

A Science We Can Believe In

December 01, 2011

APS, our Board and our Members are against scientific misconduct... at least (by my estimate — more on that below) 98.03 percent of them are. Does this sound like something newsworthy enough to devote a column to? I've decided to interrupt my planned series of opinion pieces to write a bit about (mis)conduct and the practices of science. My main motivation is to offer a large dose of optimism to some younger psychological scientists who may be in danger of slipping into cynicism about our field. This column benefited from extensive help from Rumen Iliev, a postdoctoral fellow in our lab.

There are some bad vibrations in psychological science these days, owing to two highly publicized cases of unambiguous scientific misconduct where internationally known psychological researchers were found to have faked data. These two abused the trust of scholars working with them, undermined the careers of these scholars, and raised questions about the overall integrity of our field. If two were caught, how many more have been getting by with cheating that has so far gone undetected? Let's be real: It's a good bet that there are more than a few other scientists who "help their data along," either by giving trends in the data that are just short of reliability a little boost by changing a number or two or by selectively dropping data that are impeding statistical significance. This practice is hardly different from fabricating data, though it may be harder to detect. And it is bad practice.

The first thing to be said is that when it comes to cheating (contrary to the recent article by Benedict Carey in *The New York Times*), our field is nothing special. Reports from the [US Office of Research Integrity](#) suggest that psychological science is not a major transgressor. A meta-analysis by Daniele Fanelli (2009) confirms this impression and indicates that misconduct in clinical, pharmacological and medical research is more widespread than in other fields. My estimate of 98.03 percent of APS Members being against misconduct also comes from Fanelli's review (Fanelli gives a pooled weighted average of 1.97% scientists who have admitted to have fabricating, falsifying or modifying data or results at least once).



APS President Doug Medin

So the problem isn't specific to psychology researchers, which may provide a bit of comfort. But we still should worry about other practices that may be more common and perhaps equally pernicious. For example, if a psychological scientist uses open-ended tasks like I often do, he or she has considerable leeway in developing coding categories, and then doing statistics once they have settled on a scheme. But if schemes are selected *because* they yield significant differences, aren't researchers helping themselves to more chances to produce statistically reliable results without taking into account the actual number of tests they are conducting? If all readers see is the final scheme and the statistics associated with them, the number of earlier schemes and corresponding tests remain hidden.[\[1\]](#) I recommend a careful reading of a recent article by Simmons, Nelson and Simonsohn (2011) analyzing the unaccounted for degrees of freedom researchers may exploit to produce statistically significant results.

We're also aware of issues of judgment that can readily be shaded. Aside from the well-known "file-drawer problem" where studies failing to produce some effect often go unreported, there is also the issue of what makes it into the file drawer. It is easy to dismiss a study "failure" as attributable to stimulus materials or procedures that were a bit flawed and to keep the "improved" materials and methods when they produce the desired result.

These and related factors should tend to inflate "false positives" (aka Type I errors) which leads inexorably to the pessimistic conclusion that some unknown (but considerably higher than five percent) proportion of our field's published effects are not true effects. Can we really have sound psychological science given this uncertain picture?

The answer is an emphatic "Yes!"

The first key factor is the tool of replication. Many, if not most, researchers will not seek to publish a result unless they themselves have replicated it, often with modest changes in methods and materials so that they can get a sense of the robustness of the effect. It is also common for researchers aiming to follow up on some published result of interest to get as close as they can to an exact replication of the

study before proceeding further. These successful replications often will go unreported. Furthermore, many phenomena — such as decision heuristics, which we know about thanks to Daniel Kahneman, Amos Tversky and others, — are so robust that they are commonly used as classroom demonstrations.

Another tool is statistical. It has become conventional to report effect sizes and various measures of power associated with a study. This convention not only gives an indication of the robustness of some effect, but also includes the information needed for meta-analyses that help filter out idiosyncratic results.

Yet another potent tool for converging on an accurate empirical picture is consistency within a broader picture. This tool is often referred to as “converging measures,” but the more general point is that research is never done in a vacuum. Psychological scientists always ask themselves how some result fits with what we already know, and if it doesn’t, they are extremely careful in pursuing their finding. They correctly infer that they will need a strong pattern of evidence to overturn convention, not because convention is good per se, but because convention may be based on hundreds, and perhaps thousands, of studies.

What this all adds up to is a science that we can believe in, at least with respect to the study populations, stimulus materials, methods and contexts that receive sufficient attention (as I’ve indicated in other columns and as other scholars have suggested, this attention has been disproportionately directed at undergraduates and lab studies, using materials and methods that have been selected for these favored few, but that’s an issue of generalizability, not validity).

Can we do better? It behooves us to carefully consider this issue. As far as I can tell, we don’t know how to make scientists more ethical through any sort of training program. It appears that courses in ethical behavior increase knowledge but leave ethicality unaffected (e.g. Eastwood, et al, 1996; Plemmons et al, 2006). One challenge psychological scientists might want to take on is to see if misconduct can be reduced by teaching ethical practices.

But again, we may benefit from revisiting the practices of science. Some journals, such as *Judgment and Decision Making*, now require that scholars make their data publicly available at the time of publication. This practice allows others to check the statistics and perform other analyses that may provide some additional insights (this practice may also discourage faking data). It is also unethical not to provide data from published studies when they are requested by other scientists.[\[2\]](#) Simmons et al (2011) couple their analyses of hidden degrees of freedom with a series of recommendations that make good sense.

Another useful step might be to make it easier for failures to replicate to become part of the public record. Barbara Spellman, current Editor of the APS journal, *Perspectives on Psychological Science*, is working on a plan to do just this (and it might be good to include successful replications as well). We also might want to take another look at what in marketing is called the “pioneer effect” that favors what is already published over new candidates. For example, if someone publishes some result and later on someone else finds the opposite pattern, typically the onus is on the latter to explain both sets of results before their paper will be published. Maybe we should make a place in our journals for findings that we don’t understand and apparently conflict with what has gone before.

There are doubtless many other ideas about improving the way research and the associated publication

process is conducted in our field. The APS Board is evaluating these issues on an ongoing basis and we welcome your input.

Young scholars of psychological science should be disgusted by the instances of fraud seen in a few scientists (1.97 percent of them). And I would always recommend a healthy dose of skepticism in approaching some theoretical or empirical issue. But that's just good science. I believe that intrinsic to the scientific practices of replication, drawing on converging measures, sensitivity to experimenter expectancy effects^[3] and bringing diverse perspectives to bear are the keys to helping our field to continue to flourish. ^[4]

[1] An advanced graduate student in my lab, Ananda Marin, points out that if the solution to this problem was to require coding schemes to be fixed in advance, the treatment would be worse than the disease. Finding themes and patterns in the data is important and maybe we need to figure out how to make statistical adjustments for this sort of exploration.

[2] Of course, there are some cases where this practice may not be fair or practical. One can imagine a large and rich data set collected only after years of labor that is reported piecewise. Requiring that these data be instantly available to all would undermine the effort that went into data collection.

[3] For example, using research assistants or data coders who are blind to the hypotheses under consideration.

[4] Finally, we may have an advantage over most other sciences because the practices of scientists themselves are a legitimate target of psychological research.