

# Birth of a New School: How Self-Publication can Improve Research

*Micah Allen*

Preface: What follows is an attempt to imagine a radically different future for research publishing. Apologies for any overlooked references – the following is meant to be speculative and purposely walks the line between paper and blog post. Here is to a productive discussion regarding the future of research.

Our current systems of producing, disseminating, and evaluating research could be substantially improved. For-profit publishers enjoy extremely high taxpayer-funded profit margins. Traditional closed-door peer review is creaking under the weight of an exponentially growing knowledge base, delaying important communications and often resulting in seemingly arbitrary publication decisions<sup>1-4</sup>. Today's young researchers are frequently dismayed to find their pain-staking work producing quality reviews overlooked or discouraged by journalistic editorial practices. In response, the research community has risen to the challenge of reform, giving birth to an ever expanding multitude of publishing tools: statistical methods to detect p-hacking<sup>5</sup>, numerous open-source publication models<sup>6-8</sup>, and innovative platforms for data and knowledge sharing<sup>9,10</sup>.

While I applaud the arrival and intent of these tools, I suspect that ultimately publication reform must begin with publication culture – with the very way we think of what a publication is and can be. After all, how can we effectively create infrastructure for practices that do not yet exist? Last summer, shortly after igniting [#pdftribute](#), I began to think more and more about the problems confronting the publication of results. After months of conversations with colleagues I am now

convinced that real reform will come not in the shape of new tools or infrastructures, but rather in the culture surrounding academic publishing itself. In many ways our current publishing infrastructure is the product of a paper-based society keen to produce lasting artifacts of scholarly research. In parallel, the exponential arrival of networked society has led to an open-source software community in which knowledge is not a static artifact but rather an ever-expanding living document of intelligent productivity. We must move towards “research 2.0” and beyond<sup>11</sup>.

From Wikipedia to Github, open-source communities are changing the way knowledge is produced and disseminated. Already this movement changing academia, with researchers across disciplines flocking to social media, blogs, and novel communication infrastructures to create a new movement of post-publication peer review<sup>4,12,13</sup>. In math and physics, researchers have embraced self-publication, uploading preprints to the online repository arXiv, with more and more disciplines using the site to archive their research. I believe that the inevitable future of research communication is in this open-source metaphor, in the form of pervasive self-publication of scholarly knowledge. The question is thus not where are we going, but rather how do we prepare for this radical

---

## **p-hacking**

Sub-par research practices that purposely or inadvertently undermine confidence in published findings. Typically involves unreported "researcher degrees of freedoms", e.g. collecting data until an effect is found, selective subject exclusion, and incomplete disclosure of hypothesis testing.

change in publication culture. In asking these questions I would like to imagine what research will look like 10, 15, or even 20 years from today. The following is intended as an initial step towards bringing to light specific ideas for how this transition might be facilitated. I invite you to treat what follows as an ‘open beta’ for these ideas.

## Part 1: Why self-publication?

I believe the essential metaphor is within the open-source software community. To this end over the past few months I have feverishly discussed the merits and risks of self-publishing scholarly knowledge with my colleagues and peers. While at first I was worried many would find the notion of self-publication utterly absurd, I have been astonished at the responses – many have been excitedly optimistic! I was surprised to find that some of my most critical and stoic colleagues have lost so much faith in traditional publication and peer review that they are ready to consider more radical options.

The basic motivation for research self-publication is pretty simple: *research papers cannot be properly evaluated without first being read*. Now, by evaluation, I don’t mean for the purposes of hiring or grant giving committees. These are essentially financial decisions, e.g. “how do I effectively spend my money without reading the papers of the 200+ applicants for this position?” Such decisions will always rely on heuristics and metrics that must necessarily sacrifice accuracy for efficiency. However, I believe that self-publication culture will provide a finer grain of metrics than ever dreamed of under our current system. By documenting each step of the research process, self-publication and open science can yield rich information that can be mined for increasingly useful impact measures – but more on that later.

When it comes to evaluating research, many

admit that there is no substitute for opening up an article and reading its content – regardless of journal. My prediction is that as post-publication peer review gains acceptance, some tenured researcher or brave young scholar will eventually decide to simply self-publish her research directly onto the internet. When that research goes viral, the resulting deluge of self-publications will be overwhelming. Of course, busy lives require heuristic decisions and it’s arguable that publishers provide this editorial service. While I will address this issue specifically in Part 3, for now I want to point out that growing empirical evidence suggests that our current publisher/impact-based system provides an unreliable heuristic at best<sup>14-16</sup>. Thus, my essential reason for supporting self-publication is that in the worst-case scenario, self-publications must be accompanied by the disclaimer: “read the contents and decide for yourself.” As self-publishing practices are established, it is easy to imagine that these difficulties will be largely mitigated by self-published peer reviews and novel infrastructures supporting these interactions.

Indeed, with a little imagination we can picture plenty of potential benefits of self-publication to offset the risk that we might read poor papers. Researchers spend exorbitant amounts of their time reviewing, commenting on, and discussing articles – most of that rich content and meta-data is lost under the current system. In documenting the research practice more thoroughly, the ensuing flood of self-published data can support new quantitative metrics of reviewer trust, and be further utilized in the development of rich information about new ideas and data in near real-time. To give just one example, we might calculate how many subsequent citations or retractions a particular reviewer generates, generating a reviewer impact factor and reliability index. The more aspects of research we publish, the greater the data-mining potential. Incentivizing in-depth reviews that add

"Read the contents and decide for yourself."

- the new school of research evaluation

clarity and conceptual content to research, rather than merely knocking down or propping up equally imperfect artifacts, will ultimately improve research quality. By self-publishing well-documented, open-sourced pilot data and accompanying digital reagents (e.g. scripts, stimulus materials, protocols, etc), researchers can get instant feedback from peers, preventing uncounted research dollars from being wasted. Previously closed-door conferences can become live records of new ideas and conceptual developments as they unfold. The metaphor here is research as open-source – an ever evolving, living record of knowledge as it is created.

Now, let's contrast this model to the current publishing system. Every publisher (including open-access) obliges researchers to adhere to randomly varied formatting constraints, presentation rules, submission and acceptance fees, and review cultures. Researchers perform reviews for free for often publically subsidized work, so that publishers can then turn around and sell the finished product back to those same researchers (and the public) at an exorbitant mark-up. These constraints introduce lengthy delays – ranging from 6+ months in the sciences all the way up to two years in some humanities disciplines. By contrast, how you self-publish your research is entirely up to you – where, when, how, the formatting, and the openness. Put simply, if you could publish your research how and when you wanted, and have it generate the same “impact” as traditional venues, why would you use a publisher at all?

One obvious reason to use publishers is copy-editing, i.e. the creation of pretty manuscripts. Another is the guarantee of high-profile distribution. Indeed, under the current system these are legitimate worries. While it is possible to produce reasonably formatted papers, ideally the creation of an open-source, easy to use copy-editing software is needed to facilitate mainstream self-publication.

Innovators like figshare are already leading the way in this area. In the next section, I will try to theorize some different ways in which self-publication can overcome these and other potential limitations, in terms of specific applications and guidelines for maximizing the utility of self-published research. To do so, I will outline a few specific cases with the most potential for self-publication to make a positive impact on research right away, and hopefully illuminate the ‘why’ question a bit further with some concrete examples.

## Part 2: Where to begin self-publishing

What follows is the “how-to” part of this document. I must preface by saying that although I have written so far with researchers across the sciences and humanities in mind, I will now focus primarily on the scientific examples with which I am more experienced. The transition to self-publication is already happening in the forms of academic tweets, self-archives, and blogs, at a seemingly exponential growth rate. To be clear, I do not believe that the new publication culture will be utopian. As in many human endeavors the usual brandism<sup>3</sup>, politics, and corruption can be expected to appear in this new culture. Accordingly, the transition is likely to be a bit wild and woolly around the edges. Like any generational culture shift, new practices must first emerge before infrastructures can be put in place to support them. My hope is to contribute to that cultural shift from artifact to process-based research, outlining particularly promising early venues for self-publication. Once these practices become more common, there will be huge opportunities for those ready and willing to step in and provide rich informational architectures to support and enhance self-publication – but for now we can only step into that wild frontier.

In my discussions with others I have identified three particularly promising areas where self-

Do we really need publishers to share our research?

publication is either already contributing or can begin contributing to research. These are: the publication of exploratory pilot-data, post-publication peer reviews, and trial pre-registration. I will cover each in turn, attempting to provide examples and templates where possible. Finally, Part 3 will examine some common concerns with self-publication. In general, I think that successful reforms should resemble existing research practices as much as possible: publication solutions are most effective when they resemble daily practices that are already in place, rather than forcing individuals into novel practices or infrastructures with an unclear time-commitment. A frequent criticism of current solutions such as the comments section on Frontiers, PLOS One, or the newly developed PubPeer, is that they are rarely used by the general academic population. It is reasonable to conclude that this is because already over-worked academics currently see little plausible benefit from contributing to these discussions given the current publishing culture (worse still, they may fear other negative repercussions, discussed in Part 3). Thus a central theme of the following examples is that they attempt to mirror practices in which many academics are already engaged, with complementary incentive structures (e.g. citations).

## **Example 1: Exploratory Pilot Data**

This previous summer witnessed a fascinating clash of research cultures, with the eruption of intense debate between pre-registration advocates and pre-registration skeptics. I derived some useful insights from both sides of that discussion. Many were concerned about what would happen to exploratory data under these new publication regimes. Indeed, a general worry with existing reform movements is that they appear to emphasize a highly conservative and somewhat cynical “perfect papers” culture. I do not believe in perfect papers – the scientific model

is driven by replication and discovery. No paper can ever be 100% flawless – otherwise there would be no reason for further research! Inevitably, some will find ways to cheat the system. Accordingly, reform must incentivize better reporting practices over stricter control, or at least balance between the two extremes.

Exploratory pilot data is an excellent avenue for this. By their very nature such data are not confirmatory – they are exciting in that they do not conform well to prior predictions. Such data benefit from rapid communication and feedback. Imagine an intuition-based project – a side or pet project conducted on the fly for example. The researcher might feel that the project has potential, but also knows that there could be serious flaws. Most journals won’t publish these kinds of data. Under the current system these data are lost, hidden, obscured, or otherwise forgotten.

Compare to a self-publication world: the researcher can upload the data, document all the protocols, make the presentation and analysis scripts open-source, and provide some well-written documentation explaining why she thinks the data are of interest. Some intrepid graduate student might find it, and follow up with a valuable control analysis, pointing out an excellent feature or fatal flaw, which he can then upload as a direct citation to the original data. Both publications are citable, giving credit to originator and reviewer alike. Armed with this new knowledge, the original researcher could now pre-register an altered protocol and conduct a full study on the subject (or alternatively, abandon the project entirely). In this exchange, it is likely that hundreds of hours and research dollars will have been saved. Additionally, the entire process will have been documented, making it both citable and minable for impact metrics. Tools already exist for each of these steps – but largely cultural fears prevent it from happening. *How would it be perceived? Would anyone*



*read it? Will someone steal my idea?* To better frame these issues, I will now examine a self-publication practice that has already emerged in force.

### **Example 2: Post-publication peer review**

This is a particularly easy case, precisely because high-profile scholars are already regularly engaged in the practice. As I've frequently joked on twitter, we're rapidly entering an era where publishing in a glam-mag has no impact guarantee if the paper itself isn't worthwhile – you may as well hang a target on your head for post-publication peer reviewers. However, I want to emphasize the positive benefits and not just the conservative controls. Post-publication peer review (PPPR) has already begun to change the way we view research, with reviewers adding lasting content to papers, enriching the conclusions one can draw, and pointing out novel connections that were not extrapolated upon by the authors themselves. Here I like to draw an analogy to the open source movement, where code (and its documentation) is forkable, versioned, and open to constant revision – never static but always evolving.

Indeed, just last week PubMed launched their new “PubMed Commons” system, an innovative PPPR comment system, whereby any registered person (with at least one paper on PubMed) can leave scientific comments on articles. Inevitably, the reception on Twitter and Facebook mirrored previous attempts to introduce infrastructure-based solutions – mixed excitement followed by a lot of bemused cynicism – bring out the trolls many joked. To wit, a brief scan of the average comment on another platform, PubPeer, revealed many low-quality comments, perhaps encouraged by the anonymous, low-incentive format of that venue. While many comments were on topic, most had little to no formatting and were given with little

context. On the internet, all too often comments end up trollish. A challenge for PPPR will be to incentivize quality reviews to ensure greater participation and value, going beyond merely pointing out minor flaws as if they render the paper worthless to actually adding content to research. This is a serious barrier to entry for PPR; why would anyone acknowledge such a system if the primary result is endless nitpicking of their research? The essential problem here is incentive – for reviews to be quality there needs to be a motivation for researchers to put their time and effort into them. We need a culture of PPPR that values positive and negative comments equally. This is common to both traditional and self-publication practices.

### **Example 3: Pre-registration of experimental trials**

For many researchers, self-publication of trial pre-registrations (PR) may be an excellent way to test the waters of PR in a format with a low barrier to entry. Replication attempts are a particularly promising venue for PR, and self-publication of such registrations is a way to quickly move from idea to registration to collection (as in the above pilot data example), while ensuring that credit for the original idea is embedded in the infamously hard to erase memory of the internet.

A few benefits of PR self-publication, rather than relying on for-profit publishers, is that PR templates can be easily open-sourced themselves, allowing various research fields to generate community-based specialized templates adhering to the needs of that field. Self-published PRs, as well as high quality templates, can be cited – incentivizing the creation and dissemination of both. Indeed, as self-publication practices solidify, it is likely that a multitude of preformatted templates for data sharing, code documentation, and

---

### **Forking**

Forking is the process by which open-source software differentiates into distinct packages driven by independent developers. Open-source software by definition may be forked and the resulting code published without permission from the original authors. In this way a project can be utilised far beyond it's original scope, improving long term viability and impact.

manuscript publishing/versioning will emerge as standardized tools. The rapid emergence of specialized templates within each community, tailored to the needs of that research discipline, will further help to promote greater transparency with less overhead for researchers.

### Part 3: Criticism and limitations

Here I close by considering some common concerns with self-publication:

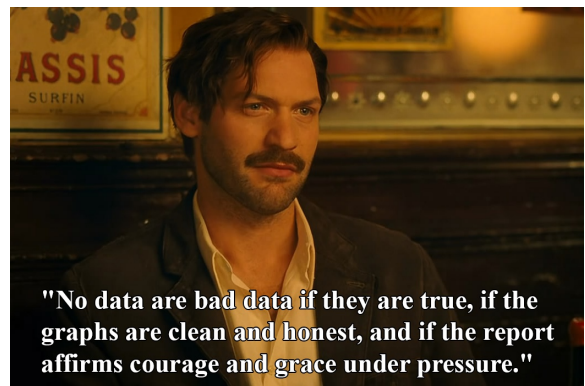
#### *Quality of data*

A natural worry at this point is quality control. How can we be sure that what is published without the seal of peer review isn't complete hokey? The primary response is that we cannot, just like we cannot be sure that peer reviewed materials are quality without first reading them ourselves. Still, it is for this reason that I tried to suggest a few particularly ripe venues for self-publication of research. The cultural zeitgeist supporting full-blown scholarly self-publication has not yet arrived, but we can already begin to prepare for it. With regards to filtering noise, I argue that by coupling post-publication peer review and social media, quality self-publications will rise to the top. Importantly, this issue points towards flaws in our current publication culture. In many research areas there are effects that are repeatedly published but that few believe, largely due to the presence of biases against null-findings. Self-publication aims to make as much of the research process publicly available as possible, preventing this kind of knowledge from slipping through the editorial cracks and improving our ability to evaluate the veracity of published effects. If such data are reported cleanly and completely, existing quantitative tools can further incorporate them to better estimate the likelihood of p-hacking within a

literature. That leads to the next concern – quality of presentation.

#### *Quality of presentation*

Many ask: how in this brave new world will we separate signal from noise? I am sure that every published researcher already receives at least a few garbage citations a year from obscure places in obscure journals with little relevance to actual article contents. But, so the worry goes, what if we are deluged with a vast array of poorly written, poorly documented, self-published crud. How would we separate the signal from the noise?



The answer is **Content, Clarity, and Presentation**. These are central guidelines for self-publication to be worth anyone's time. The Internet memesphere has already generated one rule for ranking interest: content rules. Content floats and is upvoted, blogspam sinks and is downvoted. This is already true for published articles – twitter, reddit, facebook, and email circles help us separate the wheat from the chaff at least as much as impact factor if not more. But presentation and clarity are equally important. Poorly conducted research is not shared, or at least is shared with vehemence. Similarly, poorly written self-publications, or poorly documented data/reagents are unlikely to generate positive feedback, much less impact-generating

eyeballs. I like to imagine a distant future in which self-publication has given rise to a new generation of well-regarded specialists: reviewers who are prized for their content, presentation, and clarity; coders who produce cleanly documented pipelines; behaviorists producing powerful and easily customized paradigm scripts; and data collection experts who produce the smoothest, cleanest data around. All of these future specialists will be able to garner impact for the things they already do, incentivizing each step of the research processes rather than only the end product.

*Being scooped, intellectual credit*

Another common concern is “what if my idea/data/pilot is scooped?” I acknowledge that particularly in these early days, the decision to self-publish must be weighted against this possibility. However, I must also point out that in the current system authors must also weight the decision to develop an idea in isolation against the benefits of communicating with peers and colleagues. Both have risks and benefits – an idea or project in isolation can easily over-estimate its own quality or impact. The decision to self-publish must similarly be weighted against the need for feedback. Furthermore, a self-publication culture would allow researchers to move more quickly from project to publication, ensuring that they are readily credited for their work. And again, as research culture continues to evolve, I believe this concern will increasingly fade. It is notoriously difficult to erase information from The Internet (see the “Streisand effect”) – there is no reason why self-published ideas and data cannot generate direct credit for the authors. Indeed, I envision a world in which these contributions can themselves be independently weighted and credited.

*Prevention of cheating, corruption, self-citations*

To some, this will be an inevitable point of departure. Without our time-tested guardian of peer review, what is to prevent a flood of outright fabricated data? My response is: what prevents outright fabrication under the current system? To misquote Jeff Goldblum in *Jurassic Park*, cheaters will always find a way. No matter how much we tighten our grip, there will be those who respond to the pressures of publication by deliberate misconduct. I believe that the current publication system directly incentivizes such behavior by valuing end product over process. By creating incentives for low-barrier post-publication peer review, pre-registration, and rich pilot data publication, researchers are given the opportunity to generate impact for each step of the research process. When faced with the vast penalties of cheating due to a null finding, versus doing one’s best to turn those data into something useful for someone, I suspect most people will choose the honest and less risky option.

Corruption and self-citations are perhaps a subtler, more sinister factor. In my discussions with colleagues, a frequent concern is that there is nothing to prevent high-impact “rich club” institutions from banding together to provide glossy post-publication reviews, citation farming, or promoting one another’s research to the top of the pile regardless of content. I again answer: how is this any different from our current system? Papers are submitted to an editor who makes a subjective evaluation of the paper’s quality and impact, before sending it to four out of a thousand possible reviewers who will make an obscure decision about the content of the paper. Sometimes this system works well, but increasingly it does not<sup>2</sup>. Many have witnessed great papers rejected for political reasons, or poor ones accepted for the same. Lowering the barrier to post-publication peer review means that even when these factors drive a

paper to the top, it will be far easier to contextualize that research with a heavy dose of reality. Over time, I believe self-publication will incentivize good research. Cheating will always be a factor – and this new frontier is unlikely to be a utopia. Rather, I hope to contribute to the development of a bridge between our traditional publishing models and a radically advanced not-too-distant future.

## Conclusion

Our current systems of producing, disseminating, and evaluating research increasingly seem to be out of step with cultural and technological realities. To take back the research process and bolster the ailing standard of peer-review I believe research will ultimately adopt an open and largely publisher-free model. In my view, these new practices will be entirely complementary to existing solutions including such as the p-curve<sup>5</sup>, open-source publication models<sup>6-8</sup>, and innovative platforms for data and knowledge sharing such as PubPeer, PubMed Commons, and figshare<sup>9,10</sup>. The next step is to simply start publishing. Don't wait - upload your data to figshare, document your code on Github, publish your peer reviews on blogs and Pubpeer. The frontier is now. In attempting to build a bridge to the coming technological and social revolution, I hope to inspire others to join in the conversation so that we can improve all aspects of research.

## Acknowledgments

*Thanks to Hauke Hillebrandt, Kate Mills, and Francesca Fardo for invaluable discussion, comments, and edits of this work. Many of the ideas developed here were originally inspired by this post envisioning a self-publication future. Thanks also to PubPeer, PeerJ, figshare, and others in this area for their*

*pioneering work in providing some valuable tools and spaces to begin engaging with self-publication practices.*

## Addendum

Excellent resources already exist for the many of the ideas presented here. I want to give special notice to researchers who have already begun self-publishing their work either as preprints, archives, or as direct blog posts. Parallel publishing is an attractive transitional option where researchers can prepublish their work for immediate feedback before submitting it to a traditional publisher. Special notice should be given to Zen Faulkes whose excellent pioneering blog posts demonstrated that it is reasonably easy to self-produce well formatted publications. Here are a few pioneering self-published papers you can use as examples:

The distal leg motor neurons of slipper lobsters, *Ibacus* spp. (Decapoda, Scyllaridae), Zen Faulkes  
<http://neurodojo.blogspot.dk/2012/09/Ibacus.html>

Eklund, Anders (2013): Multivariate fMRI Analysis using Canonical Correlation Analysis instead of Classifiers, Comment on Todd et al. figshare.  
<http://dx.doi.org/10.6084/m9.figshare.787696>

Automated removal of independent components to reduce trial-by-trial variation in event-related potentials, Dorothy Bishop  
[http://bishoptechbits.blogspot.dk/2011\\_05\\_01\\_archive.html](http://bishoptechbits.blogspot.dk/2011_05_01_archive.html)

Deep Impact: Unintended consequences of journal



rank

Björn Brembs, Marcus Munafò  
<http://arxiv.org/abs/1301.3748>

A novel platform for open peer to peer  
 review and publication:  
<http://thewinnower.com/>

A platform for open PPPRs:  
<http://pubpeer.com/>

Another PPPR platform:  
<http://f1000.com/>

## References

- Henderson, M. Problems with peer review. *BMJ* 340, c1409 (2010).
- Ioannidis, J. P. A. Why Most Published Research Findings Are False. *PLoS Med* 2, e124 (2005).
- Peters, D. P. & Ceci, S. J. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behav. Brain Sci.* 5, 187 (2010).
- Hunter, J. Post-publication peer review: opening up scientific conversation. *Front. Comput. Neurosci.* 6, 63 (2012).
- Simonsohn, U., Nelson, L. D. & Simmons, J. P. P-Curve: A Key to the File Drawer. (2013).  
<http://papers.ssrn.com/abstract=2256237>
- MacCallum, C. J. ONE for All: The Next Step for PLoS. *PLoS Biol.* 4, e401 (2006).
- Smith, K. A. The frontiers publishing paradigm. *Front. Immunol.* 3, 1 (2012).
- Wets, K., Weedon, D. & Velterop, J. Post-publication filtering and evaluation: Faculty of 1000. *Learn. Publ.* 16, 249–258 (2003).
- Allen, M. PubPeer – A universal comment and review layer for scholarly papers? | Neuroconscience on WordPress.com. Website/Blog (2013).  
<http://neuroconscience.com/2013/01/25/pubpeer-a-universal-comment-and-review-layer-for-scholarly-papers/>
- Hahnel, M. Exclusive: figshare a new open data project that wants to change the future of scholarly publishing. *Impact Soc. Sci. blog* (2012).  
[http://eprints.lse.ac.uk/51893/1/blogs.lse.ac.uk-Exclusive\\_figshare\\_a\\_new\\_open\\_data\\_project\\_that\\_wants\\_to\\_change\\_the\\_future\\_of\\_scholarly\\_publishing.pdf](http://eprints.lse.ac.uk/51893/1/blogs.lse.ac.uk-Exclusive_figshare_a_new_open_data_project_that_wants_to_change_the_future_of_scholarly_publishing.pdf)
- Yarkoni, T., Poldrack, R. A., Van Essen, D. C. & Wager, T. D. Cognitive neuroscience 2.0: building a cumulative science of human brain function. *Trends Cogn. Sci.* 14, 489–496 (2010).
- Bishop, D. BishopBlog: A gentle introduction to Twitter for the apprehensive academic. Blog/website (2013).  
<http://deevybee.blogspot.dk/2011/06/gentle-introduction-to-twitter-for.html>
- Hadibeenareviewer. Had I Been A Reviewer on WordPress.com. Blog/website (2013).  
<http://hadibeenareviewer.wordpress.com/>
- Tressoldi, P. E., Giofrè, D., Sella, F. & Cumming, G. High Impact! =! High Statistical Standards? Not Necessarily So. *PLoS One* 8, e56180 (2013).
- Brembs, B. & Munafò, M. Deep Impact: Unintended consequences of journal rank. (2013).  
<http://arxiv.org/abs/1301.3748>
- Eisen, J. A., Maccallum, C. J. & Neylon, C. Expert Failure: Re-evaluating Research Assessment. *PLoS Biol.* 11, e1001677 (2013).