Thinking back on the year 2010, cognitive scientists will probably remember it as the year the Hauser affair broke out after years of rumors. That summer, a Harvard investigation committee found Marc Hauser responsible for eight cases of "scientific misconduct". The most serious problem involved a missing control condition in an experiment published in Cognition. Since the experimenters were blind to the conditions they were testing, a weird and important mistake cannot be entirely ruled out. (See [here](#).) The other seven cases involved various forms of sloppiness – whether it was guided by dishonesty is still hard to know. Then informal accusations of misconduct started pouring in, from colleagues and former students. Scientists who spoke against Hauser, like Gordon Gallup or Michael Tomasello, concentrated their attacks on various failures to replicate his results.

This post is not about Marc Hauser. Whether he was guilty of true misconduct or unintentionally produced false results matters less than the fact that we ('we' stands for us, cognitive scientists with an interest in morality, religion or culture) were fooled into accepting this work. "Accepting" is an understatement. We hyped it up, we based arguments upon it, we advertised it to philosophers, anthropologists and the greater public as an example of good science.

The situation would be a little more comfortable if we had not been serving for many years as self-appointed ambassadors of the Scientific Bushido to the social sciences and humanities.

For or against Hauser, the reactions I heard or read in the community were, basically, of two kinds. First, there were those who considered scientific fraud at that level an exceptional (or incredible) evil. Others, however, noted that few researchers could have published 130+ papers over five years while keeping all their data nicely stored and cleaned for the kind of close inspection Harvard inspectors gave Hauser's archives. They pointed out that replication failures happen all the time. And they made vague hints at colleagues (in other labs, other cities, other countries) who got away with much worse.

The righteous and the cynical tended to agree on one point. Everyone expressed concern that the Hauser affair might tarnish the public reputation of cognitive science, and eager to prevent that. This, I think, is exactly the lesson not to draw from the Hauser case.

In a well-functioning culture, the public reputation of our field should be influenced by what happened, just like the public reputation of economists was harmed by recent financial scandals. Changes in reputation following such events are noisy, nasty and unfair. That is how opinion works. Did we complain of noisy, nasty, unfair opinion when it was extolling cognitive science and ignoring other fields?

Guilty or not guilty, Marc Hauser is right on one thing: [as Helen De Cruz pointed out](#), failure to replicate famous experiments is a common problem, especially in primate studies. The scale of the problem is anything but small, and it is not something we can entirely blame on fraud, nor a problem for primatologists alone. Researchers working on replicability suggest that many experimental fields suffer from a tendency to produce results that prove less and less robust with each attempt at reproducing them. In other words, wrong (unreliable and unreplicatable) experimental findings are more numerous than most of us would think.

Ioannidis' paper on the topic, five years ago, is one of the most downloaded papers in the history of PLoS. It is all the more striking that it has been so slow in grabbing headlines. Two recent pieces, one in The New Yorker, the other in the Atlantic, have called the public's attention on the poor replicability of published experiments.

The Atlantic essay covers Ioannidis' two famous papers. The one in PLoS, in which he tries to make sense of the fact that "80 percent of non-randomized studies (by far the most common type) turn out to be wrong, as do 25 percent of supposedly gold-standard randomized trials, and as much as 10 percent of the platinum-standard large randomized trials." And the second one, showing that "Of the 49 articles, 45 claimed to have uncovered effective interventions. Thirty-four of these claims had been retested, and 14 of these, or 41 percent, had been convincingly shown to be wrong or significantly exaggerated".

The average scientist's reflex is to put the problem down to sloppiness, fraud, an urge to publish or a mix of those. The next step is a call for virtue, followed by a prayer to keep the public's attention away from the mess. The Atlantic essay reflects this attitude: it blames the poor replicability on bias, and blames the bias on ambition.

"The studies were biased," he says. "Sometimes they were overtly biased. Sometimes it was difficult to see the bias, but it was there." Researchers headed into their studies wanting certain results—and, lo and behold, they were getting them. We think of the scientific process as being objective, rigorous, and even ruthless in separating out what is true from what we merely wish to be true, but in fact it's easy to manipulate results, even unintentionally or unconsciously. "At every step in the process, there is room to distort results, a way to make a stronger claim or to select what is going to be concluded," says Ioannidis. "There is an intellectual conflict of interest that pressures researchers to find whatever it is that is most likely to get them funded."

This is an attractive perspective, because it is a moral one – and we all experienced a feeling of righteousness when reading Ioannidis (or when reading about Hauser). Also because it is intuitive – yes, dear reader, you were probably thinking of some particular sloppy, ambition-driven scientists that you dislike -, and above all because it leaves our belief in the fundamental soundness of the experimental method untouched. Science being pure, individuals need virtue to keep up with its high standards. Individuals failed, their methods did not.

Sadly, there are reasons to think that praying for individual virtue is not likely to solve the problem. Even the most impartial experiments prove abnormally difficult to replicate. Something seems to be broken in our tools, not just in our conduct. Some convincing evidence is cited in Jonah Lehrer's fascinating New Yorker piece. It comes from all the studies that have been purposefully conducted to test for replication problems. In these studies, the replication trials are built to be as similar as possible to one another, and, quite often, the researchers want the replication to work. Yet the results are astoundingly noisy, much more than one or two replications would suggest.

Another problem that individual virtue alone cannot solve is the famous file-drawer effect: the results that make it into journals need to be provocative and important. Yet chance alone produces many sexy and important results that, upon replication, turn out to be weaker, or not to exist at all. A substantial part of the scientific literature might be nothing but the sexy tip of an iceberg of randomness, most of it lying hidden in the sea of unpublished findings.

The file-drawer effect is especially destructive since it is most often combined with p-value significance testing (see here, here and here). Among other flaws, p-value significance testing (at the .05 level in most fields of cognitive science) guarantees that chance alone will give us a significant result every twenty trials or so. In this post, Cosma Shalizi explains how this is sufficient in itself to

generate spectacularly counter-intuitive results, and bolster them against refutation for some time. Of course, these results stand little chances of being replicated – but it might take a while before the community realizes that. What is more, Shalizi notes, the very presence of surprising false results that take some time to get refuted might give scientists the illusion that they are getting somewhere – since the field changes all the time – and bravely applying the scientific method – since we are bold enough to repudiate the fashionable hypotheses of ten years ago. None of this requires dishonesty or sloppiness.

Make a mistake. Acknowledge it today, it is a blunder – or a fraud. Discover it ten years later, and you may call it progress.

I am not saying that experimental science has become a machine that turns randomness into sexy falsities – or that we spend most of our time trying to interpret subtle noise. But if the million-dollars drug tests of the biggest medical reviews prove fragile, I would not bet that we, pioneers on the margins of hard science (or at the cutting edge, if you prefer), will do much better. After all, cognitive science produced more than its fair share of sexy ideas that did not quite live up to the hype they generated (Strong AI… [Universal Grammar.](.).. [Mirror Neurons.](.).).. The Hauser affair itself is not utterly without precedent (you might remember the ugly controversies surrounding Daniel Povinelli). Why would the future be different?

Whatever else we do, we are part of a growing trend in the social sciences and humanities: people with a psychological background trying to win the next generation of philosophers / anthropologists / theologians (pick your target) to the experimental method. We proudly wave the flag of Science. We have (some) money. We are vocal about the shortcomings of traditional methods – and we're lucky enough that our opponents ignore most of the flaws in ours. So we might very well succeed in turning many ethnographers, philosophers or theologians into experimentalists. Popular scientist (like Marc Hauser) have already succeeded in making experimental psychology an authoritative source of insights about ethics, religion or culture.

There is little probability that Marc Hauser, who is still running his lab, will lose tenure. He is still a member of our institute, by the way, and as far as I'm concerned, his name should stay on the members list (not be surreptitiously erased as it was elsewhere). If it eventually turns up that he was forgivably sloppy and biased, we will be glad we did not rush to conclusions. If he did fraud, then his name on the website will serve as a public reminder of our gullibility.