

World view

How misconduct helped psychology to thrive



By Jelte Wicherts

Grass-roots action against bad behaviour has spurred reform – and should keep going.

Ten years ago this week, I was startled to see tweets saying that Dutch psychologist Diederik Stapel, a former colleague, had admitted to falsifying and fabricating data in dozens of articles. My inbox filled with e-mails from fellow methodologists, researchers who examine and refine research techniques and statistical tools. They expressed disbelief about the extent of the misconduct, but also a sense of inevitability. We all knew that sloppiness, low ethical standards and competitiveness were widespread.

What happened next was inspiring: an open debate that went far beyond misconduct and focused on improving research. Numerous researchers, many early in their careers, used social media to call for bias-countering practices, such as sharing data and plans for analysis. It changed the conversation. Before 2011, my applications for grants to study statistical errors and biases in psychology were repeatedly rejected as low priority. By 2012, I had received funding and set up my current research group.

This August, another incident of data fraud came to light, this time in a 2012 publication from behavioural-science superstar Dan Ariely, who agrees that the data are fabricated, but says he did not fabricate them. This case, ironically in a study assessing how to encourage honesty, is an invitation to examine how expectations for research practice have changed, and how much further reform must go.

Publication bias – the tendency for findings that confirm hypotheses to be published more often than are null results – was documented clearly in the 1950s. The 1960s and 1970s brought warnings that decisions about how data were analysed could cause bias, such as the identification of spurious or overly strong effects. The widespread failure to share psychology data for verification purposes was also decried in the 1960s and 1970s. (My group documented it in 2006.)

By the 1990s, methodologists had raised the alarm that most studies had unacceptably low statistical power – the probability that actual effects are being detected – and that researchers often misrepresented a study as being designed to test a specific hypothesis, when in fact they had spotted a trend in exploratory work. The high prevalence of statistical errors was not news, at least to methodologists. Nor was the practice of tweaking and repeating analyses until a statistical threshold (such as $P < 0.05$) was reached. In 2005, a modelling paper showed that, combined, these biases could mean that most published results were false (J. P. A. Ioannidis *PLoS Med.* **2**, e124; 2005). This provocative message generated attention, but little practical change.

Despite this history, before Stapel, researchers were



We need to strengthen a culture in which top research is rigorous and trustworthy, as well as creative and exciting.”

broadly unaware of these problems or dismissed them as inconsequential. Some months before the case became public, a concerned colleague and I proposed to create an archive that would preserve the data collected by researchers in our department, to ensure reproducibility and reuse. A council of prominent colleagues dismissed our proposal on the basis that competing departments had no similar plans. Reasonable suggestions that we made to promote data sharing were dismissed on the unfounded grounds that psychology data sets can never be safely anonymized and would be misused out of jealousy, to attack well-meaning researchers. And I learnt about at least one serious attempt by senior researchers to have me disinvited from holding a workshop for young researchers because it was too critical of suboptimal practices.

Around the time that the Stapel case broke, a trio of researchers coined the term *P*hacking and demonstrated how the practice could produce statistical evidence for absurd premises (J. P. Simmons *et al. Psychol. Sci.* **22**, 1359–1366; 2011). Since then, others have tirelessly promoted study preregistration and organized large collaborative projects to assess the replicability of published findings.

Much of the advocacy and awareness has been driven by early-career researchers. Recent cases show how preregistering studies, replication, publishing negative results, and sharing code, materials and data can both empower the self-corrective mechanisms of science and deter questionable research practices and misconduct.

For these changes to stick and spread, they must become systemic. We need tenure committees to reward practices such as sharing data and publishing rigorous studies that have less-than-exciting outcomes. Grant committees and journals should require preregistration or explanations of why it is not warranted. Grant-programme officers should be charged with checking that data are made available in accordance with mandates, and PhD committees should demand that results are verifiable. And we need to strengthen a culture in which top research is rigorous and trustworthy, as well as creative and exciting.

The Netherlands is showing the way. In 2016, the Dutch Research Council allocated funds for replication research and meta-research aimed at improving methodological rigour. This year, all universities and major funders in the country are discussing how to include open research practices when they assess the track records of candidates for tenure, promotion and funding.

Grass-roots enthusiasm has created a fleet of researchers who want to improve practices. Now the system must assure them that they can build successful careers by following these methods. Never again can research integrity become a taboo topic: that would only create more untrustworthy research and, ultimately, misconduct.

Jelte Wicherts is chair of the Department of Methodology and Statistics at the Tilburg School of Social and Behavioral Sciences in the Netherlands. e-mail: j.m.wicherts@tilburguniversity.edu