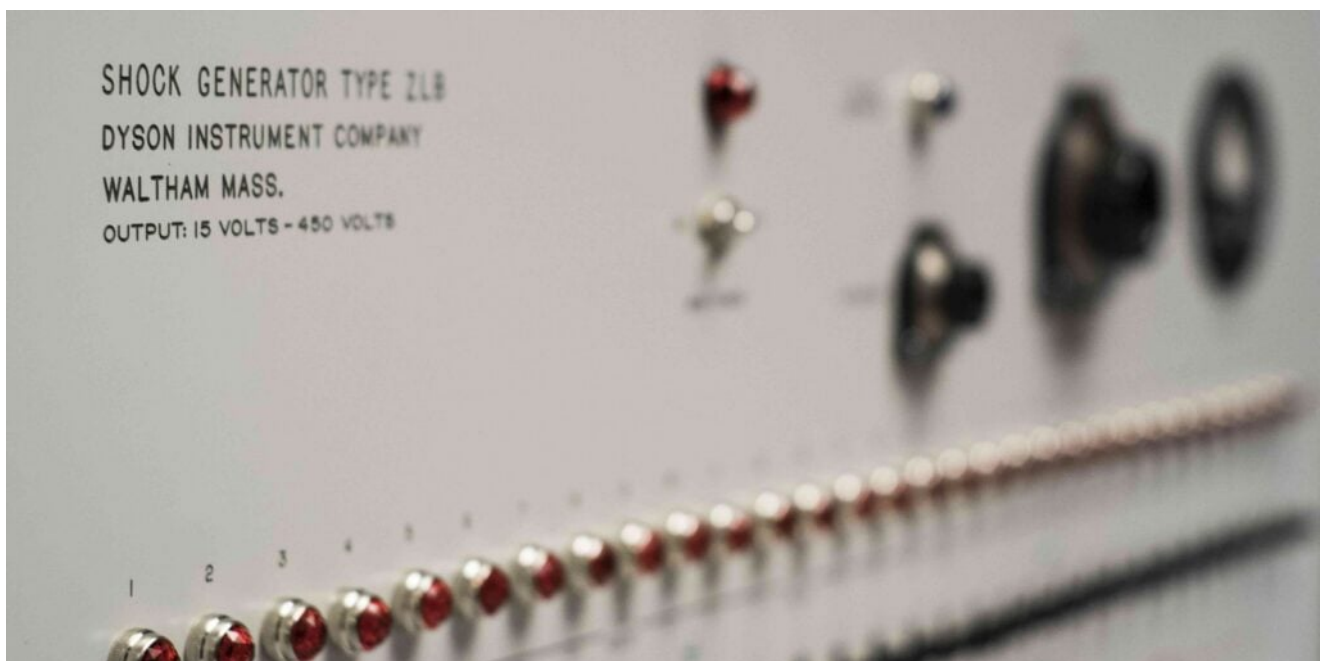


# The Replication Conundrum

By Theodore Dalrymple

Until quite recently—I cannot put an exact date on it—I assumed that everything published in scientific journals was, if not true, at least not deliberately untrue. Scientists might make mistakes, but they did not cheat, plagiarise, falsify, or make up their results. For many years as I opened a medical journal, the possibility simply that it contained fraud did not occur to me. Cases such as those of the Piltdown Man, a hoax in which bone fragments found in the Piltdown gravel pit were claimed to be those of the missing link between ape and man, were famous because they were dramatic but above all because they were rare, or assumed to be such.



Such naivety is no longer possible: instances of dishonesty have become much more frequent, or at least much more publicised. Whether the real incidence of scientific fraud has increased is difficult to say. There is probably no way to estimate the incidence of such fraud in the past by which a proper comparison can be made.

There are, of course, good reasons why scientific fraud should

have increased. The number of practising scientists has exploded; they are in fierce competition with one another; their careers depend to a large extent on their productivity as measured by publication. The difference between what is ethical and unethical has blurred. They cite themselves, they recycle their work, they pay for publication, they attach their names to pieces of work they have played no part in performing and whose reports they have not even read, and so forth. As new algorithms are developed to measure their performance, they find new ways to play the game or to deceive. And all this is not even counting commercial pressures.

Furthermore, the general level of trust in society has declined. Are our politicians worse than they used to be, as it seems to everyone above a certain age, or is it that we simply know more about them because the channels of communication are so much wider? At any rate, trust in authority of most kinds has declined. Where once we were inclined to say, "It must be true because I read it in a newspaper," we are now inclined to say, "It must be untrue because I read it in a newspaper."

Quite often now I look at a blog called *Retraction Watch* which, since 2010, has been devoted to tracing and encouraging retraction of flawed scientific papers, often flawed for discreditable reasons. Such reasons are various and include research performed on subjects who have not given proper consent. This is not the same as saying that the results of such research are false, however, and raises the question of whether it is ethical to cite results that have been obtained unethically. Whether it is or not, we have all benefited enormously from past research that would now be considered unethical.

One common problem with research is its reproducibility, or lack of it. This is particularly severe in the case of psychology, but it is common in medicine too.

Many papers in medical journals are now fundamentally epidemiological in nature. Let me give a hypothetical example. Groups of assiduous researchers have assembled a database of 5,000,000 people. (In Scandinavia, the medical records of the entire population are available for such research.) The researchers correlate, say, the self-reported consumption of bananas with a disease, let us call it bananism. They find that those who eat more than 5 bananas a week are 1.4 times more likely to suffer from bananism than those who eat fewer, even when many other factors are controlled for. What is one supposed to do with this result?

No one is ever going to reproduce the experiment. Though trying to reproduce other researchers' results is a perfectly honourable, and indeed a very useful, thing to do, the kudos attached to it is not very great. Like modern architects, scientists strive mightily to be original, therefore they add twists to the original design that make subsequent interpretations contentious. Besides, it is difficult, costly, and time-consuming to assemble population samples of 5,000,000 and ask them about their consumption of bananas.

With psychology, the difficulties are even greater because of the nature of the subject matter. Recently on *Retraction Watch*, I came across an article titled *The Replication Database: Documenting the Replicability of Psychological Science*. I quote:

*Despite its importance, replication efforts are few and far between in psychological science with many attempts failing to corroborate past findings.*

The authors have founded a database to trace efforts at replication.

This is an honourable enterprise, but it seems to me to avoid one important reason why psychological experiments are so difficult to replicate, namely the reflexive nature of the

human mind.

Let us take the late Stanley Milgram's famous experiments on obedience to authority as an example. I disregard any criticisms of Milgram's probity that have been raised; I take the experiments at face value. Certainly, their results in the wake of the Second World War were very startling. Moreover, when they were published in book form, I remember reading the book as if it were a great novel, so compelling was it.

But what now are the lessons that we can still draw from these fascinating experiments? Could we reproduce the experiments in such a way as to establish their stability and their timeless scientific validity?

The experiments would be considered unethical today because they involved gross deception of their subjects. If there had not been such deception, the experiments could not have been done. But let us suppose that the ethical objections were waived, and permission given for the experiments to be repeated.

It is extremely doubtful whether they *could* be repeated. They were carried out in the early 1960s, in social conditions very different from those of today. Apart from anything else, it is likely that a large proportion of the population that would volunteer to participate would have heard of, and possibly even know about, Milgram's original results. But even if they hadn't or didn't, so much has changed in the meantime that any difference in results might be attributable to any number of reasons, from Milgram having been mistaken in the first place, to chance, to a change in the mentalities of the population.

In other words, the problem of reproducibility in psychological science is inherent in the nature of the science itself, the more it departs from purely physiological investigation and becomes of obvious social significance. Research involving attitudinal surveys is particularly time-,

culture- and purpose-limited. Nothing is so easy, or so dangerous, as to suppose that we ourselves are models for the whole of humanity, for the whole of time.

*First published in [Law and Liberty](#)*