0.55
50	0.26	0.25	0.21	0.13	0.22	0.59
60	0.23	0.22	0.2	0.13	0.2	0.63
70	0.21	0.21	0.19	0.13	0.19	0.65
80	0.2	0.2	0.18	0.13	0.18	0.68
90	0.19	0.19	0.17	0.13	0.17	0.69
100	0.18	0.18	0.17	0.13	0.16	0.71

Table 4: A paper initially judged good and later judged bad.

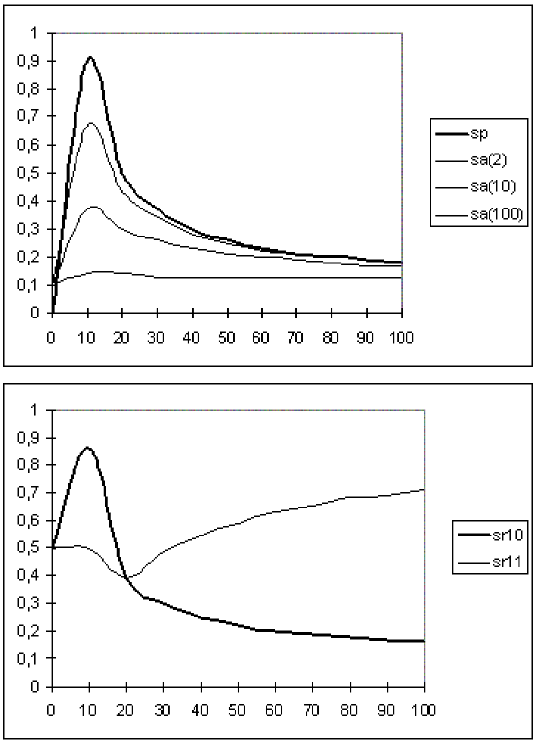

Figure 5: Graphical representation of the data in Table 4.

Lobbies

Subscribers might be tempted to behave in a malicious way to improve their scores. One possible malicious strategy is made up by *lobbies* of readers, *i.e.,* people that agree in mutually giving high scores with the aim of mutually increasing their own scores. Of course, a necessary condition for having a rewarding lobby is that the number of subscribers in the lobby is higher than the number of readers that, outside the lobby, make a sort of counter-lobby. Let us take a concrete example: a lobby with 6 subscribers in it and 60 out of it, all of them with initial score 0.5 and initial steadiness 10. If all the subscribers in the lobby publish one paper and judge each other paper with a 0.9 judgment, whereas all the subscribers in the counter-lobby express a 0.1 judgment on all the lobby papers, we have that: the average score of the lobby papers is 0.16, the average score of the lobby authors is 0.22, and the average score of the lobby readers is 0.27, whereas the average score of the counter-lobby readers is 0.92. The average of the scores is needed because the judgments are not expressed simultaneously, and the first reader will judge papers with lower steadiness.

Lazy readers

Another malicious behavior might be *laziness*: a *lazy reader* can simply confirm the previously expressed judgments, giving to each read paper a score equal to its actual score, with the aim of obtaining a high goodness.

At a first glance, this seems a serious problem. However, there are three answers to this objection. The first one is that it is not obvious at all if the strategy adopted by lazy readers is effective or not: it might lead to a low score for them, if the judgments by the following readers change the score of the paper. Therefore, the effectiveness depends on the future evolution of the score of the papers judged "lazily". Let us take again into account the example in the above subsection "Bad author, bad paper, late recognized" (see Table 4). The first ten readers, with the exception of the first one, are lazy, since they simply confirm the current score of the paper. But their score decreases. The score of reader r_{10} (the laziest one, since when she confirms the score of the paper, it has a higher steadiness) is shown in the last but one column of Table 4: it increases from 0.5 to 0.86 when r_{10} expresses the judgment, but it constantly decreases later, while $r_{11} \dots r_n$ express their judgments, reaching a 0.16 value when $n = 100$. On the contrary, $s_{r_{11}}$, the score of reader r_{11} (the less lazy one), decreases from 0.5 to 0.24 at the judgment expression time, but then increases until a 0.71 value, soon overtaking $s_{r_{10}}$ (when $n = 20$).

The second answer is that the mechanism might be improved by both giving higher scores to fast readers (those that first read the papers) and not showing the paper score for a period after its publication (for instance, until when its steadiness reaches a certain value).[3]

The third answer is that the mechanism might be improved even more substantially, by introducing a measure of the laziness of a reader, that allows a better estimate of the reliability of the score of a reader. This measure can be computed in a complete automatic way, *e.g.,* by computing the mean of the goodness values of the judgments at the time of judgment expression, weighted by the steadiness that the judged paper has at the time of judgment expression. In such a way, very good judgments on very steady papers wiil lead to higher laziness values.

Definition 8 (Laziness of a reader l_r) Given a reader r and the set of papers $P_r(t)$ judged by her before time t, we have $\forall p \in P_r(t)$ the judgment $j_{r,p}$ expressed by r on paper p at time $t_{r,p}$, the steadiness $\sigma_p(t_{r,p})$ that the paper p has when r expresses the judgment on it, and the goodness $g_{j_{r,p}}(t_{r,p})$ of the judgment $j_{r,p}$ calculated at time $t_{r,p}$. The laziness of r at time t is:

$$l_r(t) \quad = \quad \frac{\displaystyle\sum_{p \in P_r(t)} \sigma_p(t_{r,p}) \cdot g_{j_{r,p}}(t_{r,p})}{\displaystyle\sum_{p \in P_r(t)} \sigma_p(t_{r,p})}.$$

■

Remark 21 The laziness of a reader is a value in $[0, 1]$ (with 0 as the initial value). A lazy reader and an excellent one will perhaps have similar (high) scores, but the lazy reader will have a higher laziness and, therefore, will be easily singled out. We could also redefine the score of a reader, *e.g.,* as

$$\overline{s}_r(t) \quad = \quad \frac{s_r(t) + (1 - l_r(t))}{2},$$

to reward the readers with low laziness.

An updating formula for laziness, similar to the formulæ in the above subsection "Updating formulæ", can be easily defined. Also other measures might be used, for instance the quickness of a reader (on the basis of the average steadiness of the judged papers), or their activeness (more active readers will judge more papers), and used to have a more precise score for readers.

[3]Let us remark that adopting the second part only (not showing the paper score for a period after its publication) would have the opposite effect, since the lazy readers could rely on more stable paper scores.

More complex simulations

To understand what could happen in a real case, with different kinds of readers interacting among them, more complex simulations can (and have to) be run, using "software" readers that autonomously, and partly randomly, judge some "software" papers. This can be done in several ways, and it is not a simple task since many parameters have to be taken into account (distribution of readers and papers of various kinds, choice of papers to judge, and so on).

A discrete simulation: kinds of readers

One concrete example is the following. Let us take five categories of readers, differentiated on the basis of the strategy used to express their judgments: *random* readers (that express a random judgment); *constant* readers (expressing a 0.5 judgment); *lazy* readers (confirming the current score of the paper being judged); *worst* readers (if the current score of the paper is > 0.5 they give a 0.0 judgment, otherwise a 1.0 judgment); and *lazy-best* readers (like lazy readers, but they will express their judgments only at the end of the simulation).

Now let us create 60 papers and 100 readers randomly divided into five groups, roughly of the same size, corresponding to the five categories above. First, let us allow each reader in one of the first three categories to express her judgments on about 50% of the papers, randomly chosen and in random order. Then, to simulate *good* readers (that express correct judgments), let us allow all the worst readers to express their judgments (again on about 50% of the papers), followed by the lazy-best ones. For each run, about 3000 judgments are expressed.

By repeatedly running simulations of this kind, some insights can be derived, although the results are not conclusive at all. The worst readers get the lower score (usually around 0.2). The good (actually, lazy-best) readers obtain of course the highest scores (almost 1). The random readers obtain an average score of about 0.5, as expected (see Remark 11). The constant readers do unexpectedly well (often higher than 0.8), the reason being that both worst and random readers actually tend to move the scores of the papers toward 0.5. The scores of the lazy readers (about 0.7) are lower than the lazy-best readers, but usually also lower than the constant readers.

A continuous simulation: parameters of readers

The previous simulation is a "discrete" one, since each reader either belongs to a category or is out of it. To have a continuous, and more general, simulation, I have chosen the following seven parameters, all of them taking a value in the range $[0, 1]$, thus obtaining readers with different "gradations" of features: *goodness* (willingness to express a "correct" judgment—each paper has a theoretically correct s_p); *laziness* (willingness to confirm the current score of the paper); *activeness* (willingness to express many judgments); *selectiveness* (willingness to express judgments on highly rated papers only) *randomness* (willingness to express a random judgment); *quickness* (willingness to be among the first ones to express a judgment on each paper); and *constantness* (willingness to express a constant judgment).

Assuming that these parameters are uniformly distributed and independent, we get a population of readers where each reader has some random values for these parameters. A simulation can now be run by firstly allowing each reader to judge some of the papers, randomly chosen, and then measuring both the distributions of s_p and s_r and the correlations between the parameters and the final score of the readers.

Repeatedly running simulations with 300 papers and 500 readers equivalent results are obtained in each simulation. About 9500 judgments are expressed (about

32 per paper and 19 per reader), s_p has an average value of about 0.5, with a 0.3 minimum and a 0.7 maximum, and s_r has an average value of about 0.7, with a 0.5 minimum and a 0.9 maximum. The approximate correlation of the parameters with s_r is: 0.16 for goodness, 0.20 for laziness, 0.27 for both activeness and σ_r (of course, activeness and σ_r are highly correlated, with a 0.8 value), -0.5 for randomness, 0.27 for constantness, -0.1 for quickness, and 0 for selectiveness.

Some assumptions have been made in this simulations:

- 500 subscribers publishing 300 papers and expressing 9500 judgments correspond, *e.g.,* to a publication rate of 1 paper every 100 days and a judgment expression rate, for each subscriber, of 1 paper every 3 days. This is consistent with the publication and downloading rates in ArXiv (http://arXiv.org/cgi-bin/show_stats), that are about 1 publication every 133 days and 1 download every 5 days, for the average user.
- The parameters chosen are representative of users' features. Of course, there is no certainty that they are exhaustive.
- The distributions of the parameters are uniform and independent.

All these assumptions can be questioned, and they must be taken into account. Keeping that in mind, the conclusions that can be drawn are the following. On the average, laziness seems to bring to a slightly higher s_r than goodness. This should not be a big problem, anyway, since solutions have been proposed in the discussion about lazy readers in the previous subsection. As expected, constant readers are rewarded (the reason being the same just seen in the discrete simulation), whereas random readers get the lower correlation with s_r. Activeness seems to lead to a higher s_r, the reason being that most of the parameters (goodness, laziness, constantness) lead to express "good" judgments, whereas the only negative parameter is randomness. Therefore, the readers distribution is biased toward giving good judgments. Selectiveness seems to have no effect at all, and, slightly surprising, quickness is not extremely harmful.

A general remark about the issue of malicious behavior is that it is not simple at all to understand if a malicious behavior is rewarding or not. Lobbies are a good example for demonstrating this. If a lobby is successful, subscribers in the lobby will tend to have high scores. Therefore, the number of subscribers reading and judging their papers will probably increase, with a higher probability that many readers will express a (correct) judgment, thus decreasing the scores of the subscribers in the lobby. In other words, the mechanism presented in this paper involves real people and, like all biological and social systems, it is likely to exhibit unexpected behavior: see, *e.g.,* (Dawkins, 1976; Ridley, 1997) for interesting discussions of these phenomena.

The general lesson that can be learned is just that readers play a game that, because of the interactions among them, is very complex, and further simulations need to be done.

Conclusions and future developments

This paper describes an electronic scholarly journal capable of a kind of quality control which is not based on peer review but on reader's judgments. As already emphasized, this does not mean that peer review is considered an inadequate solution. On the contrary, the proposal follows the trend of increasing the number of referees: from the single editor in the 17th century, to a few reviewers in the last century, to all the readers acting as referees. Moreover, the quality control proposed here can be used in joint action with peer review: either after it, with peer reviewed papers judged by the readers, or before it, with institutional referees judging some papers only. In the first case, more judgments on papers, authors, and readers (and referees too), are available; in the second case, the papers undergoing peer review can be selected also on the basis of readers' judgments.

Generally speaking, this proposal can be seen as an improvement of the dissemination of scholarly information through on-line journals. More specifically, it can be seen as an improvement of the democratic journals proposed in (LaPorte et al., 1995; Nadasdy, 1997; Rogers and Hurt, 1990; Stodolsky, 1990; Varian, 1998), and of collaborative information retrieval and filtering (Karamuftuoglu, 1998), since it allows to distinguish among "good" and "bad" collaborators. Also the well known impact factor mechanism (Garfield, 1972) appears to be a quite poor measure if compared with this proposal: the impact factor of a paper is something similar to (actually simpler than) the steadiness of a paper σ_p, whereas the impact factor of an author is something similar to (simpler than) the steadiness of an author σ_a. And the steadiness, once seen in the whole framework proposed here, is a poor quality measure indeed.

From a socio-economical viewpoint, the standard practice, in which peer reviewers work for free (Harnad, 2001a), is improved by the mechanism proposed in this paper, in which the increased reputation of readers/referees is an explicit reward. This, besides being more fair, might help to solve the problem of lack of referees. Let us remark that, in the scenario proposed in this paper, a reader might choose to read a paper for various reasons: because of the paper score, of the paper topic, or of some commentaries on the paper. The paper might be found using a search engine, or similar tools. Even if a paper has a low score because it is flawed, a reader might read it because it has a very good literature survey, or, vice versa, a paper with a low score because it is written in a very bad way might contain a novel and interesting result.

This proposal is not free of problems itself. In general, one may wonder if democracy is a good approach to scholarly knowledge dissemination. Of course, it is difficult to have an objective opinion on that: it could be appropriate, appropriate in some fields only, or not appropriate at all. I believe that only by further studies and experiments we can find an objective answer. However, it has to be emphasized that the mechanism proposed here is different from democracy, since different readers have different importance.

Problems that fit into another category are the malicious strategies that readers and authors can adopt to improve their scores. Two strategies of this kind (lobbies and lazy readers) are discussed in the "Examples and discussion" section, where it is shown that is not simple to understand whether a strategy is rewarding or not, and some improvements to the system, to limit these malicious behaviors, are suggested. A more radical approach would be to add some "supervisor" (partly software and partly human) that monitors users' behavior and intervenes when appropriate. Supervisors would be different from referees in scholarly journals: the former would intervene only if some incorrect behavior is found, with a "punishment" approach, whereas the latter are continuously filtering the information to be published.

If this will not be enough, a last resort is to show only some of the scores. For instance, by keeping secret the scores of authors and readers we obtain a mechanism that measures the quality of papers, and can be used to improve the performances of current information retrieval systems; by keeping secret the score of papers, we obtain a mechanism that measures in an objective way the quality of researchers, an alternative to the extremely criticized impact factor; and by not showing the steadiness values of papers, laziness is hindered.

Of course there are technical difficulties too, *e.g.,* the identification of subscribers to avoid malicious tampering and spoofing, or the huge amount of storage needed for recording the papers, the subscribers' data, and the history of expressed judgments.

Appropriate database, cryptography, and security technologies must be adopted. If the complexity of computing score and steadiness values turned out to be too high, some approximations will have to be found.

It is also easy to see some mandatory *improvements*:

- To deal with papers with more than one author. This is simple: the judgments on the paper cause a modification of all the authors' scores, weighted on the basis of the importance of the contribution of each author to the paper.
- To have more scores, both for subscribers (authors and readers) and papers: comprehensibility, technical soundness, originality, and so on. In this way, a more detailed evaluation is available. If just one single number is needed, a weighted mean of all the scores of a subscriber (or a paper) can be used.
- To have more than one journal, with different acceptance thresholds: a paper is published either if its author has a score higher than the threshold, or after a peer review. With this approach, less brilliant researchers would probably subscribe to lower rated journals, and first class journals would accept only very good papers. As mentioned above, the mechanism presented in this paper can also be used as a complement of, instead of a replacement to, peer review: the initial score of a paper can be given by a standard peer review, thus allowing an author with a low score to submit her paper to an higher rated journal; or peer review might take place only after reader's judgments, *e.g.,* for controversial papers.
- To allow the subscribers know why their score is decreasing, *i.e.,* which "wrong" judgment, or paper, causes that, and eventually let them revise their judgment or withdraw their paper.
- To introduce some sort of *rent* function, for decreasing the score of subscribers that are inactive for long periods of time.
- To allow the readers to express, besides the numerical judgments, also a free text *commentary* on the paper. The commentary can then be considered as a paper itself, and judged by other readers, but it is linked to the paper it comments, and the score of the commentary can affect the score of the paper.
- To simplify the updating of previous readers' scores. With the above proposed approach, all the values assumed by scores and steadiness and all the expressed judgments on the paper being judged need to be recorded and easily accessible for updating the scores of the previous readers. This might be simplified, even by means of some approximations; for instance, one might take into account a limited history only, disregarding what happened a long time ago.

Finally, the future developments of this research are sketched. The implementation of a software simulator of the electronic journal proposed here is just finished and I am currently running some simulations to understand the general behavior of the system and the features of the system, *e.g.,* if the initial conditions—*i.e.,* the number of initial readers, their score, and so on—are a critical factor for having a stable system. I also plan to use some mathematical models and techniques (game theory seems the most adequate) for formally studying the behavior of the system and for studying other similar approaches. For example, I might use a bet-like approach, in which each reader has some money and bets on the goodness—or badness—of some papers.

These theoretical and experimental activities will allow to verify that the behavior of the system is correct and consistent and to choose in a more reliable way among the possible formulæ and parameters. After that, the software for the complete system will be implemented, tested, and evaluated. An ideal environment for these experiments is a repository of preprints, like ArXiv (http://ArXiv.org) or ResearchIndex (http://researchindex.org). I plan to perform some laboratory experiments (both with simulated papers, authors, and readers and by using data logged from real users of preprints repositories) and some real life experiments (involving real users). Also social issues need to be enquired, by means of social sciences approaches and methodologies.

These theoretical and experimental activities will also allow to determine the relations between the values of some parameters and the real situation: indeed, it is likely that the values of parameters depend on figures like the number of subscribers, the rate of papers publishing, the rate of judgments expression, and so on. These dependencies must be singled out and, if the above presented formulæ turn out to be inadequate, new ones need to be proposed and evaluated.

Appendix. Proof of Proposition 1

Each of the eight formulæ is proved independently.

1. The set of readers that have expressed a judgment on p before time t_{i+1} is $R_p(t_{i+1}) = R_p(t_i) \cup \{r\}$. Using this equality and Formula 2 we have the following equality chain:

$$\sigma_p(t_{i+1}) = \sum_{r_i \in R_p(t_{i+1})} s_{r_i}(t_{r_i,p}) = \sum_{r_i \in R_p(t_i)} s_{r_i}(t_{r_i,p}) + s_r(t_i) = \sigma_p(t_i) + s_r(t_i).$$

Actually, Formula 2 has needed a variable substitution to be used in the previous equality chain, since r cannot refer to both the reader expressing the judgment (as in Proposition 1) and all the elements of $R_p(t)$ (as in Formula 2). Therefore, Formula 2 has been used rewritten as:

$$\sigma_p(t) \quad = \sum_{r_i \in R_p(t)} s_{r_i}(t_{r_i,p}).$$

Similar variable substitutions are done in the following for other formulæ.

2. By Formula 3 we have

$$s_p(t_{i+1}) \quad = \quad \frac{\displaystyle\sum_{r_i \in R_p(t_{i+1})} s_{r_i}(t_{r_i,p}) \cdot j_{r_i,p}}{\sigma_p(t_{i+1})} =$$

that, by observing that $R_p(t_{i+1}) = R_p(t_i) \cup \{r\}$ (because r is added to set R_p at time t_i), that $t_{r,p} = t_i$, and by splitting the summation, becomes

$$= \quad \frac{\displaystyle\sum_{r_i \in R_p(t_i)} s_{r_i}(t_{r_i,p}) \cdot j_{r_i,p} + s_r(t_i) \cdot j_{r,p}(t_i)}{\sigma_p(t_{i+1})}$$

that, using again Formula 3 on the summation, is the right part of Formula 15.

3. By Formula 6

$$\sigma_a(t_{i+1}) \quad = \sum_{p_i \in P_a(t_{i+1})} \sigma_{p_i}(t_{i+1}) =$$

that, since $P_a(t_{i+1}) = P_a(t_i)$ (the set of papers published by a at time t_{i+1} is the same as the set at time t_i), and by splitting the summation extracting the term concerning p, becomes

$$= \sum_{p_i \in P_a(t_i) \setminus \{p\}} \sigma_{p_i}(t_{i+1}) + \sigma_p(t_{i+1}) =$$

that, since $\forall p_i \neq p(\sigma_{p_i}(t_{i+1}) = \sigma_{p_i}(t_i))$ and by Formula 14 is

$$= \sum_{p_i \in P_a(t_i) \setminus \{p\}} \sigma_{p_i}(t_i) + \sigma_p(t_i) + s_r(t_i) =$$

that, by re-inserting the p term in the summation and then using Formula 14 again is

$$= \quad \sigma_a(t_i) + s_r(t_i).$$

4. From Formula 8 we have

$$s_a(t_{i+1}) \cdot \sigma_a(t_{i+1}) \quad = \sum_{p_i \in P_a(t_{i+1})} \sigma_{p_i}(t_{i+1}) \cdot s_{p_i}(t_{i+1}) =$$

(where $P_a(t_{i+1})$ is the set of papers published by a before time t_{i+1}) that, by splitting the summation and by adding and subtracting the same amount becomes

$$= \sum_{p_i \in P_a(t_{i+1}) \setminus \{p\}} \sigma_{p_i}(t_{i+1}) \cdot s_{p_i}(t_{i+1}) + \sigma_p(t_{i+1}) \cdot s_p(t_{i+1}) + \sigma_p(t_i) \cdot s_p(t_i) - \sigma_p(t_i) \cdot s_p(t_i) = .$$

Now, note that $P_a(t_{i+1}) = P_a(t_i)$ and that $\forall p_i \neq p$ we have $\sigma_{p_i}(t_{i+1}) = \sigma_{p_i}(t_i)$ and $s_{p_i}(t_{i+1}) = s_{p_i}(t_i)$ (since the judgment is on p, the other papers are not affected). We therefore obtain (by rearranging the expression too)

$$= \sum_{p_i \in P_a(t_i) \setminus \{p\}} \sigma_{p_i}(t_i) \cdot s_{p_i}(t_i) + \sigma_p(t_i) \cdot s_p(t_i) + \sigma_p(t_{i+1}) \cdot s_p(t_{i+1}) - \sigma_p(t_i) \cdot s_p(t_i) =$$

that, by re-inserting the p terms in the summation and by Formula 15, is

$$= \sum_{p_i \in P_a(t_i)} \sigma_{p_i}(t_i) \cdot s_{p_i}(t_i) + s_r(t_i) \cdot j_{r,p} =$$

that, by Formula 8, is

$$= s_a(t_i) \cdot \sigma_a(t_i) + s_r(t_i) \cdot j_{r,p}.$$

5. Using Formula 12, since $P_r(t_{i+1}) = P_r(t_i) \cup \{p\}$, and since $\forall p_i \neq p$ $(\sigma_{p_i}(t_{i+1}) = \sigma_{p_i}(t_i))$, we have:

$$\sigma_r(t_{i+1}) = \sum_{p_i \in P_r(t_{i+1})} \sigma_{p_i}(t_{i+1}) = \sum_{p_i \in P_r(t_i)} \sigma_{p_i}(t_{i+1}) + \sigma_p(t_{i+1}) = \sigma_r(t_i) + \sigma_p(t_{i+1}).$$

6. By Formula 13

$$s_r(t_{i+1}) \quad = \quad \frac{\sum_{p_i \in P_r(t_{i+1})} \sigma_{p_i}(t_{i+1}) \cdot g_{j_{r,p_i}}(t_{i+1})}{\sigma_r(t_{i+1})} =$$

that, by the equality $P_r(t_{i+1}) = P_r(t_i) \cup \{p\}$, by extracting the p terms from the summation, and by $t_{r,p} = t_i$, is

$$= \frac{\sum_{p_i \in P_r(t_i)} \sigma_{p_i}(t_{i+1}) \cdot g_{j_{r,p_i}}(t_{i+1}) + \sigma_p(t_{i+1}) \cdot g_{j_{r,p}}(t_{i+1})}{\sigma_r(t_{i+1})} =$$

that, by observing that $\forall p_i \neq p$ we have $\sigma_{p_i}(t_{i+1}) = \sigma_{p_i}(t_i)$, $s_{p_i}(t_{i+1}) = s_{p_i}(t_i)$ and therefore $g_{j_{r,p_i}}(t_{i+1}) = g_{j_{r,p_i}}(t_i)$, becomes

$$\frac{\sum_{p_i \in P_r(t_i)} \sigma_{p_i}(t_i) \cdot g_{j_{r,p_i}}(t_i) + \sigma_p(t_{i+1}) \cdot g_{j_{r,p}}(t_{i+1})}{\sigma_r(t_{i+1})}$$

that, by using again Formula 13 is the right part of Formula 19.

7. By Formula 12

$$\sigma_{r'}(t_{i+1}) \quad = \sum_{p_i \in P_{r'}(t_{i+1})} \sigma_{p_i}(t_{i+1}) =$$

that, since $P_{r'}(t_{i+1}) = P_{r'}(t_i)$ (the set of papers judged by r' does not change since it is r that is expressing the judgment), and by extracting the p term from the summation, becomes

$$= \sum_{p_i \in P_{r'}(t_i) \setminus \{p\}} \sigma_{p_i}(t_{i+1}) + \sigma_p(t_{i+1}) = .$$

Now, $\forall p_i \in P_{r'}(t_i) \setminus \{p\}$ it holds that $\sigma_{p_i}(t_{i+1}) = \sigma_{p_i}(t_i)$ (because only p changes), and, by Formula 14, $\sigma_p(t_{i+1}) = \sigma_p(t_i) + s_r(t_i)$; therefore, we can write:

$$= \sum_{p_i \in P_{r'}(t_i) \setminus \{p\}} \sigma_{p_i}(t_i) + \sigma_p(t_i) + s_r(t_i)$$

that, by re-inserting the p term in the summation and by using Formula 12 again, is the right part of Formula 20.

8. From Formula 13 we have

$$\sigma_{r'}(t_{i+1}) \cdot s_{r'}(t_{i+1}) \quad = \sum_{p_i \in P_{r'}(t_{i+1})} \sigma_{p_i}(t_{i+1}) \cdot g_{j_{r',p_i}}(t_{i+1}) =$$

that by splitting the summation in the usual way becomes

$$= \sum_{p_i \in P_{r'}(t_{i+1}) \setminus \{p\}} \sigma_{p_i}(t_{i+1}) \cdot g_{j_{r',p_i}}(t_{i+1}) + \sigma_p(t_{i+1}) \cdot g_{j_{r',p}}(t_{i+1}) = .$$

Now, let us note that: $P_{r'}(t_{i+1}) = P_{r'}(t_i)$ (the set of papers judged by r' does not change since it is r that expresses the judgment); $\forall p_i \in P_{r'}(t_i) \setminus \{p\}$ $(\sigma_{p_i}(t_{i+1}) = \sigma_{p_i}(t_i))$; and the goodnesses of the judgments by r' on the other papers does not change (because $\forall p_i \in P_{r'}(t_i) \setminus \{p\}$ $(s_{p_i}(t_{i+1}) = s_{p_i}(t_i))$). Therefore, by adding and subtracting the same amount, we have

$$\begin{aligned} = & \sum_{p_i \in P_{r'}(t_i) \setminus \{p\}} \sigma_{p_i}(t_i) \cdot g_{j_{r',p_i}}(t_i) + \sigma_p(t_{i+1}) \cdot g_{j_{r',p}}(t_{i+1}) \\ & + \sigma_p(t_i) \cdot g_{j_{r',p}}(t_i) - \sigma_p(t_i) \cdot g_{j_{r',p}}(t_i) = \end{aligned}$$

that, by inserting the added p term in the summation, becomes

$$= \sum_{p_i \in P_{r'}(t_i)} \sigma_{p_i}(t_i) \cdot g_{j_{r',p_i}}(t_i) + \sigma_p(t_{i+1}) \cdot g_{j_{r',p}}(t_{i+1}) - \sigma_p(t_i) \cdot g_{j_{r',p}}(t_i)$$

that, by Formula 13 again is the numerator of Formula 21. ■

Acknowledgments

I have been discussing the idea in this paper with several people, who read drafts of this paper, provided useful suggestions, and discussed this issue. I would like to thank Paolo Coppola, Vincenzo Della Mea, Massimo Di Fant, Luca Di Gaspero, Marco Fabbrichesi, Andrea Fusiello, Stefania Gentili, Marco Mizzaro, Carla Piazza, Ivan Scagnetto, and Walter Vanzella. Special thanks to Paolo Zandegiacomo Riziò, who implemented the software simulator used to run the simulation examples. I also acknowledge the referees, especially one of them, who greatly helped in improving the paper. This research has been partially supported by the 5th Framework EU Project TIPS (IST-1999-10419).

References

Apt, K. R. (2001a). One more revolution to make: Free scientific publishing. *Communications of ACM*, 44(5):25–28.

Apt, K. R. (2001b). Towards free access to scientific literature. *Nieuw Archief voor Wiskunde*, 5/2(3):251–255.

Arms, W. Y. (2002). What are the alternatives to peer review? quality control in scholarly publishing on the web. *JEP: the Journal of Electronic Publishing*, 8(1). http://www.press.umich.edu/jep/08-01/arms.html.

Bailey, Jr., C. W. (2002). Scholarly electronic publishing bibliography—Version 45. Web document, visited 21/10/2002. http://info.lib.uh.edu/sepb/sepb.html.

Borgman, C. (2000). *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. MIT Press, Cambridge, MA.

Borgman, C. and Furner, J. (2002). Scholarly communication and bibliometrics. In Cronin, B., editor, *Annual Review of Information Science and Technology*, volume 36. Information Today Inc. ISBN: 1-57387-131-1.

Brown, C. (2001). The e-volution of preprints in the scholarly communication of physicists and astronomers. *Journal of the American Society for Information Science and Technology*, 52(3):187–200.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press, Oxford.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178:471–479. http://www.garfield.library.upenn.edu/essays/V1p527y1962-73.pdf.

Ginsparg, P. (1994). First steps towards electronic research communication. *Computers in Physics*, 8(4):390–396. http://arxiv.org/blurb/blurb.ps.gz.

Ginsparg, P. (1996). Winners and losers in the global research village. Invited contribution for Conference held at UNESCO HQ, Paris, 21 Feb 1996. http://arxiv.org/blurb/pg96unesco.html.

Harnad, S. (1985). Rational disagreement in peer review. *Science, Technology and Human Values*, 10:55–62. http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad85.peerev.htm.

Harnad, S. (1990). Scholarly skywriting and the prepublication continuum of scientific inquiry. *Psychological Science*, 1:342–343. Reprinted in *Current Contents* 45: 9-13, November 1991. http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html.

Harnad, S. (1995). The PostGutenberg Galaxy: How to get there from here. *Information Society*, 11(4):285–292. http://cogprints.ecs.soton.ac.uk/archive/00001689/index.html.

Harnad, S. (1996). Implementing peer review on the net: Scientific quality control in scholarly electronic journals. In Peek, R. and Newby, G., editors, *Scholarly Publication: The Electronic Frontier*, pages 103–108. MIT Press, Cambridge MA. http://www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad96.peer.review.html.

Harnad, S. (1997). Learned inquiry and the net: The role of peer review, peer commentary and copyright. *Antiquity*, 71:1042–1048. ftp://ftp.princeton.edu/pub/harnad/Harnad/HTML/harnad97.antiquity.html.

Harnad, S. (2000). The invisible hand of peer review. *Exploit Interactive*, 5. http://www.exploit-lib.org/issue5/peer-review/. A shorter version is published on *Nature [online]*, Nov. 1998, http://helix.nature.com/webmatters/invisible/invisible.html.

Harnad, S. (2001a). Research access, impact and assessment. *Times Higher Education Supplement*, 1487:16. http://www.cogsci.soton.ac.uk/~harnad/Tp/thes1.html.

Harnad, S. (2001b). The self-archiving initiative. *Nature*, 410:1024–1025. http://www.cogsci.soton.ac.uk/~harnad/Tp/nature4.htm.

Harter, S. P. (1998). Scholarly communication and electronic journals: An impact study. *Journal of the American Society for Information Science*, 49(6):507–516.

Harter, S. P. and Ford, C. E. (2000). Web-based analyses of e-journal impact: Approaches, problems, and issues. *Journal of the American Society for Information Science*, 51(13):1159–1176.

Holoviak, J. and Seitter, K. L. (1997). Earth Interactions—Trascending the limitations of the printed page. *JEP the Journal of Electronic Publishing*, 3(1). ISSN 1080-2711; http://www.press.umich.edu/jep/03-01/EI.html.

Karamuftuoglu, M. (1998). Collaborative information retrieval: Toward a social informatics view of IR interaction. *Journal of the American Society for Information Science*, 49(12):1070–1080.

Kling, R. and Callahan, E. (2003). Electronic journals, the internet, and scholarly communication. In Cronin, B., editor, *Annual Review of Information Science and Technology*, volume 37. Information Today Inc. To appear. ISBN: 1-57387-154-0.

Kling, R. and Covi, L. (1995). Electronic journals and legitimate media in the system of scholarly communication. *The Information Society*, 11(4):261–271. http://www.slis.indiana.edu/TIS/articles/klingej2.html.

Kling, R. and McKim, G. (1997). A typology for electronic-journals: Characterizing scholarly journals by their distribution forms. CSI Working Paper No. WP-97-07. http://www.slis.indiana.edu/csi/Wp/wp97-07B.html.

Kling, R. and McKim, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science*, 51(14):1306–1320.

Kling, R., McKim, G., and King, A. (2001). A bit more to IT: Scholarly communication forums as socio-technical interaction networks. *Journal of the American Society for Information Science and Technology*. Working Draft 4.0a, to be submitted, http://www.slis.indiana.edu/csi/wp01-02.html.

Kling, R., Spector, L., and McKim, G. (2002). Locally controlled scholarly publishing via the internet: The Guild Model. *JEP: the Journal of Electronic Publishing*, 8(1). ISSN: 1080-2711. http://www.press.umich.edu/jep/08-01/kling.html.

LaPorte, R., Marler, E., Akazawa, S., Sauer, F., Gamboa, C., Shenton, C., Glosser, C., Villasenor, C., and Maclure, M. (1995). The death of the biomedical journal. *British Medical Journal*, 31(6991):1387–1390.

Lesk, M. (1997). *Practical Digital Libraries: Books, Bytes, and Bucks*. Morgan Kaufmann, San Francisco, CA.

McKiernan, G. (2002). E-profile web-based journal manuscript management and peer-review software and systems. *Library Hi Tech News*, 19(7). http://www.emeraldinsight.com/vl=1/cl=6/nw=1/rpsv/cw/mcb/07419058/v19n7/s7003/p3l.html.

Mizzaro, S. (1999). A concrete proposal for an automatically refereed scholarly electronic journal. Research Report UDMI/07/99, Department of Mathematics and Computer Science, University of Udine, Via delle Scienze, 206 — Loc. Rizzi — Udine, Italy. Presented at a Workshop at ECDL'99 Conference, Paris.

Mizzaro, S. (2000). An automatically refereed scholarly electronic journal. In Bavec, C., Gams, M., Rajkovič, V., Rozman, I., and Györkös, J., editors, *Information Society 2000 — Proceedings of the Third International Conference*, pages 121–126, Ljubljana, Slovenia. ISBN: 961-6303-26-0.

Mizzaro, S. and Zandegiacomo Riziò, P. (2000). An automatically refereed scholarly electronic journal: Formal specifications. *Informatica - An International Journal of Computing and Informatics (ISSN 0359-5596)*, 24(4):431–438. The Slovene Society Informatika, Ljubljana, Slovenia.

Nadasdy, Z. (1997). A truly all-electronic journal: Let democracy replace peer review. *JEP the Journal of Electronic Publishing*, 3(1). ISSN 1080-2711; http://www.press.umich.edu/jep/03-01/EJCBS.html.

Odlyzko, A. M. (1995). Tragic loss or good riddance? The impending demise of traditional scholarly journals. *International Journal of Human-Computer Studies*, 42:71–122.

Odlyzko, A. M. (2002). The rapid evolution of scholarly communication. *Learned Publishing*, 15(1):7–19. Also to appear in *Bits and Bucks: Economics and Usage of Digital Collections*, W. Lougee and J. MacKie-Mason, eds., MIT Press, 2002.

Okerson, A. (1991). The refereed electronic journal: What, whence, and when. *Public Access Computer Systems Review*, 2(1):5–24.

Peek, R., Pomerantz, J., and Pailing, S. (1998). The traditional scholarly journal publishers legitimize the web. *Journal of the American Society for Information Science*, 49(11):983–989.

Resh, V. H. (1998). Science and communication: An author/editor/user's perspective on the transition from paper to electronic publishing. *Issues in Science and Technology Librarianship*, 19. ISSN 1092-1206; http://www.library.ucsb.edu/istl/98-summer/article3.html.

Ridley, M. (1997). *The Origins of Virtue*. Penguin Books, London.

Rogers, S. J. and Hurt, C. S. (1990). How scholarly publications should work in the 21st century. *College and Research Libraries*, 51(1):5–67.

Service, R. F. (2002). Scientific misconduct: Bell labs fires star physicist found guilty of forging data. *Science*, 298:30–31.

Stodolsky, D. S. (1990). Consensus journals: Invitational journals based upon peer consensus. *Psycoloquy*, 1(15). Also appeared in: Datalogiske Skrifter (Writings on Computer Science), 29, 1990, Roskilde University Centre, Institute of Geography, Socioeconomic Analysis, and Computer Science (ISSN 0109-9779-29).

Varian, H. R. (1998). The future of electronic journals. *JEP the Journal of Electronic Publishing*, 4(1). ISSN 1080-2711; http://www.press.umich.edu/jep/04-01/varian.html.

Weller, A. C. (2000). Editorial peer review for electronic journals: Current issues and emerging models. *Journal of the American Society for Information Science*, 51(14):1328–1333.

Wells, A. (1999). Exploring the development of the independent, electronic scholarly journal. Master's thesis, MSc in Information Management. http://panizzi.shef.ac.uk/elecdiss/edl0001/.

Zhang, Y. (2001). Scholarly use of internet-based electronic resources. *Journal of the American Society for Information Science*, 52(8):628–654.