How should scientific journals handle "Big if true" submissions?[1]

Andrew Gelman[2]

8 Dec 2021

Scientific journals are littered with embarrassing publications. Perhaps the most consequential and disastrous such example was from *Lancet*'s 1998 publication of a vaccine denial paper which was only retracted twelve years later. Other noteworthy examples are a 2011 article on extra-sensory perception appearing in the *Journal of Personality and Social Psychology* and a 1994 article on a purported "Bible code" published in *Statistical Science*.

Each of these papers spawned definitive refutations; for example, see Eggertson (2010), Engber (2017), and McKay et al. (1999). Indeed, all these articles were notorious in large part because they were generally considered implausible or outlandish even at the time they appeared, and this raises the question of why reputable journals published them in the first place. We can't be sure, but our guess is that the motivation for publication was a mix of aversion to censorship and some fear of missing out in the unlikely event that the papers represented real research findings.

In support of the journals' decision to publish is the larger perspective is that science is inherently speculative. True findings will be replicated while mistakes will fail to replicate, so it is better practice to publish than to suppress. That all three of these papers were discredited on both substantive and methodological grounds could be taken as an example of the self-correcting nature of science and a vindication of the journals' publication policies. Better to get an idea out in the open and have it be refuted, than to let it fester underground.

But publishing bad papers has consequences. The vaccine denial paper added fuel to a movement which, decades later, has hindered the world's response to a pandemic. The ESP and Bible code papers fall more in the "feature story" than "news" categories, but at the very least they displace more serious scientific work. Journals have page limitations, and scientists and the news media have attention limitations.

The question is, what can or should be done? How can journals satisfy an admirable desire for open-mindedness and aversion to censorship while minimizing the publication of junk science?

We consider this question in the context of the Bem (2011) paper reporting extra-sensory perception among Cornell students, which we choose partly for its high profile—the decision of a top psychology journal publish and the subsequent discussion of this research in the New York Times and elsewhere was arguably the single event that kicked

---

off the replication crisis in science—and partly because it is a clean example, without the political or religious overtones of vaccine denial or the Bible code.

The journal's decision to publish that ESP paper was, from a scientific standpoint, a mistake.

But what should the editors have done? It's a tough choice:
Option A. Publish an article when you're pretty sure its theories and conclusions are completely wrong; or
Option B. Reject an article that has no obvious methodological flaws (or, at least none that were noticed at the time; in retrospect the data analysis has big problems, see section 2.2 of Gelman and Loken, 2014, for just one example).

Before going on, we emphasize that top journals reject articles without obvious flaws *all the time*. A common reason for rejecting an article is that it's not important enough. What about that article on ESP? Well, if its claims were correct, then it would be super-important. On the other hand, if there's nothing there, it's not important at all! So it's hard to untangle the criteria of correctness and importance. Here we're just pointing out that Option B is not so unreasonable: JPSP is allowed to reject a paper that makes big claims about ESP, just as it's allowed to reject a paper that appears to be correct but is on a topic that they judge to be too boring or specialized to be of general interest.

Anyway, to continue . . . the choice between options A and B is awkward: publish something you don't really want to publish something, or decide to reject a paper largely on theoretical grounds.

But there's a third choice. *Option C,* which we'll get to it in a moment. But first, why did JPSP publish such a ridiculous paper? Here are some good, or at least reasonable, motivations:
– *Fairness.* Psychology journals routinely were publishing articles that were just as bad on other topics, so it doesn't seem fair to reject Bem's article just because its theory is implausible.
– *Open-mindedness; avoidance of censorship.* The very implausibility of Bem's theories could be taken as a reason for publishing his article: maybe it's appropriate to bend over backward to give exposure to theories that we don't really believe. The only trouble with this motivation is that there are *so many* implausible theories out there: if JPSP gives space to all of them, there will be no space left for mainstream psychology, what with all the articles about auras, ghosts, homeopathy, divine intervention, reincarnation, alien abductions, and so forth. Avoidance-of-censorship is an admirable principle, but in practice, some well-connected fringe theories seem to get special treatment. For another example, medical journals from time to time publish articles on the effectiveness of intercessory prayer, which typically seem to get more publicity than their inevitable follow-up failed replications. Surveys that 45% of Americans say they believe in ghosts, 65% believe in the supernatural, and 65% believe in God. There must be an academic literature on ghosts. It just doesn't get the respect of mainstream psychology journals.

– What if it's real? Stranger phenomena than ESP have been found in science. So another reason for publishing a paper such as Bem's is that it's possibly the scoop of the century. High-risk, high-reward.

Ok, now, here it is . . . what JPSP could have done:

Option C. Don't publish Bem's *article*. Publish his *data*. His raw data. Raw raw raw. All of it, along with a complete description of the data collection and experimental protocols, and enough computer code to allow outsiders to do whatever reanalyses they want. And then, if you must insist, you can also include Bem's article as a speculative document to be included in the supplementary material.

Our proposal—which JPSP could've done in 2010, had "just publish the raw data" been considered a live option at the time—flips the standard scheme of scientific publication. The usual way things go is to publish a highly polished article making strong conclusions, along with statistical arguments all pointing in the direction of said conclusions—basically, an expanded version of that five-paragraph essay you learned about in high school—and then, as an aside, some additional data summaries might appear in an online supplement. And, if you're really lucky, the raw data are in some repository somewhere, but that almost never happens.

Here we're saying the opposite: to the extent there's news in a psychology experiment, the news comes from the design and data collection (which should be described in complete detail) and in the data. *That's* what's important. The analysis and the write-up are the afterthoughts. Given the data, anyone should be able to do the analysis. Now apply this to the Cornell ESP research. The value, if any, in the experiments comes from the data. But that was the one thing that the journal didn't publish! Instead they published pages and pages of speculations, funky theory, and selective data analysis.

Let's go back and see how Option C fits in with JPSP's motivations:

– Fairness. Publishing Bem's data is fair, and the journal could do the same for any other research projects that it deems to be of sufficient quality and importance.

– Open-mindedness; avoidance of censorship. Again, what better argument can be offered to the skeptics than the raw data? That's the least censored thing possible.

– What if it's real? If the postulated phenomenon is, or could be, real, we want as many eyes on the data as possible. Who knows what could be learned? The very importance of the topic, which motivates publication, should also motivate full data sharing.

OK, it may be too much of a purist position to ask the journal to *only* publish the raw data and code; they'd also want to publish some basic analyses showing the key patterns in the data. But the focus should be on the data, not the statistical analysis.

Option C is not a panacea and it is not intended to resolve all the problems of scientific publication of "big if true" result. In particular, journals still have to decide what to publish, what to reject, and when to request revision. The difference is in what gets published. Or, to be more precise, in what aspects of the publication are considered necessary and which are optional. For the ESP paper as published in JPSP, the writeup, the bold claims, and the statistically significant p-values were necessary; the data were

optional. We would switch that around. But it wouldn't go that way for every project. Some projects have primary value in their data; for others, it's the analysis or the theory that are most important.

In the example of the ESP study, if anything's valuable it's the data. Publishing the data would get the journal off the hook regarding fairness, open-mindedness, and not missing a scoop, while enabling others to move on reanalyses right away, and *without* saddling the journal with an embarrassing endorsement of a weak theory that, it turns out, was not really supported by data at all.

All this is conditional on JPSP's decision to not simply reject Bem's paper. The editors of the journal had the option to reject the paper when it was submitted, and they chose not to do so. The point of the present article is that if, for whatever reason, you don't want to reject such a paper, there are other alternatives than simply publishing it as is.   We prefer the alternative of publishing the data without the conclusions—rather than what was actually done, which was to publish the conclusions without the data.

If the data are deemed potentially important, it would not be required that some exciting conclusion be drawn. This is similar to preregistration, where you get the article accepted for publication based on the research design, with the analysis coming later. But here the idea is more general. The article in the data-focused journal is not just a list of links. It would have to demonstrate the importance of the research and situate it in the literature, just like any other scientific article. The point is that if you're an experimenter with cool ideas and cool data, you can share all that with the world, without having to follow the current standard practice of (a) making strong conclusions not really supported by the data, and (b) hiding the raw data so nobody else can reanalyze.

References

Eggerson, L. (2010). Lancet retracts 12-year-old article linking autism to MMR vaccines. *Canadian Medical Association Journal* **182**, E199-E200.

Engber, D. (2017). Daryl Bem proved ESP is real. Which means science is broken. *Slate*, 7 June. https://slate.com/health-and-science/2017/06/daryl-bem-proved-esp-is-real-showed-science-is-broken.html

Gelman, A., and Loken, E. (2014). The statistical crisis in science. *American Scientist* **102**, 460-465.

McKay, B., Bar-Natan, D., Bar-Hillel, M., and Kalai, G. (1999). Solving the Bible code puzzle. *Statistical Science* **14**, 150-173.