

# Science Needs a Solution for the Temptation of Positive Results

Aaron E. Carroll

MAY 29, 2017

A few years back, scientists at the biotechnology company Amgen set out to replicate 53 landmark studies that argued for new approaches to treat cancers using both existing and new molecules. They were able to replicate the findings of the original research only 11 percent of the time.

Science has a reproducibility problem. And the ramifications are widespread.

These 53 papers were published in high-profile journals, and the 21 that were published in the highest-impact journals were cited an average of 231 times in subsequent work.

In 2011, Bayer pharmaceuticals reported similar reproduction work. Of the 67 projects they conducted to rerun experiments (47 of which involved cancer), only about 25 percent ended with results in line with the original findings.

It turns out that most pharmaceutical companies run these kinds of in-house validation programs regularly. They seem skeptical of findings in the published literature. Given that their valuable time and their investment of billions of dollars of research resources hinge directly on the success of projects, their concerns seem

warranted.

Unfortunately, the rest of us have not been quite so careful. More and more data show we should be. In 2015, researchers reported on their replication of 100 experiments published in 2008 in three prominent psychology journals. Psychology studies don't usually lead to much money or marketable products, so companies don't focus on checking their robustness. Yet in this experiment, research results were just as questionable. The findings of the replications matched the original studies only one-third to one-half of the time, depending on the criteria used to define "similar."

There are a number of reasons for this crisis. Scientists themselves are somewhat at fault. Research is hard, and rarely perfect. A better understanding of methodology, and the flaws inherent within, might yield more reproducible work.

The research environment, and its incentives, compound the problem. Academics are rewarded professionally when they publish in a high-profile journal. Those journals are more likely to publish new and exciting work. That's what funders want as well. This means there is an incentive, barely hidden, to achieve new and exciting results in experiments.

Some researchers may be tempted to make sure that they achieve "new and exciting results." This is fraud. As much as we want to believe it never happens, it does. Clearly, fabricated results are not going to be replicable in follow-up experiments.

But fraud is rare. What happens far more often is much more subtle. Scientists are more likely to try to publish positive results than negative ones. They are driven to conduct experiments in such a way as to make it more likely to achieve positive results. They sometimes measure many outcomes and report only the ones that showed bigger results. Sometimes they change things just enough to get a crucial measure of probability — the p value — down to 0.05 and claim significance. This is known as p-hacking.

How we report on studies can also be a problem. Even some studies reported

on by newspapers (like this one) fail to hold up as we might hope.

This year, a study looked at how newspapers reported on research that associated a risk factor with a disease, both lifestyle risks and biological risks. For initial studies, newspapers didn't report on any null findings, meaning those that had results without expected outcomes. They rarely reported null findings even when they were confirmed in subsequent work.

Fewer than half of the "significant" findings reported on by newspapers were later backed by other studies and meta-analyses. Most concerning, while 234 articles reported on initial studies that were later shown to be questionable, only four articles followed up and covered the refutations. Often, the refutations are published in lower-profile journals, and so it's possible that reporters are less likely to know about them. Journal editors may be as complicit as newspaper editors.

The good news is that the scientific community seems increasingly focused on solutions. Two years ago, the National Institutes of Health began funding efforts to create educational modules to train scientists to do more reproducible research. One of those grants allowed my YouTube show, Healthcare Triage, to create videos to explain how we could improve both experimental design and the analysis and reporting of research. Another grant helped the Society for Neuroscience develop webinars to promote awareness and knowledge to enhance scientific rigor.

The Center for Open Science, funded by both the government and foundations, has been pushing for increased openness, integrity and reproducibility of research. They, along with experts and even journals, have pushed for the preregistration of studies so that the methods of research are more transparent and the analyses are free of bias or alteration. They conducted the replication study of psychological research, and are now doing similar work in cancer research.

But true success will require a change in the culture of science. As long as the academic environment has incentives for scientists to work in silos and hoard their data, transparency will be impossible. As long as the public demands a constant stream of significant results, researchers will consciously or subconsciously push

their experiments to achieve those findings, valid or not. As long as the media hypes new findings instead of approaching them with the proper skepticism, placing them in context with what has come before, everyone will be nudged toward results that are not reproducible.

For years, financial conflicts of interest have been properly identified as biasing research in improper ways. Other conflicts of interest exist, though, and they are just as powerful — if not more so — in influencing the work of scientists across the country and around the globe. We are making progress in making science better, but we've still got a long way to go.

At first sight, the solution might seem obvious: each community should make a step toward the other. Yet it would be a mistake to consider this an easy task. The obstacles are many. At present, a significant number of philosophers disdain science or don't see the relevance of science to their work. Reading science is indispensable for the practice of philosophy of science, but reading philosophy can also constitute a great source of inspiration for researchers as illustrated by some of the examples above. For example, journal clubs where both science and philosophy contributions are discussed constitute an efficient way to integrate philosophy and science. We need a reinvigoration of science at all levels, one that returns to us the benefits of close ties with philosophy.

Acknowledgments. the positive conception of perfectionism was indeed. associated with more positive characteristics than the. negative conception. Because the aim was to review evidence. for the view that a positive and a negative form of per. - 297. Positive conceptions of perfectionism. Need for approval I'm concerned with whether or not other people approve of my actions. I compare my work to others and often feel inadequate. I often don't say anything, because I'm scared I might say the wrong thing.

Furthermore, given that positive results are published, whereas negative data will struggle, it is extremely difficult to correct the scientific record for false positives; controversial studies that conflict with or cannot reproduce previously published studies are seldom given space in peer-reviewed journals. For the funding bodies this holds an additional benefit: a grant funding research that resulted in negative data would then still result in publications and shared information. New Negatives in Plant Science – a pilot journal. To raise this important issue, and put the spotlight on negative and controversial data, the journal New Negatives in Plant Science was launched in 2014. It is an open access journal that publishes both research articles and commentaries.

Results: When the infection rate of the close contacts and the sensitivity and specificity of reported results were taken as the point estimates, the positive predictive value of the active screening was only 19.67%, in contrast, the false-positive rate of positive results was 80.33%. The multivariate-probabilistic sensitivity analysis results supported the base-case findings, with a 75% probability for the false-positive rate of positive results over 47%. I initially found this to be inspiring as it may mean that many cases where people are sick may actually just cases of influenza.