Psychological Inquiry, 23: 305–307, 2012

ISSN: 1047-840X print / 1532-7965 online
DOI: 10.1080/1047840X.2012.706204

Psychology Press
Taylor & Francis Group

Beginning at Nosek and Bar-Anan's End: Let's Put Open Evaluation First

Tal Yarkoni
Institute of Cognitive Science, University of Colorado, Boulder, Colorado

Brian A. Nosek and Yoav Bar-Anan's (this issue) target article offers an exceptionally clear, insightful, and ambitious vision of a future in which scientists no longer have to worry very much about day-to-day annoyances like access restrictions, publication delays, or arbitrary editorial decisions and can instead focus their efforts on the process of actually producing and evaluating science. Because I share virtually all of Nosek and Bar-Anan's concerns about current publishing practices (Yarkoni, in press), and welcome their utopian vision of the future, this commentary brooks no substantive disagreement about where we're collectively headed. Instead, it focuses on the navigational problem of how to best get there from here. Whereas Nosek and Bar-Anan propose a series of six cumulative steps that ultimately culminate with platforms for open, continuous review, I instead argue that if we focus the bulk of our efforts on developing open evaluation platforms right now, all of the other steps are likely to follow in short order.

Putting Postpublication Evaluation First

Nosek and Bar-Anan take their six steps to be cumulative and partially dependent. But their rationale for emphasizing sequential dependency isn't always clear. As Nosek and Bar-Anan themselves concede, in many cases it appears that "it is possible to initiate aspects of later stages prior to complete adoption of earlier ones." In fact, in some cases it's not clear that there's *any* dependency between steps. Particularly problematic, in my view, is that Nosek and Bar-Anan see the implementation of open, continuous peer review as the very last of their six steps to be fully implemented. I have recently argued that centralized and open peer review is probably the *easiest* part of a next-generation scientific communication model to implement (Yarkoni, in press), and quite possibly also the most beneficial. Yarkoni (in press) lays out the case for open review platforms in detail (see also Kravitz & Baker, 2011; Walther & van den Bosch, 2012); as Nosek and Bar-Anan make many of the same points, my focus here is on demonstrating that full-featured open review platforms don't need to be preceded by any of Nosek and Bar-Anan's other steps. To the contrary, I argue that focusing on the last step first will have three major benefits. First, open peer review systems can be implemented much more quickly than almost any of Nosek and Bar-Anan's other suggestions. Second, it will likely be easier to incentivize researchers to participate in open peer review systems than to elicit many of the other desired behaviors. Third, successful implementation of open postpublication review will naturally accelerate progress toward all of the other steps Nosek and Bar-Anan describe, without requiring substantial additional effort. I discuss each of these points in turn.

We Can Build It—Right Now

Building an engaging, full-featured postpublication evaluation platform is clearly not a trivial enterprise (Yarkoni, in press). However, the challenges involved in building such platforms arguably pale compared to those associated with most of the other problems Nosek and Bar-Anan discuss. As Nosek and Bar-Anan note, the biggest barriers to change are typically social rather than technological. For example, you can't have universal open access (OA) until you can convince a large proportion of key stakeholders to invest a considerable amount of time and money realizing such goals. But given that commercial publishers like Elsevier are unlikely to preserve their characteristic 30%+ profit margins (Elsevier, 2011) under an open-access model, it's unlikely that they'll voluntarily opt to go gentle into that dark night. Although the fights for universal open access and fully digital communication are clearly worth fighting, they seem likely to take a long time to win. In contrast, a small group of competent software developers could build a sophisticated and effective postpublication evaluation platform in a matter of months. In fact, such a platform wouldn't even need to be built from scratch; as I've previously suggested, and Nosek and Bar-Anan also discuss, open-source platforms like reddit already provide many if not most of the features the scientific community would want in an evaluation platform (Yarkoni, in press). Consequently, developing such platforms will probably be more a matter of judiciously borrowing existing code than of having to solve any major new engineering problems.

In fairness, Nosek and Bar-Anan do note that a number of publishers, like PLoS, already have rudimentary open evaluation platforms. But they don't push this point far enough. If nonprofit OA publishers like PLoS and Frontiers (with much smaller budgets than most commercial publishers) can readily implement evaluation platforms tied to their own content, what's to prevent an enterprising group of scientists or developers from creating an evaluation platform that encompasses *all* of the scientific literature? One might object that a comprehensive open peer review system isn't possible until we get rid of copyright restrictions and access barriers (Step 2). But in practice, there's no reason a postpublication evaluation platform has to take any kind of stance on access restrictions. As Nosek and Bar-Anan themselves note in their Step 3, the publication and evaluation processes can be cleanly separated. All an open evaluation platform needs to do is *link to* articles elsewhere on the web; it's not the developers' concern whether every end user can access those articles. (In fact, as I discuss next, the fact that many users wouldn't have access to a large proportion of articles should, if anything, incentivize publishers to move to OA.) From a technical standpoint, Step 6 is probably the easiest of all of Nosek and Bar-Anan's steps to actually implement.

Incentivization Is Relatively Easy

Step 6 is probably also the easiest step to achieve from a social engineering perspective. Most of the problems Nosek and Bar-Anan identify have deep systemic roots; it's hard to envision universal OA or fully digital communication being achieved within a decade, let alone a couple of years. But postpublication evaluation is a different story. There's at least one very good reason to think that a well-designed open evaluation platform would attract a very large and very active userbase very quickly: It's already worked in any number of other domains. Evaluation platforms are a central reason for the success of commercial websites like Amazon and Netflix, as well as social news sites like reddit. At the core of these platforms is a simple formula: first, allow users to review or comment on virtually any product in the database; second, allow the resulting reviews and comments to be recursively rated by other users; and third, tie each user's reputation directly to the quality and number of their contributed reviews and comments. This formula has been successfully replicated so often on the web that we no longer find it all noteworthy; in fact, at this point there are few if any online merchants that *haven't* implemented evaluation and/or recommendation systems along these lines.

One might object that reviewing a TV on Amazon or commenting on a posted link on reddit is a far cry from the delicate act of reviewing a scientific paper. Fortunately, we need look no further than sites like Math Exchange (http://math.stackexchange.com)—a mathematics question-and-answer site based on the popular Stack Exchange platform—for evidence to the contrary. Math Exchange has nearly 30,000 registered users, a large proportion of who are tenured or tenure-track math professors. The Cognitive Sciences Stack Exchange (http://cogsci.stackexchange.com), which launched in February 2012, has managed to accumulate nearly 1,000 registered users in just 4 months. Clearly, busy academics seem quite happy to spend much of their time asking and answering questions, reviewing arguments, and providing advice online for free—just as long as there's a well-designed reputation system in place to incentivize such contributions. Moreover, as both Yarkoni (in press) and Nosek and Bar-Anan (2012) note, a major benefit of open review systems is that many trained scientists working in nonacademic positions could make potentially valuable contributions to the research enterprise without having to have labs or write papers. These potential users represent an enormous squandered resource that is currently powerless to help effect any of the other changes Nosek and Bar-Anan discuss but would be in an excellent position to help evaluate the scientific literature. In sum, social considerations, much like technical considerations, would seem to call for a much greater focus on postpublication evaluation.

All Else Will Follow

It is important to note that focusing on open evaluation platforms wouldn't require us to ignore any of the other steps that Nosek and Bar-Anan discuss. To the contrary, a well-designed and well-implemented postpublication evaluation platform would go a long way toward addressing virtually all of the other problems Nosek and Bar-Anan raise—no extra effort required. For example, Step 5—making peer review public—is virtually guaranteed in an open review system. Similarly, the goal of Step 3 is to cleanly separate publication from evaluation, and this is true almost by definition for postpublication evaluation platforms, which would only link to, and not host, their content. Although it's true that an open review platform wouldn't feature the kind of expert review service that Nosek and Bar-Anan propose in Step 3, it's not obvious that that kind of review offers any real benefits over a completely open system. An enormous literature—some of it cited by Nosek and Bar-Anan—demonstrates convincingly that the reliability of conventional expert-based peer review is surprisingly low (Bornmann, Mutz, & Daniel, 2010). And we know from decades of work by Paul Meehl and others that expert judgment in domains where decisions are subjective and

feedback is infrequent is notoriously poor (Dawes, Faust, & Meehl, 1989; Hanson & Morton-Bourgon, 2009; White, 2006). Put simply, we have little reason to think that the kind of expert peer review service that Nosek and Bar-Anan envision in Step 3 would be any more reliable than the current standard. What open review loses in expertise it will almost certainly more than make up for in volume, interactive discussion, and collaborative filtering.

But the benefits don't stop there. Consider the impact on Nosek and Bar-Anan's Step 2—achieving universal OA. Once a postpublication evaluation platform achieves critical mass, there will be an enormous incentive for publishers to convert their journals to OA, because an article that only a fraction of the userbase can read is an article that only a fraction of the userbase will rate. The net effect is that articles published in OA journals are likely to be ranked considerably higher—and ultimately cited more often—than those published in closed-access journals. It's hard to envision a better argument for OA than "all your journals' impact factors are going to keep dropping unless you do something fast." To some degree this trend is already evident in what some have termed the "OA advantage" (Eysenbach, 2006; Lawrence, 2001; Norris, Oppenheim, & Rowland, 2008), but the centralization of postpublication review and commentary will accelerate this process by making barriers to access—and the penalties for failing to remove them—more salient than ever.

Similarly, once a postpublication evaluation system achieves critical mass, the "one article, one journal" model that Nosek and Bar-Anan target in Step 4 is likely to dissolve much more rapidly. In fact, the very idea of a *journal* becomes archaic. Why spend several months getting your article rubber stamped (or, worse, rejected!) by three people behind closed doors when you can have it interactively evaluated and critiqued by the entire community at the push of an Upload button? In a world of open commentary, collaborative filtering algorithms, and threaded discussions, there is little reason to retain even so much as the notion of an "officially published" article.

The bottom line is that Nosek and Bar-Anan's Step 6 need not languish at the end of the utopian chain. There is currently no major barrier to the rapid implementation of comprehensive, full-featured, open postpublication evaluation platforms, and if anything, development of such platforms is likely to substantially hasten Nosek and Bar-Anan's other steps. I share virtually all of Nosek and Bar-Anan's concerns about current scientific communication practices and greatly admire the utopian vision they articulate; however, I suspect that we will realize that vision much faster if we begin at their end and work our way backward.

Note

Address correspondence to Tal Yarkoni, Institute of Cognitive Science, University of Colorado Boulder, 344 UCB, Boulder, CO 80309. E-mail: tal.yarkoni@colorado.edu

References

Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A reliability-generalization study of journal peer reviews: a multilevel meta-analysis of inter-rater reliability and its determinants. *PloS one*, *5*, e14331. doi:10.1371/journal.pone.0014331

Dawes, R., Faust, D., & Meehl, P. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668–1674. doi:10.1126/science.2648573

Elsevier, R. (2011). *Annual reports and financial statements*. Retrieved from http://reporting.reedelsevier.com/staticreports/Reed_AR_2011.pdf

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, *4*, e157. doi:10.1371/journal.pbio.0040157

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, *21*, 1–21. doi:10.1037/a0014421

Kravitz, D. J., & Baker, C. I. (2011). Toward a new model of scientific publishing: Discussion and a proposal. *Frontiers in Computational Neuroscience*, *5*, 55. doi:10.3389/fncom.2011.00055

Lawrence, S. (2001). Online or invisible. *Nature*, *411*, 521.

Norris, M., Oppenheim, C., & Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology*, *59*, 1963–1972. doi:10.1002/asi

Walther, A., & van den Bosch, J. J. F. (2012). FOSE: A framework for open science evaluation. *Frontiers in Computational Neuroscience*, *6*, 32. doi:10.3389/fncom.2012.00032

White, M. J. (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*, 341–382. doi:10.1177/0011000005285875

Yarkoni, T. (in press). Designing next-generation platforms for evaluating scientific output: What scientists can learn from the social web. *Frontiers in Computational Neuroscience*.