

A Year of Horrors

Eric-Jan Wagenmakers

For social psychologists, the year 2011 can go in the books as a true annus horribilis. First, the flagship journal in the field, the Journal of Personality and Social Psychology, decided to publish an article claiming that people can look into the future. Going from silly to bizarre, this ability was reported to be strongest for extravert women confronted with erotic pictures. The resulting media frenzy centered on questions such as “should JPSP ever have accepted such an article?” and, more to the point, “is there something wrong with the way social psychologists conduct their experiments and analyze their data?”. The author of the infamous article, Dr. Daryl Bem, was a guest on the Colbert Report, where the host mocked the effect as “extrasensory pornception”. And then, as if the reputation of JPSP had not yet been tarnished quite enough, the journal rejected (without external review) all manuscripts that reported failures to replicate the Bem results. As it turns out, JPSP has a long-standing policy not to publish “mere” replication studies. A terrible policy to espouse, of course – apparently, JPSP believes it can pollute the field and then leave the clean-up effort to the lesser journals.

Second, there was the Stapel saga. For those of you who have been living in a cave, Diederik Stapel is one of social psychology’s brightest young stars. At least he was until he had to admit that he had fabricated data on a large scale, affecting at least 30 publications and probably many more. Consequently, Stapel was forced to leave Tilburg University and later voluntarily relinquished his doctoral degree. Immediately after the Stapel bomb exploded, some social psychologists tried to limit the damage to their field by putting the blame squarely on the shoulders of the one man: Diederik Stapel. And indeed, the committee Levelt concluded that Stapel, and Stapel alone, has “culpa”. Nevertheless, the shock waves of the Stapel saga run much deeper. One may well wonder about the scientific status of a field in which an academic serial-killer can wander loose, undetected, for decades. Why weren’t his results subjected to replication? You may call Stapel deranged, but the fact that he was not caught by replication points to a major system malfunction. Hindsight is always 20/20, but this makes it no less remarkable that, to the best of my knowledge, failed replications of Stapel’s work have not made it in print (if such replications were even conducted).

For social psychology, the year 2012 has not gone off to a good start either. The influential John Bargh was recently confronted with a failed replication of his experiment where participants walk more slowly after being primed with the “old” stereotype (Doyen et al., PloS ONE). His response (<http://www.psychologytoday.com/blog/the-natural-unconscious/201203/nothing-in-their-heads>) has been described as a “scathing personal attack” and may make graduate students think twice before they attempt to publish a nonreplication of a leading figure in the field.

Two articles bring us closer to home – back to experimental psychology and cognitive neurosciences. The first article is by Simmons et al. (*Psychological Science*) and its title speaks volumes: “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant”. In this article, the authors state clearly what many researchers already know: using creative outlier-rejection, selective reporting, post-hoc theorizing, and optional stopping, researchers can very likely obtain a significant result even if the null hypothesis is exactly true. Or, in other words, if you set out to torture the data until they confess, you will more likely than not obtain some sort of confession – even if the data are perfectly innocent. The second article, by John et al.

(also *Psychological Science*) uses a questionnaire to measure the extent to which psychologists actually use these tricks. The results are remarkable, and, if true, one would not allow researchers anywhere near their data. For instance, there exists such a practice as “rounding a p-value”, meaning that if you find $p=.054$, you report $p<.05$. Of course, this is really lying about a p-value, and you may be surprised to learn that 22% of psychologist admitted having used this method of reporting.

What I’m driving at is this. It is easy to feel superior to social psychologists with their counterintuitive effects, speculative explanations, and lack of formal theory. But when we do our statistics, we feel the same pressure to publish, we are subject to the same confirmation biases, and I doubt that we can lay claim to a higher morality. I suspect, therefore, that we torture the data just as much as social psychologists do. In the past, I myself have certainly looked at intermediate results and tested participants until the ambiguous effect became a clear effect. When you use p-values, this amounts to plain and simple cheating. Such cheating may help our publication records, but it hurts the field, and it makes one curious how many results are spurious.

The key problem, I believe, with the way we do our research is that almost all of it (my subjective estimate is at least 99%) is to some extent exploratory. This means that researchers have not fully committed themselves to a specific method of data analysis before they see the data. It then

becomes very tempting, perhaps almost irresistible, to fine-tune the analyses to the data. Some researchers succumb to this temptation more easily than others, and from presented work it is often completely unclear to what degree the data were tortured to obtain the reported confession. The only solution to this problem is to separate more strictly the analyses that are confirmatory from those that are exploratory. A good method, already used in medicine, is to pre-register your experiments and indicate exactly what analyses you intend to carry out. Only those analyses deserve the label “confirmatory” in the final article. Of course these are not the only analyses you can carry out, and nobody will stop you from freely running additional analyses as well. These additional analyses, however, would need to be reported under the heading “exploration”. Clearly, confirmatory analyses (the kind that our statistics were designed for, after all) have much greater evidential impact than exploratory analyses, and reviewers and readers deserve full access to this information.

Last year, I conducted a purely confirmatory study, with all analyses pre-registered online. It was scary but I can recommend the experience – you will notice the difference with how you normally carry out your research. I am more and more convinced that the only way to obtain clear answers from Nature is to ask her clear questions. Now let’s see what the reviewers say when I submit the study for publication...