

Close

THE CONVERSATION

Academic rigor, journalistic flair



Experiment design affects the quality of the results. IAEA Seibersdorf Historical Images, CC BY-SA

Why isn't science better? Look at career incentives

September 20, 2016 8.01pm EDT

There are often substantial gaps between the idealized and actual versions of those people whose work involves providing a social good. Government officials are supposed to work for their constituents. Journalists are supposed to provide unbiased reporting and penetrating analysis. And scientists are supposed to relentlessly probe the fabric of reality with the most rigorous and skeptical of methods.

All too often, however, what should be just isn't so. In a number of scientific fields, published findings turn out not to replicate, or to have smaller effects than, what was initially purported. Plenty of science does replicate – meaning the experiments turn out the same way when you repeat them – but the amount that doesn't is too much for comfort.

Much of science is about identifying relationships between variables. For example, how might certain genes increase the risk of acquiring certain diseases, or how might certain parenting styles influence children's emotional development? To our disappointment, there are no tests that allow us to perfectly sort true associations from spurious ones. Sometimes we get it wrong, even with the most rigorous methods.

But there are also ways in which scientists increase their chances of getting it wrong. Running studies with small samples, mining data for correlations and forming hypotheses to fit an experiment's results

Author



Paul Smaldino

Assistant Professor of Cognitive and Information Sciences, University of California, Merced

after the fact are just some of the ways to increase the number of false discoveries.

It's not like we don't know how to do better. Scientists who study scientific methods have known about **feasible remedies for decades**. Unfortunately, their advice often falls on deaf ears. Why? Why aren't scientific methods better than they are? In a word: incentives. But perhaps not in the way you think.

Incentives for 'good' behavior

In the 1970s, psychologists and economists began to point out the danger in relying on quantitative measures for social decision-making. For example, when public schools are evaluated by students' performance on standardized tests, teachers respond by teaching "to the test" – at the expense of broader material more important for critical thinking. In turn, the test serves largely as a measure of how well the school can prepare students for the test.

We can see this principle – often summarized as “when a measure becomes a target, it ceases to be a good measure” – playing out in the realm of research. Science is a competitive enterprise. There are **far more credentialed scholars and researchers** than there are university professorships or comparably prestigious research positions. Once someone acquires a research position, there is additional competition for tenure, grant funding, and support and placement for graduate students. Due to this competition for resources, scientists must be evaluated and compared. How do you tell if someone is a good scientist?

An oft-used metric is the number of publications one has in peer-reviewed journals, as well as the status of those journals (along with related metrics, such as the *h-index*, which purports to measure the rate at which a researcher's work is cited by others). Metrics like these make it straightforward to compare researchers whose work may otherwise be quite different. Unfortunately, this also makes these numbers susceptible to exploitation.

If scientists are motivated to publish often and in high-impact journals, we might expect them to actively try to game the system. And certainly, some do – as seen in recent high-profile cases of scientific fraud (including in physics, social psychology and clinical pharmacology). If malicious fraud is the prime concern, then perhaps the solution is simply heightened vigilance.

However, most scientists are, I believe, genuinely interested in learning about the world, and honest. The problem with incentives is they can shape cultural norms without any intention on the part of individuals.

Cultural evolution of scientific practices

In a recent paper, anthropologist Richard McElreath and I considered the incentives in science through the lens of cultural evolution, an emerging field that draws on ideas and models from evolutionary biology, epidemiology, psychology and the social sciences to understand cultural organization and change.

In our analysis, we assumed that methods associated with greater success in academic careers will, all else equal, tend to spread. The spread of more successful methods requires no conscious evaluation of how scientists do or do not “game the system.”

Recall that publications, particularly in high-impact journals, are the currency used to evaluate decisions related to hiring, promotions and funding. Studies that show large and surprising associations tend to be favored for publication in top journals, while small, unsurprising or complicated results are more difficult to publish.

But most hypotheses are probably wrong, and performing rigorous tests of novel hypotheses (as well as coming up with good hypotheses in the first place) takes time and effort. Methods that boost false positives (incorrectly identifying a relationship where none exists) and overestimate effect sizes will, on average, allow their users to publish more often. In other words, when novel results are incentivized, methods that produce them – by whatever means – at the fastest pace will become implicitly or explicitly encouraged.

Over time, those shoddy methods will become associated with success, and they will tend to spread. The argument can extend beyond norms of questionable research practices to norms of misunderstanding, if those misunderstandings lead to success. For example, despite over a century of common usage, the *p*-value, a standard measure of statistical significance, is still widely misunderstood.

The cultural evolution of shoddy science in response to publication incentives requires no conscious strategizing, cheating or loafing on the part of individual researchers. There will always be researchers committed to rigorous methods and scientific integrity. But as long as institutional incentives reward positive, novel results at the expense of rigor, the rate of bad science, on average, will increase.

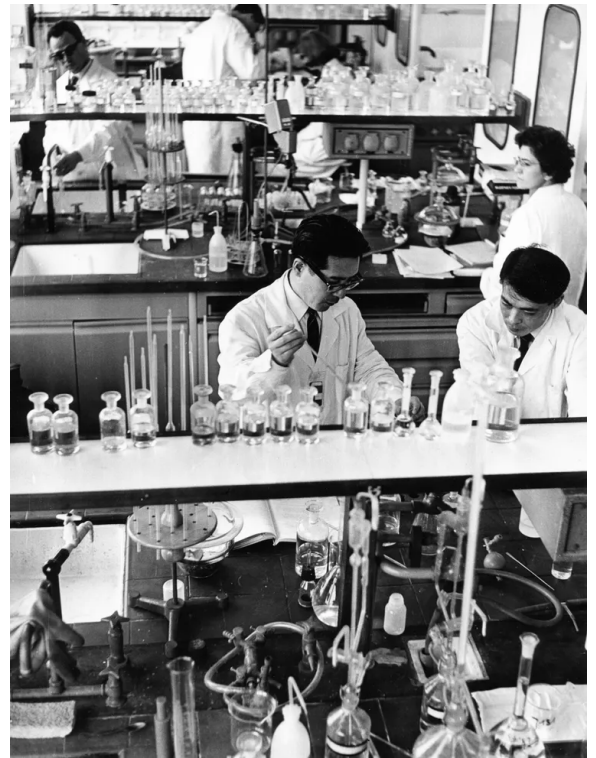
Simulating scientists and their incentives

There is ample evidence suggesting that publication incentives have been negatively shaping scientific research for decades. The frequency of the words “innovative,” “groundbreaking” and “novel” in biomedical abstracts increased by 2,500 percent or more over the past 40 years. Moreover, researchers often don't report when hypotheses fail to generate positive results, lest reporting such failures hinders publication.



We reviewed statistical power in the social and behavioral science literature. Statistical power is a quantitative measurement of a research design's ability to identify a true association when present. The simplest way to increase statistical power is to increase one's sample size – which also lengthens the time needed to collect data. Beginning in the 1960s, there have been repeated outcries that statistical power is far too low. Nevertheless, we found that statistical power, on average, has not increased.

The evidence is suggestive, but it is not conclusive. To more systematically demonstrate the logic of our argument, we



Scientists work within a culture of research. IAEA, CC BY-SA

There doesn't need to be anything nefarious going on for scientists to stick with the suboptimal methods that help them get ahead. IAEA, CC BY-SA

built a computer model in which a population of research labs studied hypotheses, only some of which were true, and attempted to publish their results.

As part of our analysis, we assumed that each lab exerted a characteristic level of "effort." Increasing effort lowered the rate of false positives, and also lengthened the time between results. As in reality, we assumed that novel positive results were easier to publish than negative results. All of our simulated labs were totally honest: they never cheated. However, labs that published more were more likely to have their methods "reproduced" in new labs – just as they would be in reality as students and postdocs leave successful labs where they trained and set up their own labs. We then allowed the population to evolve.

The result: Over time, effort decreased to its minimum value, and the rate of false discoveries skyrocketed.


And replication – while a crucial tool for generating robust scientific theories – isn't going to be science's savior. Our simulations indicate that more replication won't stem the evolution of bad science.

Taking on the system

The bottom-line message from all this is that it's not sufficient to impose high ethical standards (assuming that were possible), nor to make sure all scientists are informed about best practices (though spreading awareness is certainly one of our goals). A culture of bad science can evolve as a result of institutional incentives that prioritize simple quantitative metrics as measures of success.

There are indications that the situation is improving. Journals, organizations, and universities are increasingly emphasizing replication, open data, the publication of negative results and more holistic evaluations. Internet applications such as Twitter and YouTube allow education about best practices to propagate widely, along with spreading norms of holism and integrity.

There are also signs that the old ways are far from dead. For example, one regularly hears researchers discussed in terms of how much or where they publish. The good news is that as long as there are smart, interesting people doing science, there will always be some good science. And from where I sit, there is still quite a bit of it.

 [Scientific method](#) [Research ethics](#) [Reproducibility](#) [Scientific publishing](#) [Cultural evolution](#) [Scientific research](#)
[Replicability](#) [Science research](#) [Data mining](#) [Science careers](#) [Science publishing](#) [Research assessment](#) [Postdocs](#)
[Research incentives](#) [Research quality](#) [Research methods](#)