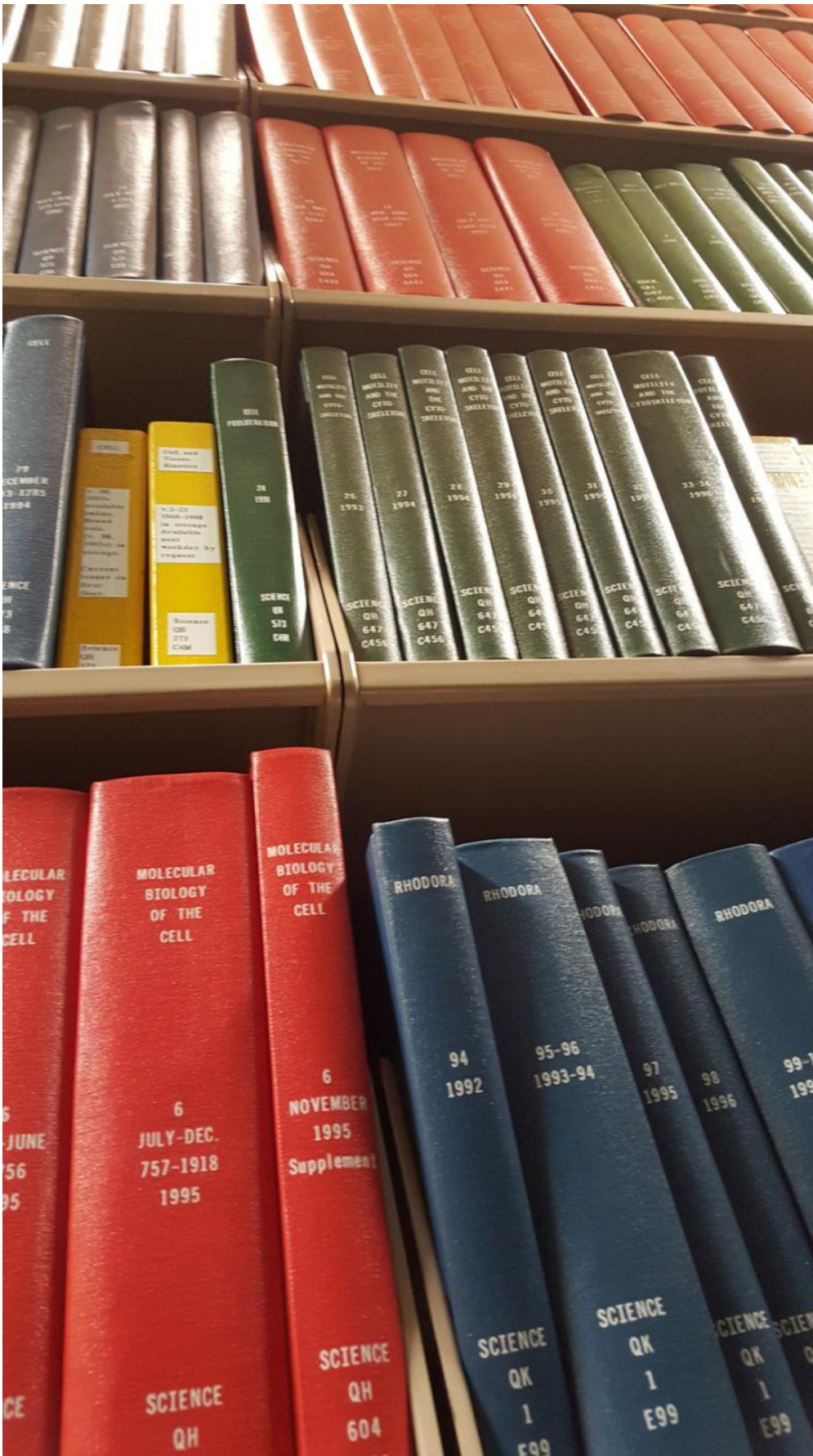


Only one-third of published psychology research is reliable – now what?

August 28 2015, by Elizabeth Gilbert And Nina Strohminger



The results are bound and shelved – but are they reproducible? Credit: Maggie Villiger, CC BY-ND

The ability to repeat a study and find the same results twice is a prerequisite for building scientific knowledge. Replication allows us to ensure empirical findings are reliable and refines our understanding of when a finding occurs. It may surprise you to learn, then, that scientists do not often conduct – much less publish – attempted replications of existing studies.

Journals prefer to publish novel, cutting-edge research. And professional advancement is determined by making new discoveries, not painstakingly confirming claims that are already on the books. As one of our colleagues recently put it, "Running replications is fine for other people, but I have better ways to spend my precious time."

Once a paper appears in a peer-reviewed journal, it acquires a kind of magical, unassailable authority. News outlets, and sometimes even scientists themselves, will cite these findings without a trace of skepticism. Such unquestioning confidence in new studies is likely undeserved, or at least premature.

A small but vocal contingent of researchers – addressing fields ranging from [physics](#) to [medicine](#) to [economics](#) – has maintained that many, [perhaps most](#), published studies are wrong. But how bad is this problem, exactly? And what features make a study more or less likely to turn out to be true?

We are two of the 270 researchers who together have just [published in](#)

[the journal Science](#) the first-ever large-scale effort trying to answer these questions by attempting to reproduce 100 previously published psychological science findings.

Attempting to re-find psychology findings

Publishing together as the [Open Science Framework](#) and coordinated by social psychologist Brian Nosek from the [Center for Open Science](#), research teams from around the world each ran a replication of a study published in three top psychology journals – *Psychological Science*; *Journal of Personality and Social Psychology*; and *Journal of Experimental Psychology: Learning, Memory, and Cognition*. To ensure the replication was as exact as possible, research teams obtained study materials from the original authors, and worked closely with these authors whenever they could.

Almost all of the original published studies (97%) had statistically significant results. This is as you'd expect – while many experiments fail to uncover meaningful results, scientists tend only to publish the ones that do.

What we found is that when these 100 studies were run by other researchers, however, only 36% reached statistical significance. This number is alarmingly low. Put another way, only around one-third of the rerun studies came out with the same results that were found the first time around. That rate is especially low when you consider that, once published, findings tend to be held as gospel.

The bad news doesn't end there. Even when the new study found evidence for the existence of the original finding, the magnitude of the effect was much smaller—half the size of the original, on average.

One caveat: just because something fails to replicate doesn't mean it isn't

true. Some of these failures could be due to luck, or poor execution, or an incomplete understanding of the circumstances needed to show the effect (scientists call these "moderators" or "boundary conditions"). For example, having someone practice a task repeatedly might improve their memory, but only if they didn't know the task well to begin with. In a way, what these replications (and failed replications) serve to do is highlight the inherent uncertainty of any single study – original or new.

More robust findings more replicable

Given how low these numbers are, is there anything we can do to predict the studies that will replicate and those that won't? The results from this [Reproducibility Project](#) offer some clues.

There are two major ways that researchers quantify the nature of their results. The first is a p-value, which estimates the probability that the result was arrived at purely by chance and is a false positive.

(Technically, the p-value is the chance that the result, or a stronger result, would have occurred even when there was no real effect.)

Generally, if a statistical test shows that the p-value is lower than 5%, the study's results are considered "significant" – most likely due to actual effects.

Another way to quantify a result is with an effect size – not how reliable the difference is, but how big it is. Let's say you find that people spend more money in a sad mood. Well, *how much* more money do they spend? This is the effect size.

We found that the smaller the original study's p-value and the larger its effect size, the more likely it was to replicate. Strong initial statistical evidence was a good marker of whether a finding was reproducible.

Studies that were rated as more challenging to conduct were less likely to

replicate, as were findings that were considered surprising. For instance, if a study shows that reading lowers IQs, or if it uses a very obscure and unfamiliar methodology, we would do well to be skeptical of such data. Scientists are often rewarded for delivering results that dazzle and defy expectation, but extraordinary claims require extraordinary evidence.

Although our replication effort is novel in its scope and level of transparency – the methods and data for all replicated studies are [available online](#) – they are consistent with previous work from other fields. Cancer biologists, for instance, have reported replication rates as low as [11%-25%](#).

We have a problem. What's the solution?

Some conclusions seem warranted here.

We must stop treating single studies as unassailable authorities of the truth. Until a discovery has been thoroughly vetted and repeatedly observed, we should treat it with the measure of skepticism that scientific thinking requires. After all, the truly scientific mindset is critical, not credulous. There is a place for breakthrough findings and cutting-edge theories, but there is also merit in the slow, systematic checking and refining of those findings and theories.

Of course, adopting a skeptical attitude will take us only so far. We also need to provide incentives for reproducible science by rewarding those who conduct replications and who conduct replicable work. For instance, at least one top journal has begun to give special "[badges](#)" to articles that make their data and materials available, and the Berkeley Initiative for Transparency in the Social Sciences has [established a prize](#) for practicing more transparent social science.

Better research practices are also likely to ensure higher replication

rates. There is already evidence that [taking certain concrete steps](#) – such as making hypotheses clear prior to data analysis, openly sharing materials and data, and following transparent reporting standards – decreases false positive rates in published studies. Some funding organizations are already demanding [hypothesis registration](#) and [data sharing](#).

Although perfect replicability in published papers is an unrealistic goal, current [replication](#) rates are unacceptably low. The first step, as they say, is admitting you have a problem. What scientists and the public now choose to do with this information remains to be seen, but our collective response will guide the course of future scientific progress.

This story is published courtesy of [The Conversation](#) (under Creative Commons-Attribution/No derivatives).

Source: The Conversation

Citation: Only one-third of published psychology research is reliable – now what? (2015, August 28) retrieved 9 April 2024 from <https://phys.org/news/2015-08-one-third-published-psychology-reliable.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.