



# Scientific Regression: Big Science Is Broken!

TN Note: An observant and critical analysis of the state of science indirectly reveals the source of error in false science like that associated with global warming. The last sentence is a good conclusion: "The greatest friends of the Cult of Science are the worst enemies of science's actual practice."

The problem with science is that so much of it simply isn't. Last summer, the Open Science Collaboration announced that it had tried to replicate one hundred published psychology experiments sampled from three of the most prestigious journals in the field. Scientific claims rest on the idea that experiments repeated under nearly identical conditions ought to yield approximately the same results, but until very recently, very few had bothered to check in a systematic way whether this was actually the case. The OSC was the biggest attempt yet to check a field's results, and the most shocking. In many cases, they had used original experimental materials, and sometimes even performed the experiments under the guidance of the original researchers. Of the studies that had originally

reported positive results, an astonishing 65 percent failed to show statistical significance on replication, and many of the remainder showed greatly reduced effect sizes.

Their findings made the news, and quickly became a club with which to bash the social sciences. But the problem isn't just with psychology. There's an unspoken rule in the pharmaceutical industry that half of all academic biomedical research will ultimately prove false, and in 2011 a group of researchers at Bayer decided to test it. Looking at sixty-seven recent drug discovery projects based on preclinical cancer biology research, they found that in more than 75 percent of cases the published data did not match up with their in-house attempts to replicate. These were not studies published in fly-by-night oncology journals, but blockbuster research featured in *Science*, *Nature*, *Cell*, and the like. The Bayer researchers were drowning in bad studies, and it was to this, in part, that they attributed the mysteriously declining yields of drug pipelines. Perhaps so many of these new drugs fail to have an effect because the basic research on which their development was based isn't valid.

When a study fails to replicate, there are two possible interpretations. The first is that, unbeknownst to the investigators, there was a real difference in experimental setup between the original investigation and the failed replication. These are colloquially referred to as "wallpaper effects," the joke being that the experiment was affected by the color of the wallpaper in the room. This is the happiest possible explanation for failure to reproduce: It means that both experiments have revealed facts about the universe, and we now have the opportunity to learn what the difference was between them and to incorporate a new and subtler distinction into our theories.

The other interpretation is that the original finding was false. Unfortunately, an ingenious statistical argument shows that this second interpretation is far more likely. First articulated by John Ioannidis, a professor at Stanford University's School of Medicine, this argument proceeds by a simple application of Bayesian statistics. Suppose that there are a hundred and one stones in a certain field. One of them has a diamond inside it, and, luckily, you have a diamond-detecting device that

advertises 99 percent accuracy. After an hour or so of moving the device around, examining each stone in turn, suddenly alarms flash and sirens wail while the device is pointed at a promising-looking stone. What is the probability that the stone contains a diamond?

Most would say that if the device advertises 99 percent accuracy, then there is a 99 percent chance that the device is correctly discerning a diamond, and a 1 percent chance that it has given a false positive reading. But consider: Of the one hundred and one stones in the field, only one is truly a diamond. Granted, our machine has a very high probability of correctly declaring it to be a diamond. But there are many more diamond-free stones, and while the machine only has a 1 percent chance of falsely declaring each of them to be a diamond, there are a hundred of them. So if we were to wave the detector over every stone in the field, it would, on average, sound twice—once for the real diamond, and once when a false reading was triggered by a stone. If we know only that the alarm has sounded, these two possibilities are roughly equally probable, giving us an approximately 50 percent chance that the stone really contains a diamond.

This is a simplified version of the argument that Ioannidis applies to the process of science itself. The stones in the field are the set of all possible testable hypotheses, the diamond is a hypothesized connection or effect that happens to be true, and the diamond-detecting device is the scientific method. A tremendous amount depends on the proportion of possible hypotheses which turn out to be true, and on the accuracy with which an experiment can discern truth from falsehood. Ioannidis shows that for a wide variety of scientific settings and fields, the values of these two parameters are not at all favorable.

For instance, consider a team of molecular biologists investigating whether a mutation in one of the countless thousands of human genes is linked to an increased risk of Alzheimer's. The probability of a randomly selected mutation in a randomly selected gene having precisely that effect is quite low, so just as with the stones in the field, a positive finding is more likely than not to be spurious—unless the experiment is unbelievably successful at sorting the wheat from the chaff. Indeed, Ioannidis finds that in many cases, approaching even 50 percent true

positives requires unimaginable accuracy. Hence the eye-catching title of his paper: “Why Most Published Research Findings Are False.”

Read full story here...