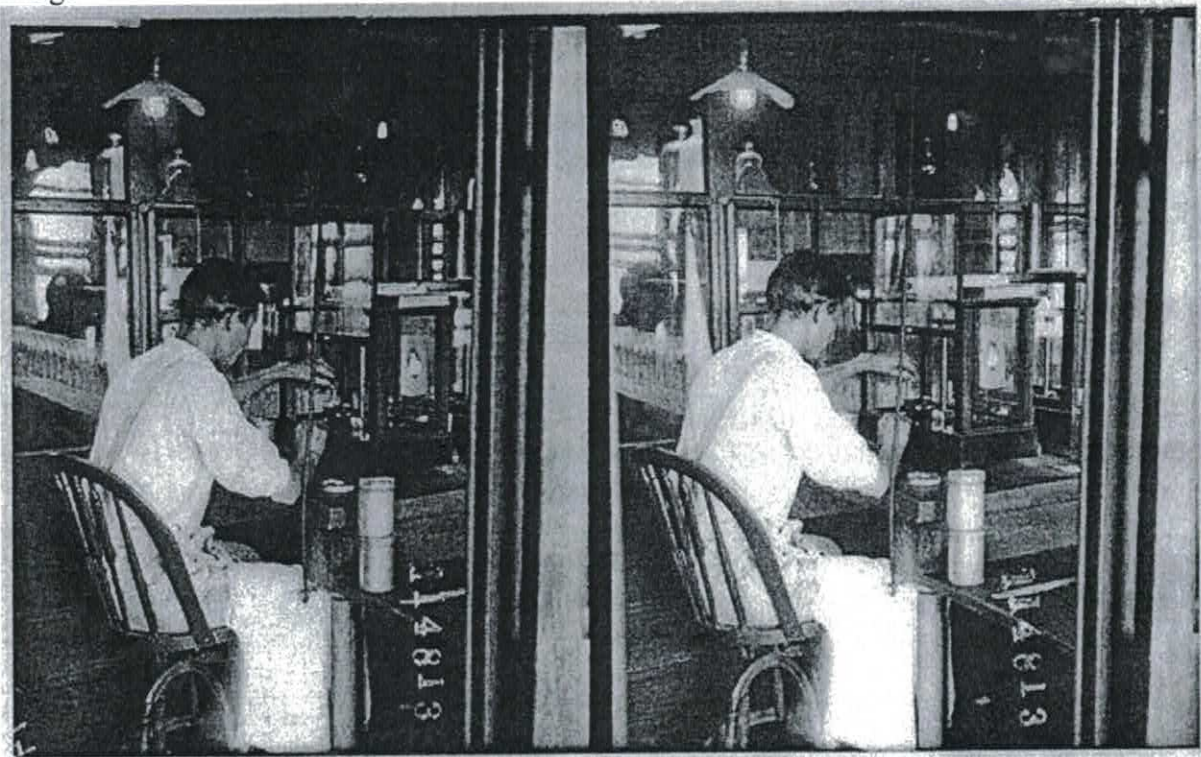


Essay: The Experiments Are Fascinating. But Nobody Can Repeat Them.

Science is mired in a “replication” crisis. Fixing it will not be easy.

Replication is important to scientists, because it means the finding might just be real. But scientists are wary of a replication crisis and fear its corrosive effects on public trust in science. Credit via Library of Congress

Image



Replication is important to scientists, because it means the finding might just be real. But scientists are wary of a replication crisis and fear its corrosive effects on public trust in science. Credit via Library of Congress

By Andrew Gelman

- Nov. 19, 2018

At this point, it is hardly a surprise to learn that even top scientific journals publish a lot of low-quality work — not just solid experiments that happen, by bad luck, to have yielded conclusions that don't stand up to replication, but poorly designed studies that had no real chance of succeeding before they were ever conducted.

Studies that were dead on arrival. We've seen lots of examples.

In 1996, a psychology study claimed that unobtrusive priming — the insertion of certain innocuous words in a quiz — could produce consistent behavioral change.

That paper got cited by other scientists a few thousand times — before failed replications many years later made it clear that this finding, and much of the subsequent literature, was little more than researchers chasing patterns in noise.

As a political scientist, my personal favorite was the survey finding in 2012 that women were 20 points more likely to support Barack Obama for president during certain days of their monthly cycle.

ADVERTISEMENT

In retrospect, this claim made no sense and was not supported by data. Even prospectively, the experiment had no chance of working: the way the study was conducted, the noise in estimating any effect — in this case, any average difference in political attitudes during different parts of the cycle — was much larger than any realistically possible signal (real result).

We see it all the time. Remember the claims that subliminal smiley faces on a computer screen can cause big changes in attitudes toward immigration? That elections are decided by college football games and shark attacks? These studies were published in serious journals or promoted in serious news outlets.

Scientists know this is a problem. In a recent paper in the journal *Nature Human Behaviour*, a team of respected economists and psychologists released the results of 21 replications of high-profile experiments.

Replication is important to scientists, because it means the finding might just be real. In this study, many findings failed to replicate. On average, results were only about half the size of the originally published claims.

Here's where it gets really weird. The lack of replication was predicted ahead of time by a panel of experts using a "prediction market," in which experts were allowed to bet on which experiments were more or less likely to — well, be real.

Similar prediction markets have been used for many years for elections, mimicking the movement of the betting line in sports. Basically, the results in this instance indicated that informed scientists were clear from the get-go that what they were reading would not hold up.

So yes, that's a problem. There has been resistance to fixing it, some of which has come from prominent researchers at leading universities. But many, if not most, scientists are aware of the seriousness of the replication crisis and fear its corrosive effects on public trust in science.

The challenge is what to do next. One potential solution is preregistration, in which researchers beginning a study publish their analysis plan before collecting their data.

Preregistration can be seen as a sort of time-reversed replication, a firewall against “data dredging,” the inclination to go looking for results when your first idea doesn’t pan out.

But it won’t fix the problem on its own.

The replication crisis in science is often presented as an issue of scientific procedure or integrity. But all the careful procedure and all the honesty in the world won’t help if your signal (the pattern you’re looking for) is small, and the variation (all the confounders, the other things that might explain this pattern) is high.

From this perspective, the crisis in science is more fundamental, and it involves moving beyond the existing model of routine discovery.

Say you wish to study the effect of a drug or an educational innovation on a small number of people. Unless the treatment is very clearly targeted to an outcome of interest (for example, a math curriculum focused on a particular standardized test), then your study is likely to be too noisy — there will too many variables — to pinpoint real effects.

If something at random does turn up and achieve statistical significance, it is likely to be a massive overestimate of any true effect. In an attempt at replication, we’re likely to see something much closer to zero.

The failed replications have been no surprise to many scientists, including myself, who have lots of experience of false starts and blind alleys in our own research.

The big problem in science is not cheaters or opportunists, but sincere researchers who have unfortunately been trained to think that every statistically “significant” result is notable.

When you read about research in the news media (and, as a taxpayer, you are indirectly a funder of research, too), you should ask what exactly is being measured, and why.

Andrew Gelman is a professor of statistics and political science at Columbia University.

A version of this article appears in print on Nov. 19, 2018, on Page D10 of the New York edition with the headline: That’s Fascinating. Now Do It Again..