

Concepts and Guidelines for an “Open Publication Revision” System

Horea Christian^{1,2}

1 Abstract

Currently, journals use peer review of context-data-evaluation-discussion monoliths to make a go/no-go choice regarding publication. We propose a non-binary model which harnesses the power of open collaboration, immediate publication, and transparent revision. We name this model, accordingly, Open Publication Revision (OPR). We believe that the worthiness of findings is never static, and can only be interpreted in light of their ability to support further developments. The delineated system is tailored to add value even to the least promising publications, rather than derive value for itself from the most promising ones. We believe our system would provide an immediate service to the scientific community by making otherwise lost findings viable.

2 Background

Peer review is a widely acclaimed procedure for quality control, the principle of which has circulated in human societies for millennia [22]. Its rationale is both solid and simple: Only people familiar with a certain part of the knowledge network can decide whether and how a new part fits in.

The exact implementation of peer review in the context of publishing has changed in accordance with the technological possibilities of the time [22], the most recent notable changes including online publishing and review, and the more controversial open access publishing[24, 17] and open peer review. While open access seems to have gained a foothold in the scientific community (notably demonstrated by the recent success of the Public Library of Science journals), the same cannot be said of open peer review.

Many opinion and research articles concern themselves with the benefits[14, 15] and the applicability[25, 26] of

open peer review. It should be noted, however, that this is but the most prominent of many incentives to change peer review. A landmark of the discussion of modern publishing was *Nature*’s 2006 series on the trial and debate of peer review [4]. The debate illustrates that the scientific community is bursting with ideas regarding the future of publishing. Some of the articles also show how novel systems have already been implemented [20, 21, 11], and how one in particular, arXiv[1], has even become a standard in a number of fields.

The article series shows that seasoned members of the scientific publishing community are vividly aware of a great many shortcomings of current knowledge mediation in general and peer review in particular. Such disappointments, complemented by many others, are also shared by less established authors [15].

3 Issues

From the vast pool of opinions in the relevant literature, we have selected a number of knowledge mediation issues which we consider most pressing, and to which we presume to offer pertinent solutions. For ease of overview we use “peer review” to refer to the predominant current implementation thereof (including such aspects as there being 3 reviewers per article).

1. Peer review is time-consuming. This can cause excessive delay of rapidly outdated findings [20], or conflicts between authors publishing similar studies.
2. Peer review commonly imposes *a priori* gauged “interest” as a filter to publishing [6]. This may encourage the formation of dogmata [5]. Anecdotally, the unreliability of such foresight has lead to the rejection of Nobel-level articles [3].
3. Reviewers are expected to contribute without receiving compensation for their work - but risk losing face if consistently refusing. This lack of positive incentive for review makes it a major bottleneck in publishing, and a burden for researchers [11].

4. Peer review creates hierarchic value for journals [10, 27], rather than for articles. There is little incentive for peer review coordinators to make an article as good as it possibly can be at a great expense of effort; the incentive is rather to select the probably most citable articles with minimum effort.
5. Peer review is conducted by the most reputable of scientists, who often lack both time and hands-on experience. Practical expertise of technicians and plentiful time resources of students are utterly disregarded for the process [12].
6. The peer review of their own work is often opaque to researchers, but even more so to the public [7].
2. By removing any filter and assigning the task of evoking interest to the evolving article; articles would fairly compete for citation and furthering of their findings.
3. Revisions are directly attributed to their authors, so there is a strong incentive to invest time in other people's work (authors of significant revisions become co-authors of future versions). Researchers can choose their own personal mix of review and primary publishing work, and receive true effort-proportional recognition.
4. Value is assigned to articles directly, and the chief incentive for all who revise is to make their work as useful as possible.
5. Everyone is able to revise articles and contribute their strongest personal expertise to the knowledge network.
6. OPR as a system would be completely transparent - to members of the public, journalists, and policy makers.

4 Methods

4.1 Git

The main method we use to tackle these issues is electronic revision control - a concept most widespread in software development. This powerful system allows authors of scientific articles to benefit from the same features which enable programmers around the world to collaboratively and asynchronously build functioning code online. The history of revision control exceeds the scope of this article; though it should be noted that more basic forms of revision control (such as book editions) already have a strong tradition in knowledge management.

Of the modern implementations of revision control we have chosen to use Git[23] - owing to its distributed and decentralized nature, and support for non-linear development.

For online publishing we encourage using GitHub[8] - an internet platform for hosting Git repositories. Owing to the nature of Git, however, users will be able to remain integrated in the network regardless where they choose to host.

4.2 LaTeX

Our method of choice for keeping information organized within an article (while supporting easy human readability and a pleasurable visual experience) is the LaTeX[13] document markup language. LaTeX is already a standard for a great number of publishers; and implementing it would potentially facilitate collaboration with journals.

5 Results

5.1 Overview

In brief, we believe the issues we address would be resolved in the following manner:

1. Immediate publication allows anyone who is ready to sacrifice polish for rapid access to data the means to do so. Articles would receive revision time proportional to how many people they motivate; and authors publishing similar studies would rather collaborate than try to obscure each other.

5.2 Workflow

The aforementioned emergent properties owe to the following workflow:

- One or multiple initial authors publish an article whenever they feel they should.
- Researchers interested in the topic find the article (via the web, or via advertisement from the original authors), and read it.
- Depending on how reliable they find the work and interesting the finding, other researches may cite, further recommend, or *fork* the article into their own repository and revise it.
- Authors of revisions would be interested in furthering the work which is now also their own, and would again propagate the article (reiterating the workflow at the second point).
- The original authors may choose to adopt none, one, or multiple of the revisions and *merge* them into their repository.
- Authors of several articles may also *merge* their work to increase its scope and visibility.

It is worth noting that the above process, though intricate and non-linear, is expeditious. Contingent on the number of interested researchers, an article could undergo manifold circles of revision, forking, and merging in the months usually required for validation through peer review.

5.3 Copyright

We suggest licensing any article intended for OPR under a "Creative Commons Attribution-ShareAlike Unported License" (preferably of the most recent version). This

allows anyone - including journal editors, educators, and policy makers - to build upon an article, and ensures both the continuity of attribution to the researcher(s), and the persistent freedom of the information.

5.4 Authorship

In Git each revision of every single line is automatically attributed to its specific author. Article authorship is thus defined clearer in OPR than in any other means of research publication. We do realize, though, that this very semantic notion of authorship is difficult to reconcile with the currently prevailing concept in scientific publishing. We are also aware that article authorship is not identical with research authorship.

We propose an explicit mention of authors (much like in most scientific journals) beneath the title. On the occasion of each revision, the author thereof should append his name, and those of any who may have helped him (in whatever order he deems fair), to the end of that list. Names should, of course, not be repeated if they are already on the authors' list.

Whenever two or more articles are merged, determining the most appropriate order in which their authors' lists are concatenated becomes the responsibility of the person performing the operation.

The human input specification of authors could be eased in the future by automatic calculation of quantitative authorship. Git indexes all the information required for this, though a script which quantitatively determines authorship has yet to be written.

Independently of the initial number of authors we encourage using the editorial "we" in the wording of articles.

5.5 Article Structuring

We propose structuring articles along the following sections:

- **Abstract** - delineating the scope of the article.
- **Background** - detailing the context of whatever the article discusses.
- **Issues** - detailing precisely and explicitly what the article presumes to tackle.
- **Methods** - detailing the (technical) means by which the issues are to be tackled.
- **Results** - detailing how usage of the methods materializes as far as the author was able to experiment.

Subsections of these items are left to the author's discretion, but we do insist on everyone strictly adhering by these five overhead categories. Rigorous implementation of this model is expected to assure homogeneity, clarity, and relative ease of performing more complex operations such as merging, forking, or text data mining.

We explicitly and strongly discourage using the popular "Discussion" category, as we expect it to provide a playground for highly speculative debate. We encourage brevity in each category.

In the hope of supporting articles from the humanities in OPR, thought experiments may also be considered valid sources of information for the **Results** section.

5.6 File Management

The LaTeX template which we propose for all OPR texts is the one demonstrated by this article. We believe it is of utmost importance that we maintain a consistent visual appearance.

The template should be copied and pasted, and the contents replaced. Original content is written in *article.tex* and the abstract in *README*. References are listed under *bib.bib*. Metadata such as comments on the template should be left untouched, and vital markup modifications should be referred to upstream.

In order to better facilitate prospective data mining and semantic authorship every sentence or subsentence (for instance following the usage of a semicolon), should be placed on a new line. The LaTeX interpreter conveniently does not introduce a break in this context.

The Git repository name for OPR articles should begin with "OPR." followed by the title, contracted as the author sees fit. The reasons for contraction are mostly visual, since Git hosting services tailor their websites to short repository names.

Exported PDF files and figures should only be pushed to a repository for milestone versions (they are explicitly excluded via *.gitignore*). It should be noted that Git keeps track of all changes made to files, and binaries lead to very rapid bloating of repositories. Image files should be tailored to the minimum adequate size for quality PDF printing - 85 µm/dot (300 DPI). Overall, the PDF files in a repository should only be considered a convenience for readers.

All citations of OPR articles should link to a specific revision via its hash (and preferably give a link). In case readers should decide to disseminate loose PDF files, they should append the version hash (and repository) to the file name. Example: `github-repository-ArticleName-2e0ba199b01d55f242ac337baae1df8b6398b988.pdf`

5.7 Supplementary Information

OPR only makes sense if contributors can carefully scrutinize the data and accurately retrace its processing and analysis. We therefore encourage authors to publish their data as well as their post-processing and analysis scripts online, and license them as liberally as possible (ideally also under a "Creative Commons Attribution-ShareAlike Unported License").

Authors should version data post-processing and analysis scripts with Git - so that contribution to those is also possible. We encourage the use of purpose-built scripts in a transparent language (such as Python). For the sake of transparency and accessibility, we would recommend FOSS alternatives to proprietary packages, for instance:

- Scipy-Numpy-Matplotlib[16] over MATLAB
- R[9] or PSPP[2] over SPSS
- PsychoPy[18, 19] over E-Prime and Presentation

If given a choice, OPR materials should cite open access articles rather than restricted access articles of similar quality. Doing so maximizes the extent to which potential contributors and readers can assess whether or not assumptions used in the article are reliable.

5.8 Implementation

We acknowledge that our publishing system does not provide any alleged certificate of quality. While we expect OPR articles to become superior in quality to comparable peer-reviewed static articles, we are aware that the burden of proof still rests with us. It would currently be unwise for researchers to commit articles with high-profile potential to OPR. Therefore, we recommend to use OPR first and foremost to publish work which journals are reticent to accept, such as:

- Negative findings
- Discontinued work
- Ongoing work
- Informal studies conducted in researchers' free time

By acting as a hub for such findings OPR can already and immediately provide a valuable service to the scientific community - since work of this type would otherwise be lost. In fact, revision is most needed for precisely these types of articles; and improving them is an excellent challenge for our system to face. If OPR can make a name by adding value to less promising articles, then indeed it is more laudable than any publishing approach which derives value for itself from more promising ones.

Notes

¹Interdisciplinary Centre for Neurosciences (IZN), University of Heidelberg, INF 364 69120 Heidelberg, Germany.

²Department of Experimental Psychology, Oxford University, Oxford, UK.

References

- [1] General information about arxiv. URL: <http://arxiv.org/help/general>.
- [2] Statistical analysis of sampled data. URL: <http://www.gnu.org/software/pspp/>.
- [3] Coping with peer rejection. *Nature*, 425(6959):645, Oct 2003. URL: <http://dx.doi.org/10.1038/425645a>, doi:10.1038/425645a.
- [4] 2006. URL: <http://www.nature.com/nature/peerreview/debate/nature05535.html>, doi:doi:10.1038/nature05535.
- [5] Richard Akerman. Technical solutions: Evolving peer review for the internet. *Nature*, 2006. URL: <http://dx.doi.org/10.1038/nature04997>, doi:10.1038/nature04997.
- [6] Theodora Bloom. Online frontiers of the peer-reviewed literature. *Nature*, 2006. URL: <http://dx.doi.org/10.1038/nature05030>, doi:10.1038/nature05030.
- [7] Tracey Brown. I don't know what to believe. *Nature*, 2006. URL: <http://dx.doi.org/10.1038/nature04998>, doi:10.1038/nature04998.
- [8] Inc. GitHub. Github - social coding. URL: <https://github.com/>.
- [9] Ross Ihaka and Robert Gentleman. Free software environment for statistical computing and graphics. URL: <http://www.r-project.org/>.
- [10] Charles Jennings. Quality and value: The true purpose of peer review. *Nature*, 2006. URL: <http://dx.doi.org/10.1038/nature05032>, doi:10.1038/nature05032.
- [11] Thomas Koop. An open, two-stage peer-review journal. *Nature*, 2006. URL: <http://dx.doi.org/10.1038/nature04988>, doi:10.1038/nature04988.
- [12] Debomoy Lahiri. Perspective: The case for group review. *Nature*, 2006. URL: <http://dx.doi.org/10.1038/nature05033>, doi:10.1038/nature05033.
- [13] Leslie Lamport. Latex a document preparation system. URL: <http://www.latex-project.org/>.
- [14] Jeffrey T. Leek, Margaret A. Taub, and Fernando J. Pineda. Cooperation between referees and authors increases peer review accuracy. *PLoS One*, 6(11):e26895, 2011. URL: <http://dx.doi.org/10.1371/journal.pone.0026895>, doi:10.1371/journal.pone.0026895.
- [15] Gaell Mainguy, Mohammad R. Motamedi, and Daniel Mitichen. Peer review—the newcomers' perspective. *PLoS Biol*, 3(9):e326, Sep 2005. URL: <http://dx.doi.org/10.1371/journal.pbio.0030326>, doi:10.1371/journal.pbio.0030326.
- [16] Travis Oliphant. Fundamental package for scientific computing with python. URL: <http://www.numpy.org/>.
- [17] Michael Parker. The ethics of open access publishing. *BMC Med Ethics*, 14(1):16, Mar 2013. URL: <http://dx.doi.org/10.1186/1472-6939-14-16>, doi:10.1186/1472-6939-14-16.
- [18] Jonathan Peirce. Open-source application to allow the presentation of stimuli and collection of data for neuroscience, psychology, and psychophysics. URL: <http://www.psychopy.org/>.
- [19] Jonathan W. Peirce. Psychopy—psychophysics software in python. *J Neurosci Methods*, 162(1-2):8–13, May 2007. URL: <http://dx.doi.org/10.1016/j.jneumeth.2006.11.017>, doi:10.1016/j.jneumeth.2006.11.017.
- [20] Brenda Riley. Trusting data's quality. *Nature*, 2006. URL: <http://dx.doi.org/10.1038/nature04993>, doi:10.1038/nature04993.
- [21] Erik Sandewall. Opening up the process. *Nature*, 2006. URL: <http://dx.doi.org/10.1038/nature04994>, doi:10.1038/nature04994.
- [22] Ray Spier. The history of the peer-review process. *Trends Biotechnol*, 20(8):357–358, Aug 2002.
- [23] Linus Torvalds. git –distributed-is-the-new-centralized. URL: <http://git-scm.com/>.
- [24] Richard Van Noorden. Open access: The true cost of science publishing. *Nature*, 495(7442):426–429, Mar 2013. URL: <http://dx.doi.org/10.1038/495426a>, doi:10.1038/495426a.
- [25] S. van Rooyen, F. Godlee, S. Evans, N. Black, and R. Smith. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *BMJ*, 318(7175):23–27, Jan 1999.
- [26] Susan van Rooyen, Tony Delamothe, and Stephen J W. Evans. Effect on peer review of telling reviewers that their signed reviews might be posted on the web: randomised controlled trial. *BMJ*, 341:c5729, 2010.
- [27] Elizabeth Wager. What is it for? analysing the purpose of peer review. *Nature*, 2006. URL: <http://dx.doi.org/10.1038/nature04990>, doi:10.1038/nature04990.