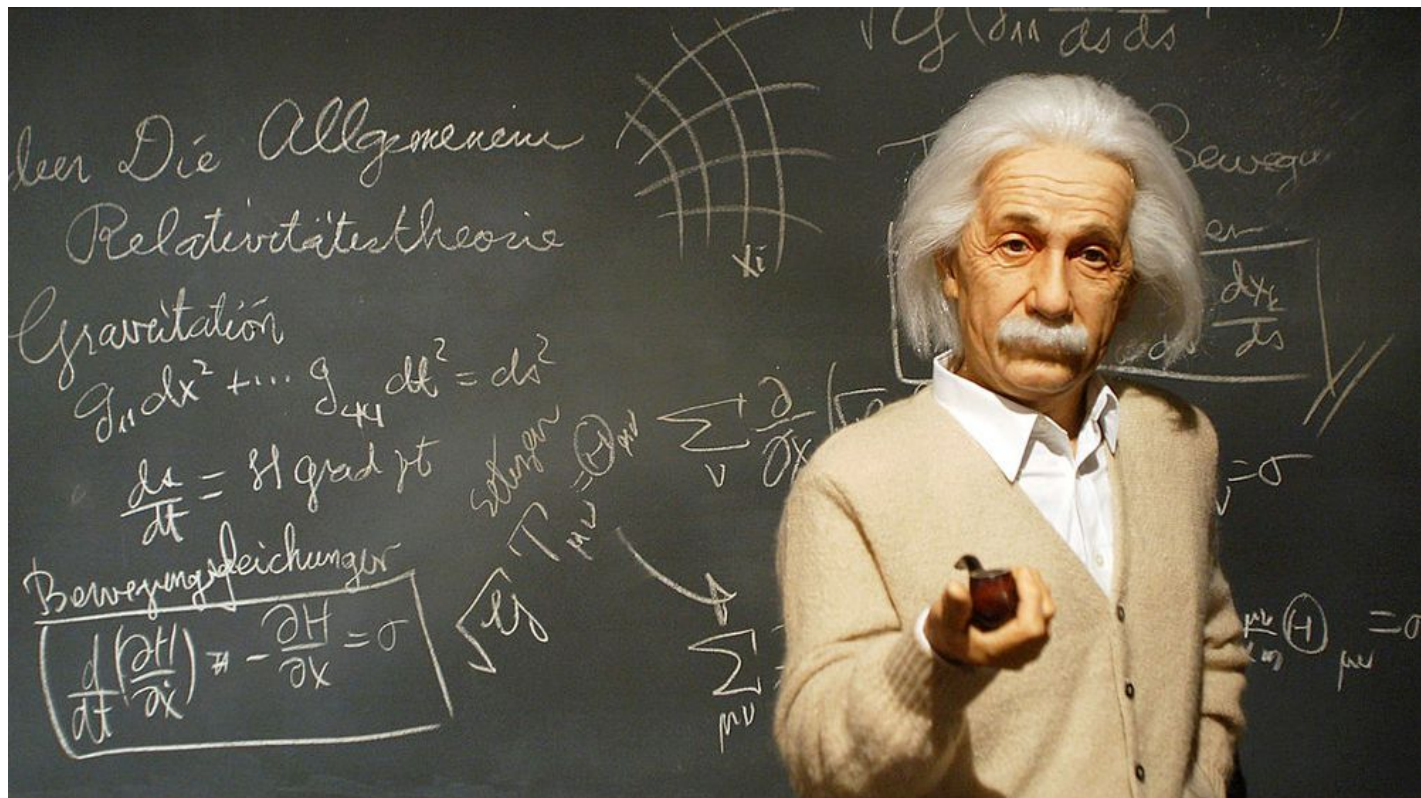


Peer Review Is Science's Wheel of Misfortune

The system for evaluating the quality of research papers works little better than flipping coins. Budding scholars pay the price.

By [Ariel Procaccia](#)

January 29, 2020, 11:00 AM EST



Peerless. Photographer: Chung Sung-Jun/Getty Images

In 1936, Albert Einstein submitted a paper, coauthored with his assistant Nathan Rosen, to the journal *Physical Review*. A month later, he received a critical report from an anonymous reviewer together with a polite request from the journal's editor to address it. Outraged by what was apparently his one and only brush with peer review, Einstein wrote back: "We (Mr. Rosen and I) had sent you our manuscript for publication and had not authorized you to show it to specialists before it is printed. I see no reason to address the – in any case erroneous – comments of your anonymous expert."

Modern scientists can only marvel at Einstein's contempt for peer review. Over the years, this process has become so central to scientific publishing that nowadays even the most distinguished researchers must regularly subject themselves to its trials and tribulations.

We can, however, easily identify with Einstein's disregard for the opinions of the so-called expert. ¹ The same time-honored sentiment drives the popular Facebook group "[Reviewer 2 Must Be Stopped](#)," which is devoted to the proposition that reviewers are too often ignorant, careless, petty or downright evil.

Unfortunately, peer review has problems that run deeper than the quality of any particular reviewer. The process is inconsistent and subjective to the degree that – in the words of Richard Smith, a former editor of the *British Medical Journal* – it's "something of a lottery." Smith wrote that Robbie Fox, a one-time editor of the *Lancet*, went so far as to question "whether anybody would notice if he were to swap the piles marked 'publish' and 'reject.'" There's a mountain of evidence that these claims aren't far from the truth.

The situation is especially grim in artificial intelligence, where most impactful publications appear in conference proceedings. Every year, each of several large conferences in the field receives thousands of submissions in a single day, which are then reviewed simultaneously by a “program committee” consisting of thousands of volunteers. It’s obvious that enforcing consistency at this scale is all but impossible.

Still, AI researchers were shocked by the results of an experiment conducted in 2014 by the organizers of the influential Conference on Neural Information Processing Systems. A portion of the submissions 2 were evaluated by two different committees, which made independent decisions to accept or reject. It turned out that 57% of the papers accepted by one committee were rejected by the other. That’s unnervingly close to what you’d expect from purely random selection.

Even the most alarming cases – papers that are blatantly wrong or fraudulent – are rarely caught in the peer review net. One of the most egregious examples is that of Jan Hendrik Schoen, a German physicist who published a slew of supposedly groundbreaking – but actually fraudulent – papers in the early 2000s. He was exposed when colleagues who were trying to build on his work noticed duplicated figures in one of his papers, leading to discoveries of additional anomalies and ultimately a full-blown investigation. In the aftermath, dozens of Schoen’s meticulously peer-reviewed papers were retracted, including an eye-popping total of 16 published in two of the most prestigious journals, Science and Nature.

The Schoen scandal mainly serves as a cautionary tale, but it also hints at why the scientific enterprise is so successful despite the shortcomings of peer review. Publication is just one part of a much larger process in which important papers are identified and then heavily scrutinized by the relevant scientific community. That’s doubly true in today’s scientific ecosystem, where online preprint repositories like arXiv make it possible for papers to achieve widespread fame or notoriety before they’re even submitted for publication.

My concern, then, is not for the integrity of science, but for the welfare of scientists.

The question of how many papers a scientist published, and where, plays a huge role in decisions about hiring, promotion, funding and – in disciplines like computer science – even admission into Ph.D. programs. A scientist’s career may depend on whether a few reviewers choose to accept or reject a single paper.

To receive tenure at a leading economics department, for example, candidates are expected to have published two or three papers in the discipline’s most prestigious journals, imaginatively called the “top 5.” Three of these journals famously rejected “The Market for Lemons,” a seminal paper that upended economic thinking and won its author a Nobel Prize, with one reviewer complaining, “If this paper was correct, economics would be different.”

It seems paradoxical that scientists – ostensible paragons of evidence-based reasoning – would give such weight to the outcomes of peer review, despite the growing evidence of the system’s limitations. One reason is laziness: nothing’s easier than skimming through a colleague’s list of publications and noting where they appeared. But another may well be that relatively few scientists recognize just how flawed peer review is. It’s up to universities and academic associations, therefore, to examine the evidence and initiate an honest discussion of this question: Assuming the way in which we evaluate papers stays fundamentally the same, how should we evaluate each other?

Scientists should also work on solutions to the problems of peer review, as many are already doing. My own contribution to this effort is reported in a recent manuscript coauthored with two former colleagues at Carnegie Mellon University, Ritesh Noothigattu and Nihar Shah. I am especially fond of this footnote: “Even papers about peer review are subject to peer review, the irony of which has not escaped us.” That irony, however, was apparently lost on our esteemed peers, who have thrice rejected the paper. To paraphrase a great scientist, I see no reason to address the – in any case erroneous – comments of these anonymous experts.

¹ Amusingly, in this case the reviewer was actually right.

2 In 2014, the conference received “only” 1,678 papers. By 2019, the number of submissions had skyrocketed to 6,743.

This column does not necessarily reflect the opinion of Bloomberg LP and its owners.

To contact the author of this story:

Ariel Procaccia at arielpro@seas.harvard.edu

To contact the editor responsible for this story:

Jonathan Landman at jlandman4@bloomberg.net

Ariel Procaccia is Gordon McKay Professor of Computer Science at Harvard University. His areas of expertise include artificial intelligence, theoretical computer science and algorithmic game theory.

[Read more opinion](#)