# Perspectives on Psychological Science

http://pps.sagepub.com/


**Scientific Misconduct and the Myth of Self-Correction in Science**


Wolfgang Stroebe, Tom Postmes and Russell Spears



*Perspectives on Psychological Science* 2012 7: 670

DOI: 10.1177/1745691612460687

The online version of this article can be found at:
http://pps.sagepub.com/content/7/6/670

Published by:

SAGE

http://www.sagepublications.com

On behalf of:

Association For Psychological Science

Perspectives on Psychological Science
7(6) 670–688

DOI: 10.1177/1745691612460687
http://pps.sagepub.com

# Scientific Misconduct and the Myth of Self-Correction in Science

**Wolfgang Stroebe[1,2], Tom Postmes[2], and Russell Spears[2]**
[1]Utrecht University, The Netherlands, and [2]University of Groningen, The Netherlands

## Abstract

The recent Stapel fraud case came as a shattering blow to the scientific community of psychologists and damaged both their image in the media and their collective self-esteem. The field responded with suggestions of how fraud could be prevented. However, the Stapel fraud is only one among many cases. Before basing recommendations on one case, it would be informative to study other cases to assess how these frauds were discovered. The authors analyze a convenience sample of fraud cases to see whether (social) psychology is more susceptible to fraud than other disciplines. They also evaluate whether the peer review process and replications work well in practice to detect fraud. There is no evidence that psychology is more vulnerable to fraud than the biomedical sciences, and most frauds are detected through information from whistleblowers with inside information. On the basis of this analysis, the authors suggest a number of strategies that might reduce the risk of scientific fraud.

## Keywords

fraud, scientific misconduct, research integrity, replication, peer review

The news that the highly respected Dutch social psychologist Diederik Stapel had been accused of scientific misconduct and had admitted large-scale research fraud came as a terrible shock to the scientific community of social psychologists. Stapel was internationally renowned, and his work had received prestigious awards from the European Association of Social Psychology and the U.S. Society of Experimental Social Psychology. This scandal provided a field day for the international press, and psychology was portrayed as being highly vulnerable to scientific misconduct. The field responded with suggestions on how the risk of fraud could be reduced in the future (e.g., Crocker & Cooper, 2011; Mummendey, 2012; Roediger, 2012). However, the Stapel case, although very high profile, is only one of many fraud cases that were discovered in recent years. Instead of proposing changes on the basis of one case, it would seem useful to study several of these cases together in order to evaluate whether the measures that are now being suggested would have been effective in preventing these frauds.

In this article, we analyze the Stapel fraud and other notorious fraud cases to address four questions: First, how prevalent is fraud? Second, is (social) psychology more vulnerable to fraud than other scientific disciplines? Third, how well do the peer review process and replications work in practice to detect fraud? Fourth, what lessons can we draw from past fraud detection for the effectiveness of strategies to reduce the risk of scientific fraud? We first present the facts of the Stapel case and of a sample of other well-known fraud cases. In analyzing the way that most of these frauds have been detected, we will demonstrate that the idea of the self-correcting nature of science is a myth.

## Notorious Cases of Research Fraud: A Review

The National Science Foundation (2001) defined scientific misconduct as fabrication, falsification, or plagiarism in proposing, performing, or reviewing research or in reporting research results. Such misconduct must be committed intentionally, knowingly, or in disregard of accepted practices. Whereas the fabrication of data involves totally inventing a data set, falsification refers to manipulation of equipment or changing data such that the research is not accurately represented in the research report (National Science Foundation, 2001).

Research fraud can have four types of external effects: First, it can damage the careers of the students and of colleagues, who unknowingly coauthored articles based on fraudulent research. Second, in the case of clinical research, patients can suffer because of misinformation about the efficacy of different treatment options. For example, the fraudulent *Lancet* article published by Wakefield et al. (1999) that linked the

**Corresponding Author:**
Wolfgang Stroebe, Utrecht University, Faculty of Social Sciences, P.O. Box 80.140, 3508 TC Utrecht, the Netherlands
E-mail: w.stroebe@uu.nl

vaccine against measles, mumps, and rubella to autism resulted in a substantial drop in vaccinations, which may have caused multiple deaths among unprotected children (e.g., Braunstein, 2012; Deere, 2012). And the fraudulent study of Bezwoda (Bezwoda, Seymour, & Dansey, 1995) indicating the superiority of high-dose chemotherapy and bone marrow transplantation over other treatments of metastatic breast cancer has led thousands of breast cancer patients to undergo expensive, debilitating, and often fatal bone marrow transplants (Maugh & Mestel, 2001). Third, fraud can delay scientific progress, because researchers waste valuable resources (research funds as well as time) by trying to follow leads suggested by fraudulent research. For example, the fraudulent claims of physicist Jan Hendrik Schön to have built high-performance transistors made of plastic and other materials and even to have built the world's first organic laser led many laboratories to waste years attempting to replicate his findings (Reich, 2009). Finally, news about scientific fraud damages the image of the field in which the fraud was committed and reduces trust in science in general.

## The Stapel case

Stapel's fraud was uncovered by three of his doctoral students. Their suspicion had first been aroused in the summer of 2009, when reading a manuscript written by Stapel describing research supposedly based on school children whose mean age at 19 years was simply too high. They discussed this with each other and then asked Stapel to clarify. He explained it as a mistake. They accepted his explanation and thought nothing further. But their suspicion was reawakened later at a research meeting at which Stapel presented data of a new study. These data fitted the hypotheses so perfectly that one member of the audience jokingly stated: "It is as if he made up these data himself" (Keulemans, 2012a; our translation). They began to check raw data of some of Stapel's studies and found suspiciously high effect sizes. The most conclusive evidence of fraud, however, was that a row of scores in one study was identical with a row of scores in another study. In late August 2011, they contacted the chair of the department who then informed the rector. Stapel could have attributed all of these mistakes to sloppiness and gotten away with it, had he not claimed that these data had been collected by his contacts in schools. When the rector wanted to contact these schools, Stapel admitted that they did not exist. This was the "smoking gun" that convicted him.

Stapel is alleged to have reported data that are fraudulent for his own publications from 1996 onward (Keulemans, 2012b), although he denies this himself. He began sharing fabricated data with PhD students as early as 2002. Around that time, he also began offering suspect data to more senior colleagues while arguing that he did not have time to write them up. He appears to have used four strategies to manipulate or fabricate data. One was the (alleged) data collection in schools, via his sister (who was a schoolteacher) and via other schools who were happy to collaborate, sometimes in return for computing equipment. His second strategy was to arrange for the data that had been collected by students and assistants in the computer lab to be sent to him first, before passing the data on to others, which gave him an opportunity to manipulate the data. His third strategy was to provide PhD students or collaborators with data or results that he had allegedly collected at other universities (either when he worked there or through contacts). And finally, he claimed to have collected data in the field (e.g., for the retracted *Science* paper; Stapel & Lindenberg, 2011). Had he used only the latter three strategies, fraud would have been difficult to prove.

The rector took immediate action. Stapel was suspended from his duties and has since been dismissed. Committees examined all of his publications during the years he spent at the Universities of Tilburg, Groningen, and Amsterdam, where he completed his PhD. The Tilburg committee, which published preliminary conclusions in April 2012 (Stapel Investigation, 2012), has continued their work. By now they have found that 34 publications and three dissertations were based on fabricated data (Witlox, 2012). The committee based their judgment not only on admissions by Stapel but also examined the data for conspicuous oddities. And there appear to have been sufficient "oddities," including "highly unlikely design," "too high effect sizes and correlations given reliability of the variables concerned," and "implausible results, for example too high factor loading, correlations, and too strong effect sizes, and no missing data."[1] The committee also emphasized that none of the coauthors was aware of these falsifications.

The Groningen and the Amsterdam committees have not yet published their complete findings. But according to preliminary information, there are indications that in Groningen 12 publications and five dissertations contained falsified data, and the committee in Amsterdam has expressed strong concerns about seven further publications, some of which were part of Stapel's own dissertation (Keulemans, 2012b; Witlox, 2012).

In evaluating the damage of Stapel's fraud, the consequences for some of his PhD students, whose dissertations were partly based on data (allegedly) collected by Stapel, are probably most severe. Even though they were allowed to keep their degree, if it had already been conferred (in one case, this was to happen 2 weeks after the fraud had been discovered), it must be heartbreaking to have worked for years for a degree, henceforth considered of little value by the academic community. Nearly as severe were the consequences for his coauthors, often young colleagues, whose articles will be retracted because they are partly based on falsified data. This was particularly painful in cases where Stapel's "empirical" contribution to the article had been minimal.

Because Stapel did not do clinical research, his fraud had no consequences for patient populations. It is difficult to assess the consequences his fraud had on the field in terms of research time lost in trying to replicate or to build on Stapel's research. Certainly some PhD student research had to be redirected after a year or two, because findings reported by Stapel could not be replicated. But because he worked mainly on somewhat esoteric problems, this damage is probably limited. He also received more than 2 million Euros in research funding from the Netherlands Organisation for Scientific Research (NWO), which could otherwise have been used for real research. NWO has given notice that it may sue Stapel for misappropriation of funds.

There was certainly great damage to the image of the field of social psychology. Not surprisingly with such a major scandal, the international press had a field day and various blogs joined in. In November 2011, Benedict Carey of the *New York Times* published an article that lambasted psychology research. Under the headline "Fraud Case Seen as a Red Flag for Psychology Research," he referred to expert opinion "that the case exposes deep flaws in the way science is done in a field, psychology, that has only recently earned a fragile respectability." He concluded that "The scandal . . . is the latest in a string of embarrassments in a field that critics and statisticians say badly needs to overhaul how it treats research results" (Carey, 2011). Along similar lines, Hank Campbell (2011) wrote a column in *Science 2.0*, entitled "Diederik Stapel: Another World-Class Psychology Fraud," suggesting that Stapel is one of many fraud cases in psychology. In it, Campbell asked, "Why didn't they catch it sooner?" His answer was that "Unlike physics or biology, social psychology is too scientifically fuzzy to say someone is wrong or demand data; if his results weren't replicated, other researchers assumed they were doing something wrong." And on September 13, 2011, Han Oud, a statistician working at the University of Nijmegen published an opinion piece in the Dutch national newspaper *NRC Handelsblad* with the headline "Fraud Is Too Easy in the Social Sciences" (our translation; Oud, 2011). He concluded that the "social sciences are more susceptible to fraud than are the natural sciences for lack of replicability."

## Other notorious cases of research fraud

To examine whether the Stapel case is unique to social psychology and how fraud cases have been uncovered in other areas, we searched for other cases of research fraud. Our starting point was publications on misconduct (e.g., Broad & Wade, 1982; Diekmann, 2004; Goodstein, 2010) as well as information on the internet and annual reports and case summaries of the U.S. Office of Research Integrity (ORI). We included cases only of research fraud (i.e., fabrication and falsification of data) that resulted in journal articles that have either been retracted or should have been retracted. We did not include cases that involved fraud only in grant proposals. Because one of our main interests was to find out how fraud is initially suspected and detected, we included only those cases on which we could find information on this, typically from newspaper reports. Because ORI does not provide this information, this last criterion reduced the number of usable cases from the ORI reports. Table 1 presents a list of the fraud cases that met these criteria. These cases were sufficiently serious to attract media attention.[2] There are likely to be many other cases that were dealt with quietly by university authorities, who shun publicity about fraud because this reflects badly on the university's image.

There are a number of similarities between the Stapel case and the other fraud cases on our list. First, Stapel was a highly respected researcher, and he usually published his papers with coauthors of considerable respect in the field. This pattern is quite typical for all of these cases. Either the researchers committing the fraud were highly respected, or in the case of young researchers, they published their work with highly respected senior colleagues. Finally, as in most other fraud cases, Stapel's fraud was discovered as a result of inside information by whistleblowers and not through the procedures by which science is supposed to identify fraudulent research.

The Stapel case bears closest resemblance to the fraud cases of Darsee, Schön, Slutsky, Spector, and Sudbø and perhaps also Breuning. Whereas other fraudulent researchers conducted occasional fraud or started later in their careers, these researchers apparently conducted fraud from the beginning. As a result, they experienced a meteoric rise as young superstars (until they came too near to the sun and like Icarus crashed down to earth). All of them were considered brilliant and highly talented by the senior researchers who were their mentors. The Nobel laureate Braunwald, who was the head of the Harvard laboratory at which Darsee committed some of his fraudulent research, thought of Darsee as "brilliant" and "creative," a young man with whom one could discuss research projects as a scientific equal (Culliton, 1983, p. 31). Schön was considered a likely candidate for a Nobel prize at Bell Laboratories (Reich, 2009). Stapel was well known in social psychology and considered by many to be a brilliant researcher, an opinion we shared.

All of these superstars were unusually productive, publishing articles at an incredible rate (certainly aided by the fact that they did not waste time on collecting data). And their data seemed often "to be too good to be true," having large effect sizes and low error rates. Schön once reported data that fitted so unrealistically close to a Gauss curve that a colleague concluded that there was a greater than 90 percent chance that Schön's data had been distorted by some kind of human factor (Reich, 2009). Finally, when asked for their raw data, these "wunderkinder" were often unable to produce them.

So why are researchers, who in all probability would have had fine careers without committing fraud, taking the risk of falsifying or fabricating data and publishing them? Nobody really knows, not even the people who committed the fraud. In his only public statement about his fraud, Stapel wrote, "I will have to dig deeply to find out myself, why all of this happened, and what motivated me to do this" (our translation from Dutch; Stapel, 2011). We suspect that people like Stapel probably start out by slightly altering their data to make them

**Table 1.** A Sample of Notorious Cases of Scientific Fraud

| Year of discovery | Name | Discipline | Affiliation | Mode of discovery | Number of articles fraudulent or questionable | Source |
|---|---|---|---|---|---|---|
| 1974 | William Summerlin | Biomedicine | Sloan-Kettering Institute for Cancer Research, New York, NY | Fraud obvious, (painted skin to fake transplant) | 0 | LaFollette, 2000 |
| 1977 | Robert Gullis | Biomedicine | University of Birmingham, UK, & Max Planck Institute for Biochemistry, Martinsried, Germany | Nonreplication | 11 | Broad & Wade, 1982 |
| 1980 | Vijay Soman | Biomedicine | Yale University, New Haven, CT | Plagiarism & fraud leading to audit | 0 | Broad & Wade, 1982 |
| 1981 | John Darsee | Biomedicine | Emory University, Atlanta, GA; and Harvard University, Cambridge, MA | Caught in act of falsifying | 82 | Broad & Wade, 1982; Culliton, 1983 |
| 1981 | Mark Spector | Biomedicine | Cornell University, Ithaca, NY | Colleague who could not replicate got hold of doctored material | 2 | Broad & Wade, 1982 |
| 1987 | Stephen Breuning | Biomedicine | University of Pittsburgh, PA | Whistleblower | Unknown | Edsall, 1995; J. Scott, 1988 |
| 1987 | Robert Slutsky | Biomedicine | University of California at San Diego | Referee in promotion case | 68 | J. Scott, 1987 |
| 1993 | Tian-Shing Lee | Biomedicine | Harvard Medical School, Boston, MA | Audit (unable to provide data) | 4 | NIH, 1993 |
| 1994 | Malcolm Pearce | Biomedicine | St. George's Hospital, London, UK | Whistleblower | 6 | Lock, 1995. |
| 1997 | Friedhelm Herrmann & Marion Brach | Biomedicine | Max Delbruck Center for Molecular Medicine, Berlin; and University of Ulm, Germany | Whistleblower | 94 | DFG, 2000; Horstkotte, 2004 |
| 1999 | Werner Bezwodaa | Biomedicine | University of Witwatersrand, South Africa | Audit due to doubts about study findings | 1 | Maugh & Mestel, 2001 |
| 1999 | Robert Liburdy | Biomedicine | Lawrence Berkeley National Laboratory, Berkeley, CA | Whistleblower | 2 | Broad, 1999 |
| 2001 | Karen Ruggiero | Social psychology | Harvard University and University of Texas at Austin | Whistleblower | 2 | Murray, 2002; Office of Research Integrity, 2001 |
| 2002 | Jan Hendrik Schön | Physics | Bell Laboratories, Murray Hill, New Jersey | Outside researchers discover duplicate data | 29 | Reich, 2009 |
| 2002 | Marc Hauser | Psychology/biology | Harvard University | Whistleblower | 1 | http://ori.hhs.gov/content/case-summary-hauser-marc |
| 2004 | Andrew Wakefield | Biomedicine | Royal Free Hospital and School of Medicine, London, UK | Journalist | 1 | Braunstein, 2012; Deere, 2012 |
| 2005 | Ranjit Chandra | Biomedicine | Memorial University, Newfoundland and Labrador, Canada | Whistleblower & journal reviewer | 1 | Kumar, 2008; O'Neill-Yates, 2006 |
| 2005 | Eric Poehlman | Biomedicine | University of Vermont, Burlington | Whistleblower | 10 | Payne, 2005; http://ori.hhs.gov/poehlman_notice |

*(continued)*

**Table 1.** (continued)

| Year of discovery | Name | Discipline | Affiliation | Mode of discovery | Number of articles fraudulent or questionable | Source |
|---|---|---|---|---|---|---|
| 2006 | Elisabeth Goodwin | Biomedicine | University of Wisconsin–Madison | Whistleblower | 3 | Couzin, 2006; Retrieved from http://ori.hhs.gov/content/case-summary-goodwin-elizabeth |
| 2006 | Steven A. Leadon | Biomedicine | University of North Carolina at Chapel Hill | Whistleblower | 4 | Wallace, 2006; http://ori.hhs.gov/content/case-summary-leadon-steven |
| 2007 | John Sudbø | Biomedicine | Oslo University, Norway | Outside researcher discovers inconsistencies in publication | 16 | http://ori.hhs.gov/content/case-summary-sudbo-jon; http://en.wikipedia.org/wiki/Jon_Sudb%C3%B8 |
| 2006 | Woo Suk Hwang | Biomedicine | Seoul National University, South Korea | Whistleblower | 2 | Cyranosky, 2009a, 2009b; http://en.wikipedia.org/wiki/Hwang_Woo-suk |
| 2007 | Kristin Roovers | Biomedicine | University of Pennsylvania, Philadelphia | Editors | 3 | Katsnelson, 2007; http://ori.hhs.gov/content/case-summary-roovers-kristin |
| 2008 | Roxana Gonzales | Social psychology | Carnegie Mellon University, Pittsburgh, PA | Whistleblower | 2 | Scudellari, 2008; http://ori.hhs.gov/content/case-summary-gonzalez-roxana |
| 2008 | Mai Nguyen | Biomedicine | University of California at Los Angeles | Whistleblower | 2 | Ghose (2009); http://ori.hhs.govhttp://www.nsf.gov/oig/session.pdf |
| 2008 | Peili Gu | Biomedicine | Baylor College of Medicine, Waco, TX | Whistleblower | 3 | Dolgin, 2008; http://ori.hhs.gov/content/case-summary-gu-peili |
| 2008 | Judith Thomas & Juan Luis Contreras | Biomedicine | University of Alabama, Birmingham | Audit | 15 | Grant, 2009; http://ori.hhs.gov/content/case-summary-thomas-judith |
| 2009 | Scott Reuben | Biomedicine | Tufts University, Boston, MA | Routine audit | 21 | Adams, 2010; Borrell, 2009 |
| 2009 | Luk van Parijs | Biomedicine | MIT, Cambridge, MA | Whistleblower | 4 | Reich, 2011; http://ori.hhs.gov/content/case-summary-van-parijs-luk |
| 2010 | Sezen Bengü | Chemistry | Columbia University, New York, NY | Failure to replicate; sting operation | 6 | Schulz, 2011; http://ori.hhs.gov/content/case-summary-sezen-bengu |
| 2010 | John Boldt | Biomedicine | University of Giessen, Germany | Journal reader found data too perfect | 88 | Blake, 2011; Marcus, 2011a |
| 2010 | Scott Brodie | Biomedicine | University of Washington, Seattle | Whistleblower | 3 | Shephard, 2007; http://ori.hhs.gov/content/case-summary-brodie-scott-j |

*(continued)*

**Table 1.** (continued)

| Year of discovery | Name | Discipline | Affiliation | Mode of discovery | Number of articles fraudulent or questionable | Source |
|---|---|---|---|---|---|---|
| 2010 | Hua Zhong & Tao Liu | Biomedicine | Jinggangshan University, China | Outside researcher discovers inconsistencies | 70 | Marcus, 2011; Stafford, 2010 |
| 2011 | Victor Ninov | Nuclear chemistry | Lawrence Berkeley National Laboratory | Failure to replicate | 1 | Goodstein, 2011 |
| 2011 | Milena Penkova | Biomedicine | Copenhagen University, Denmark | Dissertation committee & whistleblower | 3 (so far) | Leake & Campbell, 2011; Young, 2011 |
| 2011 | Diederik Stapel | Social psychology | Tilburg University, the Netherlands | Whistleblower | 53 | Stapel Investigation, 2012; Witlox, 2012 |
| 2011 | Don Poldermans | Biomedicine | Erasmus University, Rotterdam, the Netherlands | Whistleblower | 0 | Smit, 2012; Visser, 2011 |
| 2012 | Dipak Das | Biomedicine | University of Connecticut, Storrs | Whistleblower | 17 so far | Bartlett, 2011; Oransky, 2012b |
| 2012 | Michael W. Miller | Biomedicine | State University of New York, Upstate Medical University | Whistleblower | 2 | Mulder, 2012; http://ori.hhs.gov/content/case-summary-miller-michael-w |
| 2012 | Yoshitaka Fujii | Biomedicine | Toho University, Japan | Outside researchers discover inconsistencies | 172 | Marcus, 2012; Normile, 2012 |

statistically significant or to make them fit their hypotheses even better. After all, just having good data that support an interesting hypothesis is often not sufficient. The data have to fit the hypotheses pretty perfectly for an article to be accepted by a top journal. Once they score early successes and become known as highly promising, researchers have to keep on publishing at a high rate in top journals to meet these high expectations. And if studies do not work out at all, they have to make greater and greater changes to their data, until they decide to abandon data collection and to invent the total data set. As Stapel (2011) wrote, “In the past years the pressure became too much for me. I have not been able to withstand the pressure to score, to publish, to be better and better. I wanted too much too fast.” Although this sounds reasonable, it does not really explain his behavior. Many top scientists feel this pressure, and yet, as far as is presently known, few try to score by fabricating their data. Furthermore, it is likely that Stapel began committing fraud as early as his dissertation. This points to another factor raised by Goodstein (2010): These fraudulent researchers often thought they already knew the answers to the research problem they were considering and therefore felt no need to do the research to find the answer.

## Some Conclusions About Research Fraud

### *How prevalent is research fraud?*

We have no exact information about the prevalence of fraud in science or even about the number of fraud cases that are uncovered each year. Some indication regarding prevalence comes from surveys in which researchers were anonymously asked whether they ever committed research fraud. According to a meta-analysis of 18 such surveys, nearly 2% of respondents admitted to having fabricated or falsified data at some point (Fanelli, 2009). However, as none of these surveys were representative and because they do not always provide information on how often these researchers had committed fraud, they do not tell us a great deal about the proportion of published research that might be based on fraudulent data.

Probably the most precise estimate of the frequency of research misconduct comes from a survey conducted by the Gallup Organization (2008) and funded by the ORI. Instead of asking individuals to report on their own misconduct, they asked investigators to report instances of misconduct they had observed in their own departments during the last 3 years. By asking only one investigator per department, they tried to reduce the risk that the same incidence of misconduct would be reported several times. The report estimated that 1.5% of all research conducted each year would be fraudulent. And on the basis of an estimate that 155,000 researchers are supported by National Institutes of Health (NIH) grants, they suggested that there would be a total of 2,335 incidents of possible misconduct per year, with 60% of these incidents involving falsification or fabrication of data. This estimate exceeds by far the cases of misconduct that are reported to ORI.

### *Is psychology more susceptible to fraud than other sciences?*

The fact that more than two thirds of the cases on our list come from the biomedical sciences might suggest that it is not psychology but the biomedical sciences that are most susceptible to scientific fraud, a conclusion that Goodstein (2010) also drew in his recent book on scientific fraud. Of the cases listed by us, only Ruggiero, Gonzales, Hauser and Stapel held positions in psychology. (Breuning was trained as a psychologist but worked in a medical department, whereas Hauser was trained as a biologist, but worked at a psychology department.) We did not include the recently discovered cases of Smeesters (Erasmus University of Rotterdam, 2012) and Sanna (Yong, 2012), who are both social psychologists, even though both cases might finally be categorized as fraud. At present, there is not sufficient evidence to include them in our list.

The percentage of frauds committed by psychologists compared with biomedical researchers in our sample does not allow one to draw conclusions about the proportion of medical or psychological researchers committing fraud. For this conclusion, we would need to know the size of the population of scientists in each of these disciplines (i.e., base rates).[3] Related to this is the fact that many of our cases were investigated by ORI. As there are likely to be many more federally funded research projects in medicine than psychology, there will also be more medical research cases investigated by ORI.

It is also possible that there are differences between disciplines in the ease with which fraud can be detected and proved, at least with regard to medical research involving patients: Institutional approval procedures are required, and patient records need to be kept. In audits of research, all of these documents are checked. For a fraudulent medical researcher, it is practically impossible to have all of this paperwork in order. The paper trail may be less extensive in cases of animal research, but there should still be some documentation. Many medical fraud cases were detected because the fraudulent researcher did not go through the institutional approval procedures or because patients could not be traced. However, audits are usually conducted only once there is sufficient reason to suspect fraud. Furthermore, audits are less likely to be effective in the case of (nonclinical) psychological research. Participants in psychology experiments are mostly anonymous and untraceable, once the experiment has been concluded. Thus, the only trace remaining of such an experiment is the data, and these can be falsified by a fraudulent researcher.

However, a meta-analysis by Fanelli (2009) provides some convergent evidence that fraud is indeed more frequent in the biomedical sciences than in other disciplines. Fanelli (2009) concluded that “surveys among clinical, medical and pharmacological researchers appeared to yield higher rates of misconduct than surveys in other fields or in mixed samples” (p. 8). Because the meta-analysis is based on admissions by researchers and because the sample size is known, it does not suffer the shortcomings of conclusions that are based on numbers of confirmed fraud cases.

One explanation for a higher incidence of fraud cases in medicine than in psychology could be that medical research is much more likely than psychological studies to result in financial rewards. It is therefore surprising that in the medical fraud cases in the present research, there is little direct evidence of monetary gains, suggesting that ambition might have been the main motive. There are a few exceptions. For example, the *Sunday Times* journalist Brian Deere (2012) reported evidence that Wakefield had received payments in excess of $600,000 from lawyers who were preparing lawsuits against drug companies that manufactured these vaccines. He received this money before he published a *Lancet* paper that linked the measles, mumps, and rubella vaccine to a new form of autism (Braunstein, 2012). And the falsification of the data by Chandra for the study published in 2001 in *Nutrition* (Chandra, 2001), in which he claimed that a multivitamin formula that he had patented could reverse memory loss in geriatric populations, was likely to be influenced by financial motives (Kumar, 2008). Finally, Scott Reuben, a professor of anesthesiology and pain medicine at Tufts University, obtained thousands of dollars from pharmaceutical companies for research he never performed (see http://en.wikipedia.org/wiki/Scott_Reuben).

## The Myth of the Self-Correcting Nature of Science

It has always been assumed that science is self-correcting in the sense that findings that are based on falsification will eventually be discovered and rejected (Goodstein, 2011). As Philip Handler, then president of the National Academy of Sciences, declared in a hearing of the U.S. House Committee on Science and Technology that was held from March 31 to April 1, 1981, scientific fraud happens rarely, and when it does, "it occurs in a system that operates in an effective, democratic and self-correcting mode" that makes detection inevitable (cited in Broad & Wade, 1982, pp. 11–12). The principal mechanisms of this self-correction are generally assumed to be the peer review system and replications (Broad & Wade, 1982). And it is probably these mechanisms the executive director of the Association for Psychological Science (APS) had in mind, when he stated the following in an article entitled "Despite Occasional Scandals, Science Can Police Itself" (Kraut, 2011): "Scientific inquiry is guided by a set of laboratory conventions and publishing rules that promote integrity and minimize the publication of false conclusions." Trust in the peer review system was recently reaffirmed in an editorial in the journal *Circulation*. The editor expressed his belief that "in the many layers of review a manuscript receives in parallel and beyond peer review, including discussion at our weekly editorial board meeting, . . . although not eliminating the risk of publishing data that are irreproducible in articles that are later retracted, clearly offers the care necessary to minimize this risk" (Loscalzo, 2012, p. 1213) And the conviction that replications can identify fraud was expressed by Crocker and Cooper (2011) in an editorial to a special section in *Science* on data replication and reproducibility: "Scientists generally trust that fabrication will be uncovered when other scientists cannot replicate (and therefore validate) findings." Along similar lines, the former APS president Henry Roediger (2012) commented on the Stapel case that "if others had tried to replicate his work soon after its publication, his misdeeds might have been uncovered much more quickly." It is therefore disconcerting that hardly any of the fraud cases on our list were uncovered by the two "principal mechanisms of self-correction," which lends support to the doubts already expressed by Broad and Wade (1982).

### The peer review process as fraud detector

The aim of the peer review process is to support journal editors in identifying good research that constitutes scientific progress. Fraud detection is not a primary aim of peer review. However, as fraudulent research impairs rather than aids scientific progress, it is reasonable to expect that reviewers should also recognize signs of fraud in evaluating manuscripts. Peer review has already come under critique, because low interrater agreement between reviewers of the same manuscript raises doubt about its efficacy in identifying good research (e.g., Daniel, 1993; Marsh & Ball, 1989; Petty, Fleming, & Fabrigar, 1999; W. A. Scott, 1974). It will therefore not come as a major surprise that peer reviewers are also not very successful in uncovering scientific fraud. Even the most reputable journals appear to accept articles that contain glaring inconsistencies overlooked by reviewers. For example, in a publication by Darsee and Heymsfield (1981) in the *New England Journal of Medicine*, a family tree is reproduced that shows one 17-year-old father with children ages 8, 7, 5, and 4 (Diekmann, 2004). Although this is biologically not impossible, it is highly unlikely and should have resulted in questions to the authors. And the investigation committee checking Darsee's publication while he was at Emory University noted that in another article, he had claimed to have obtained at least a dozen human hearts for experimentation just hours after death, which is quite clearly impossible (Broad, 1983). Schön et al. (2000, 2001) published two articles in *Science* with nearly identical graphs, supposedly reflecting the measured performance of different devices (Diekmann, 2004; Reich, 2009). The fraud of Robert Slutsky was discovered by a referee who reviewed his work for a promotion hearing and realized that the same statistics were replicated in two articles (J. Scott, 1987). Finally, the committee that assessed the articles Stapel published during his time at Tilburg University found "oddities" in many of the articles, which had been overlooked by reviewers and journal editors who accepted these articles for publication (Stapel Investigation, 2012).

Expert readers of journal articles also often discover problems overlooked by reviewers and editors. For example, the suspicion about John Sudbø was raised by the head of the epidemiology division of the Norwegian Institute of Public Health, who had realized that the cancer patient base, which Sudbø et al. (2005) claimed to have used in a study published in the *Lancet*, had not yet been available in 2005 (see http://en.wikipedia.org/wiki/Jon_Sudb%C3%B8). The fraud of Yoshitaka Fujii could have been discovered in 2000, when Kranke, Apfel, and Roewer (2001) published a letter in the journal *Anesthesia & Analgesia* with the title "Reported Data on Granisetron and Postoperative Nausea and Vomiting by Fujii et al. Are Incredibly Nice!" However, the journal did not retract any of the 48 papers cited in the letter and continued to publish another 11 articles authored by Fujii in the following years. It took 10 more years until the new editor of *Anesthesia & Analgesia* decided to write to Toho University (in Japan), where Fujii was working at the time. As a result, Fujii was dismissed in February 2012, having published 172 articles declared to be fraudulent during his career (Marcus, 2012). This would make it the most prolific fraud case so far.

In a similar case, in 1997 a neuroimmunologist at Queen Mary University of London e-mailed concerns to the editor of the *Journal of Experimental Medicine* about a manipulated figure in one of the articles of Luk van Parijs that the journal had published (Reich, 2011). Again, the journal took no action, and Parijs's fraud was discovered only in 2004 when some of his students brought fraud allegations against their mentor. In contrast, the Boldt fraud unraveled because an expert reader of *Anesthesia & Analgesia* wrote to the editor that the pattern of data in one of the published studies was implausibly perfect. As a result, the editor contacted the Rhineland State Medical Association (in Germany), which started an investigation (Blake, 2011). Finally, Schön's fraud came to light because two colleagues from outside Bell Laboratories discovered that he had published duplicated and inconsistently captioned data throughout his work (Reich, 2009). In all of these cases, the scientific review process had failed to detect these problems.

There are a few exceptions to this apparent blindness of reviewers and editors. For example, the falsification of images in panels of Western blot data by Roevers was discovered in the spring of 2005 by editors at the *Journal of Clinical Investigation*. However, the manuscript in which they noted these irregularities had already been accepted for publication and was undergoing final revisions (Katsnelson, 2007). When Chandra submitted a manuscript that proved the efficacy of his patented vitamin for improving memory loss among the aged to the *British Medical Journal*, a reviewer rejected it, because it showed all the hallmarks of having been completely invented (Margetts, 2006).[4] At that time, Chandra was an eminent nutrition researcher, editor in chief of the international journal *Nutrition Research* and founding head of the World Health Organization Centre for Nutritional Immunology (Margetts, 2006). The *British Medical Journal* informed his university (Memorial University, Canada) and requested that the case be investigated. When Chandra was asked by the university for the data from the study, he could not produce them. He claimed that the university had lost them and resigned (see http://en.wikipedia.org/wiki/Ranjit_Chandra).[5]

How was it possible that in all the other cases, reviewers (and, for that matter, also the coauthors) did not recognize the fraudulent nature of the data, even though investigation committees later recognized clear indications of fraud in several of the published articles? One reason is that (known) fraud is extremely rare. Reviewers typically do not consider their task to be one of fraud detection when reviewing a manuscript. A reviewer may wonder about a surprisingly strong effect produced by an experimental manipulation or that the manipulation produced any effects at all, but these are not really acceptable reasons for rejecting a manuscript. If they were, surprising results would never be published and progress in science would be impaired. Furthermore, in the absence of an adequate control sample of comparable studies, judgments about the magnitude of effects in individual studies are difficult to make.

The fact that, in contrast to the investigating committees studying these cases, reviewers and editors do not consider fraud as a potential explanation for reported data patterns partly explains why they might not discover fraud. Although pursuing specific hypotheses can lead to the well-known confirmation bias in information processing and information search (Klayman & Ha, 1987), the failure to consider such hypotheses can have the opposite effect (i.e., that clear signs of fraud are overlooked). Thus, thinking the unthinkable (or the worst) and entertaining the possibility of fraud may turn out to be quite functional when it comes to simply detecting it, especially when this has a very low base rate.

The confirmation bias may be prompted or exacerbated by social motives. In general, there is a generic bias in favor of the in-group (Tajfel & Turner, 1979) and for in-group members to be considered more trustworthy than out-group members (Kramer & Brewer, 1984; Turner, 1985). Group norms may also play an important role: They may make fraud not just unexpected but unthinkable. In situations where injunctive group norms proscribe particular forms of behavior (fraud is antisocial), where descriptive norms suggest this behavior is exceptional (fraud is rare), and where open discussion of suspect behavior is discouraged (fraud is a taboo topic), thoughts of fraud are highly problematic to entertain, let alone discuss with peers and confidantes.

Further reasons why reviewers hardly ever suspect fraud are that in most of the known fraud cases, the researchers committing the fraud took great care to predict effects that were highly plausible on the basis of past research. For example, many of the effects reported by Stapel are so plausible that he might have supported his hypotheses had he really tried. Indeed, more than one social psychologist informally commenting on the Stapel case has suggested that some of the findings suspected of being fraudulently produced or embellished might well be true when tested properly. Even Jan Hendrik Schön, the physics "wunderkind," produced mostly results that were within the range of what could be expected (Reich, 2009). Finally, fraudulent researchers appear to always find highly respected scientist to coauthor their articles.

## *Replications as fraud detector*

In psychological research, there are always a multitude of potential causes for the failure to replicate a particular research finding. Therefore, even repeated failures to replicate do not indicate that fraud had occurred.[6] However, consistent failure to replicate multiple findings of a particular researcher can be seen as a warning signal. It is therefore justified that replications are considered a powerful weapon in science's armor for the protection against fraud (see the remainder of the special section in this issue). However, even though in rare cases the consistent failure to replicate findings aroused suspicion (e.g., in the case of Schön [Reich, 2009] and in the case of Sezen [Schulz, 2011]), replications rarely played a decisive role in uncovering the cases of fraud on our list. There are a number of reasons for this: First, there are few incentives for researchers to replicate studies. Regardless of whether such replications are successful or fail, they are difficult to publish. Thus, even in the cases where replications have been conducted, they often remained unpublished and did not become known to the wider scientific community.

Second, the failure of replications is not very informative. There are always numerous reasons why a researcher might not replicate a study. They might fail for lack of skills or because not all theoretical variables that are necessary to produce an effect were specified in the published manuscript. For example, when researchers could not replicate the experiments of Mark Spector testing his kinase-cascade theory of cancer, they invited Spector to their laboratories, to get advice on how to do these experiments. As Broad and Wade (1982) described, "one by one became aware of a pattern familiar to Spector's colleagues at Cornell, that often the experiments would work only in Spector's hands and could not be repeated without him. But like Spector's colleagues, they found a simple explanation: Mark was just so good at making experiments go" (p. 66). His fraud was discovered only when a colleague proved that he had fraudulently altered some crucial ingredient used in his research. Similarly, repeated failure by other physicists to replicate the findings of Jan Hendrik Schön were attributed for a long time to differences in equipment or to his superior technical skills (Reich, 2009). As Goodstein (2010) remarked, Schön worked in a field of physics in which results are notoriously sample specific and depend very much on the skills of the person who prepares the sample.

In social psychology, failures to replicate are frequently due to the fact that descriptions of experimental procedures are incomplete or that important theoretical variables are not clearly spelled out by a theory. A classic example of the latter case is the study of Festinger and Carlsmith (1959) on the role of incentives in attitude change following counterattitudinal advocacy. At the time, the Festinger and Carlsmith (1959) study was quite controversial, and numerous failures to replicate their findings were reported (e.g., Rosenberg, 1965). It was only when it was demonstrated that it was essential for the production of the dissonance effect that participants felt free to refuse participation in the experiment (Linder, Cooper, & Jones, 1967) and that they further expected their behavior to have negative consequences (Cooper & Worchel, 1970) that the Festinger and Carlsmith (1959) experiments could be replicated. If one had declared the original experiment to be fraudulent and stopped trying to replicate it, these important theoretical insights would never have been made. Finally, social psychological research is culture dependent, and study findings in one culture will not always be replicable in another culture (Smith, 2012).

Although single failures to replicate findings might not be informative, one could expect that meta-analyses that statistically summarize multiple replications would play a major part in uncovering fraud. In the fraud cases we reviewed, a meta-analysis raised alarm bells on one occasion but was nevertheless ineffective in uncovering the fraud. Kranke, Apfel, Eberhart, Georgieff, and Roewer (2001) published a meta-analysis that clearly showed that the findings published by Fujii and colleagues were significantly different from data of all other centers. Although they did not explicitly accuse Fujii and colleagues of fraud, the message could be read clearly between the lines; Kranke, Apfel, Eberhart, et al. (2001) concluded: "Further, if data of a dominating centre do not appear to be valid for other centres, it may seem advisable to either exclude them from the analysis or perform sub-group analyses so that results without the data from the dominating centre are available" (p. 659). And yet, as we reported earlier, Fujii was allowed to publish fraudulent research for another decade.

There are two fraud cases that were discovered as a result of failure to replicate, but the circumstances of these fraud cases are quite unique and therefore instructive. One occurred at the Max Planck Institute for Chemistry in Germany, where Gullis worked as a postdoc. After he left, his colleagues at Max Planck, who were coauthors on four of his articles, had the impression that there was something wrong with this research and tried to replicate it. When they could not do so, they asked Gullis to return and replicate his experiments himself. When he did not succeed either, he finally admitted having faked the data (Broad & Wade, 1982). Thus, it was his own failure to replicate his findings that forced him to admit his fraud.

The other case occurred at the nuclear chemistry group at the Lawrence Berkeley National Laboratory. Viktor Ninov, a nuclear physicist, was involved in research that resulted in the (alleged) discovery of the heavy element 118. The rule for such discoveries is that for a new element to be accepted, the results must be reproduced by another group. Three groups undertook to do so, working with equipment that was even more powerful than that used at Berkeley. When none of these groups could replicate the findings, an investigation was started at Berkeley, which finally concluded that the results had been fabricated and that the only person in a position to fake them was Viktor Ninov (Goodstein, 2010). Viktor Ninov proclaimed his innocence but was dismissed from his post in 2002. This case is highly atypical in two aspects, namely, that the finding had to be replicated to be accepted and, second, that failure to replicate was a clear sign that the original finding was falsified.

## Conclusions

So how can we explain the failure of the self-correcting processes of science? One major reason is that science is based on trust. As already mentioned, scientists do not expect their colleagues to falsify their data and therefore do not look for signs of fraud when reading manuscripts or articles. This is probably the reason why hardly any fraud cases are discovered by the peer review process. It is only in extreme cases of data patterns that are simply too good to be true that fraud is considered as a possibility. And then the signs of fraud are typically not found in a single manuscript but by looking over multiple publications of an author. Another reason is that the accusation of fraud is so serious that it cannot be made lightly. Thus, even if one suspects that the work of a colleague is fraudulent, one needs unambiguous proof for this suspicion: the well-known "smoking gun."

But even if there is no doubt that fraud has occurred, it is often difficult in research published by multiple authors to identify the person or persons responsible for the fraud. A striking example is the case of the internationally renowned German cancer researchers Friedhelm Herrmann and Marion Brach, who had worked at the Max-Delbrück Centre for Molecular Medicine in Berlin and later the Universities of Ulm (Herrmann) and Lübeck (Brach). As the result of accusations of data falsification by a whistleblower in 1997, an investigation committee was instituted by the German research funding organization Deutsche Forschungsgemeinschaft (DFG). The committee concluded that data falsification had clearly been involved in 94 publications between 1988 and 1996 in which Herrmann was coauthor (DFG, 2000;see http://de.wikipedia.org/wiki/Friedhelm_Herrmann). Furthermore, the two had plagiarized a research proposal that Herrmann had reviewed for the Dutch Wilhelmina Fonds and submitted it to the Thyssen Stiftung. Whereas Brach (who had also lived with Herrmann for most of the critical period) admitted fraud and emphasized that Herrmann was involved as well, Herrmann denied involvement in the fraud and blamed it on Brach.[7] In the meantime, Brach had emigrated to the United States and could not be contacted (Horstkotte, 2004). It appears that it will remain impossible to prove who was responsible for the fraud.

A similar case that is still ongoing is the alleged fraud of two former postdocs of Professor Silvia Bulfone-Paus of the University of Lübeck (Germany) and until recently chair of the Immunology Department at the Research Centre Borstel. Triggered by an accusation of misconduct by a whistleblower, the centre instituted an external investigation (Schiermeier, 2010). On the basis of spot checks of the group's research output, the investigation committee found manipulation of images (i.e., using picture of protein blots from unrelated experiments) in six papers produced between 2001 and 2009, on which Bulfone-Paus is listed as either senior or corresponding author (Schiermeier, 2010). The fraud was attributed to her two Bulgarian postdocs, who had left the country and could not be reached for comment. However, the committee also concluded that as supervisor of the pair, Bulfone-Paus bore "substantial responsibility" (Schiermeier, 2010). In the meantime, the number of retracted articles has risen to 13, and on one of those, the accused postdocs do not feature as coauthors (Jump, 2011; Oransky, 2011a,b). In 2010, Bulfone-Paus was persuaded by the Research Centre to step down as department head.[8]

## Institutional Changes to Curb Fraud

Before we discuss measures to reduce the risk of fraud, we should acknowledge that a great deal has already been done in the last decades to curb the incidence of fraud. Fraud was not really an issue of public interest until notable cases of research misconduct emerged in the late 1970s and early 1980s (Gallup, 2008). Because cases of fraudulent research reflect poorly on universities, these institutions were in the past quite reluctant to take action when fraud allegations were made. For example, when Robert L. Sprague, a professor of psychology at the University of Illinois, wrote a letter to the University of Pittsburgh accusing Breuning of fraud, the university took no action, even though Breuning's research had important treatment implications. Sprague had worked with Breuning for over 3 years before Breuning moved to the University of Pittsburgh. Sprague had realized that it would have been physically impossible for Breuning to have conducted the experiments he claimed to have done (Edsall, 1995). Sprague then wrote to the National Institute of Mental Health (NIMH), which did not take any action for 9 months. Only in 1987, 3 and a half years after Sprague's initial letter did NIMH release a report stating that Breuning had committed research fraud.

Since then the situation has changed dramatically, at least in the United States, where institutions have been developed at university as well as federal level to deal with research fraud. Most universities have offices or committees of research integrity, where individuals who suspect scientific misconduct can report their suspicions. Allegations of research fraud are first investigated by the institution at which the fraud is alleged to have happened, and this investigation is monitored and reviewed by ORI. To the best of our knowledge, no such institutions have evolved in Europe. In the Netherlands, most universities have committees or "persons of trust" to whom suspicions of research fraud can be reported. At Tilburg University, this person was the rector, which provided a very high threshold for such reports. In Germany, the DFG now has ombudspersons who can be contacted in cases where research fraud is suspected. The Stapel case has been investigated by an ad hoc committee without involvement of the major research funding agencies (NWO and the Royal Netherlands Academy of Arts and Sciences).[9] But being a whistleblower is still not without risk. The three PhD students who uncovered the

Stapel case were therefore wise to insist that their names not be revealed.

Another positive development is that major European and American journal publishing houses funded a Committee on Publication Ethics (COPE) in 1997, which advises editors and publishers on all aspects in this area. According to the code of conduct for journal editors, "editors have a duty to act if they suspect misconduct or if the allegation of misconduct is brought to them. Editors should not simply reject papers that raise concerns about possible misconduct. They are ethically obliged to pursue alleged cases. Editors should first seek a response from those suspected of misconduct. If they are not satisfied with the responses, they should ask the relevant employers, or institutions, or some appropriate body . . . to investigate" (COPE, 2012).

## How Can We Further Reduce the Risk of Fraud?

All of these changes take effect only once fraud has been discovered or at least strongly suspected. However, as Stapel (2011) remarked in his letter of self-justification, scientific fraud is too easy, because there are too few control mechanisms in science. People are tempted to commit fraud when the expected rewards are great and punishment is unlikely because the risk of discovery is small. This suggests three possible strategies to reduce the risk of research fraud, namely, to reduce the rewards associated with fraud, to increase the costs of fraud to the perpetrators, and, most important, to increase the chance of discovery.

### Rewards

Throughout their career, researchers are rewarded for publishing in high-impact journals. Researchers who are highly productive are likely to be considered rising stars who not only get tenure easily but also qualify for early or midcareer awards by one of the respected professional societies. (Stapel received several early and midcareer awards. Ironically, his acceptance speech for one such award was on the right way to do research.) And these intangible rewards heaped on young superstars are typically converted into tangible rewards, such as salary increases and better research facilities. To publish in high-impact journals, one has to develop novel theoretical hypotheses and support them with well-conducted empirical research. But this alone is not enough: The data have to provide strong, unambiguous support for the hypothesis. Obviously, researchers who fabricate or falsify their data have an advantage here. It is therefore not surprising that the impact of journals and the number of retractions are significantly positively correlated (Fang & Casadevall, 2011). Unfortunately, there is not a great deal that can be done to change this system. One major reason is that this system rewards not only fraudsters but also brilliant and deserving researchers. Furthermore, this system is shaped by market forces. There are more researchers than there are good jobs and more articles produced than there is space in high-impact journals. Thus, there has to be a selection process, and a selection process based on criteria of quality is preferable to all other criteria one can think of (e.g., the old boys' network).

But it is also clear that in determining what is good (and bad) quality research, institutions can do more to encourage ethical research practices. Research on student fraud suggests that social norms toward fraud and perceived moral obligations are strong predictors of fraudulent behavior (e.g., Crown & Spiller, 1998; Whitley, 1998). And failure to maintain a principled ethical climate, research suggests, is one of the main predictors of misconduct in corporate settings more generally (Kish-Gephart, Harrison, & Trevino, 2010). On this dimension, institutions can almost certainly encourage good practice more than they do at present. At the outset, clear norms and codes of conduct should exist,[10] and procedures for investigating fraud should be known, sound, and impartial.

### Costs

The costs of research fraud to the perpetrators, if it is discovered, vary across countries. But generally, researchers lose their jobs and become outcasts in the scientific community. Otherwise, consequences are very variable. Poehlman, Breuning, Hwang, and Reuben received prison sentences; Sudbø, Wakefield, and others had their medical licenses revoked, either for some time (Sudbø) or permanently (Wakefield). Sudbø had his medical doctorate revoked by the University of Oslo, because it was based on falsified data. Schön had his PhD taken away by the University of Konstanz (even though it was not based on fraudulent data). Stapel relinquished his PhD voluntarily, when the University of Amsterdam threatened to start a procedure to revoke it. Conversely, Herrmann was not only allowed to keep all his degrees but even continues to use his title of professor. And Woo Suk Hwang of South Korea is still conducting research and publishing it in scientific journals (see http://en.wikipedia.org/wiki/Hwang_Woo-suk).

Following the Stapel affair, there have been suggestions in the Netherlands that like any other fraud, research fraud should become a criminal offense. After all, researchers committing research fraud not only do a great deal of damage (some of it quite material); they usually also profit financially (e.g., tenure, salary increases). Furthermore, the threat of a prison sentence would be a very strong disincentive. However, it is questionable whether this would help rather than complicate matters. For one thing, the burden of proof in court is often greater than the kind of proof accepted by the scientific community. For example, when DFG sued Herrmann in court for repayment of the millions in research funding he had received from them, they lost the court case. Herrmann blamed it all on his colleague (and former partner) Brach, and because she was in the United States and could not be heard in court, his claim could not be disproved. To bring criminal charges against them in a German court, one would have had to prove that research funds were applied for with the intention of conducting fraudulent research or to misappropriate the funds in other ways (Die Zeit, 2005). In the United States, it is sufficient to prove that falsified or fabricated data have been used in the application of a research grant from government sources.

Finally, increasing the severity of the penalties for fraud would help only if the researchers committing fraud expected to be caught. However, as Stapel stated in his letter of excuse, fraud in science is easy. And as the discrepancy between the number of annual fraud cases estimated by the Gallup Organization (2008) and the number that are actually identified by ORI suggests, the probability of being found out is minimal. It is therefore quite reasonable that fraudsters do not seriously consider potential penalties in their decision to commit research fraud. Therefore, the most effective strategies to reduce the risk of research fraud would be those that increased the chance of discovery.

## Increasing the chance of discovery

From the current research, it appears that in past cases, whistleblowers were more central to fraud detection than any other method: They are usually close to the research in question and have the inside knowledge to provide the "smoking gun." Although other methods should definitely be strengthened, whistleblowers are likely to remain the single most effective instrument against scientific cheating: They are ideally placed and better informed than outsiders. Moreover, just as the fraud cases in this review are likely to be the tip of an iceberg of fraudsters, the whistleblowers that helped identify them are only a few drops in a sea of potential self-regulation.

Given that most scientists value the honest pursuit of knowledge above all, how can we aid whistleblowers without creating a culture of distrust and without exacerbating the risk of false accusations? One major responsibility for research institutions lies in enhancing the awareness of the prevalence and methods of fraud, so that potential whistleblowers (a) know that this kind of behavior is not abnormal in the sense that it is prevalent, (b) know what to look for, and (c) are reassured that both they and the suspected person will be treated well when they decide to discuss their suspicions in confidence with a colleague. Central to increasing this awareness is that institutions become more transparent about cases where fraud has been proven. Too often, and for too long, details of proven cases of fraud have been kept from public knowledge. Awareness of fraud is not helped, either, by the tendency of those who encounter it to attribute it either to pathologies of the perpetrator or to some specific scientific out-group.[11] Both forms of attribution reduce the perceived self-relevance of scientific fraud and are thus likely to decrease vigilance and self-regulation within the scientific community and within research groups.

Several responses to the Stapel case have suggested alternative mechanisms for fraud detection that need strengthening, principally greater transparency with data, including depositing data in repositories where they can be accessed by other scientists, and facilitation with replications (e.g., Crocker & Cooper, 2011). Regarding transparency and repositories, it is certainly indicative that fraudsters are usually reluctant to make available the data they allegedly collected. Inventing a whole data set is time-consuming and risky. Even though this requirement would not prevent researchers from fabricating or falsifying data, the fact that they usually make mistakes would aid detection.

Public availability of data sets would also facilitate the application of statistical methods to discover scientific fraud. Such methods have been used by Simonsohn (2012) in identifying problems in several articles of Smeesters and Sanna. However, since the precise method applied in this case has only just been disclosed in an unpublished manuscript, more time is needed to evaluate its efficacy. An important criterion in such an evaluation, in addition to the extent to which such methods identify fraudulent data, is the extent to which they misidentify nonfraudulent data as fraudulent. Such a method will be useful only if the proportion of false positives is minimal. However, the development of statistical methods to identify fraudulent data sets will certainly reduce fraud in science by increasing the risk of detection (for other statistical methods of fraud detection, see Diekmann, 2004, 2007).

In view of our earlier discussion of the shortcomings of replications as fraud detectors, the suggestion to encourage replications might seem less likely to prevent fraud. However, apart from the fact that in psychology there are always numerous reasons why a replication might have failed, the fact that failed replications are not typically published and thus do not become widely known to the scientific community is certainly also likely to be responsible for their minor role in uncovering frauds. Multiple reports of failures to replicate a specific study might certainly have served as a warning signal. The recent creation of a Web site where researchers can upload and view results of replication attempts in experimental psychology may help address this problem (PsychFileDrawer.org, 2012).

We would argue that an important lesson can be learned from the fact that the committees that were instituted to investigate fraud allegations typically found many indications of fraud in the publications of the research suspected of fraud. One reason these cues were overlooked by coauthors, peer reviewers, and editors is that those committees look at a large sample of an author's publications, whereas reviewers look only at one article (even though they are likely to have read other publications by that author). Clearly suspicious patterns, or more general contingencies, are easier to detect when all the information is available rather than dispersed over time and processed in a piecemeal or case-by-case basis (Arkes & Harkness, 1983). However, a second reason is that those committees were alerted to fraud, whereas reviewers rarely consider fraud as a potential alternative explanation for research findings. Therefore, another change that needs to be made is that the possibility of fraud be added to the criteria reviewers are given by journals about which aspects of an article they should evaluate. Given that fraudsters appear often to make quite obvious mistakes when fabricating data, such mistakes might be detected earlier if reviewers (and coauthors) were more open to the possibility that data might have been falsified.

## Conclusion

The Stapel fraud has shocked the field of psychology in general and social psychology in particular. Many journalists and other commentators in the mass media have painted the field as a discipline that is particularly susceptible to fraud. And although journalists, and especially commentators in the blogosphere, do not always adhere to the same principles of scientific rigor and sampling in making their claims as scientists do, when scientists are caught red-handed, they are in a very weak position to defend themselves against innuendo as well as the well-founded accusations so damaging to their collective reputation. Although the Stapel case hardly exposed deep flaws in the way psychologists conduct their science, it clearly demonstrated that any trust-based system, as science is, is open to exploitation. However, any system can be improved, and the lessons from the Stapel case, along with other cases of fraud more generally, have to be learned and acted on.

In using the Stapel case to stimulate a more general analysis of research fraud (i.e., when and how it occurs, its prevalence, and its prototypical character), we have tried to move beyond the ad hoc and ad hominem agenda that single but spectacular case studies afford, to try to assess some general patterns and processes. This may put us in a better position to propose more effective solutions and prophylactic procedures. Our provisional conclusion based on this analysis is that, rather than changing the incentive system, the most efficient and effective approach is to improve fraud detection. In this, we can perhaps strengthen the self-correcting nature of science through peer review. But given that the vast majority of scientists value integrity and honesty, it is likely to be more effective to ensure that the whistleblowers are heard and heeded (while guaranteed anonymity). To get to this position, we need to cultivate the same kind of critical mindset to findings and to data that look perfect as to findings and data that do not. We also need to consider the patterns of perfection from a more panoptic perspective of research careers over time and also the painful possibility that fraudsters may be among "us." We need to entertain the hypothesis as readers, reviewers, and editors that when data look "too good to be true," sometimes they just might be. If one believes the estimate of annual fraud cases of the Gallup Organization (2008), more than 1,000 such cases remain undetected each year in NIH-supported science alone. Therefore, as unpalatable as it is, to complete the culture change initiated in the second half of last century, we have to accept the fact that fraud can happen in our midst and that we have to look out for it. Only then will we recognize the signs when we see them.

### Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

### Notes

1. It would certainly be interesting to apply these criteria also to research reports, where falsification has not been suspected, to see how many false positives would emerge. But as Stapel admitted that he had falsified the data of those articles, this point is moot.
2. We did not include the cases of the social psychologists Dirk Smeesters and Lawrence Sanna in our list, because we found the evidence for fraud not yet conclusive. These cases were identified by Simonsohn (2012) through statistical analysis of means and standard deviations in published articles. He examined whether estimates across independent samples were too similar to have originated in random samples. In the case of Smeesters, the means of conditions predicted to be similar were too similar given the size of the standard deviations. Smeesters admitted to "massaging" data in at least two studies (since then retracted; Oransky, 2012a) but denied fraud. Smeesters claimed that he used the blue dot task, which is a validated and well-established technique to identify people for whom the manipulation could not succeed because they have not read the instructions (see Oppenheimer, Meyvis, & Davidenko, 2009). However, he did not report this in his article, which is clearly a violation of scientific integrity rules. Furthermore, it is doubtful whether this would explain the similarity in means. The Committee for Inquiry into Scientific Integrity of the Erasmus University of Rotterdam (2012), instituted after concerns raised by Simonsohn, concluded, "These patterns are probably the result of data selection by Smeesters. The committee has no confidence in the scientific integrity of the results in these three articles" (p. 8). This statement falls short of the explicit conclusion of fraud. The other case is that of Lawrence Sanna. Here Simonsohn's (2012) suspicion was triggered by the fact that the means in three experiments differed substantially across conditions while the standard deviations were nearly identical. Sanna's work was investigated by a committee at the University of North Carolina (his previous employer, where the alleged fraud had been committed). Sanna resigned his position at the University of Michigan and withdrew three of his published articles. However, since the report of the North Carolina investigation was not released, it is at present uncertain whether this is a clear case of scientific fraud.
3. A similar baseline fallacy would probably be the conclusion that research fraud is a man's job. The fact that the majority of frauds have been conducted by men could merely reflect the preponderance of men in science. Some support for this assumption comes from the fact that women are increasingly represented in more recent cases of fraud.
4. Chandra then submitted the manuscript to the journal *Nutrition*, which published it in 2001. He also published in the journal *Nutrition Research*, of which he was editor in chief, an article by Amrit Jain (2002). This article claimed that a certain vitamin supplement improved immunity and lowered infection risks in the elderly and recommended that all individuals older than 50 years should be given regular supplements. The author of the article could never be traced

and is suspected to have been Chandra himself (Kumar, 2008).

5. It is doubtful, however, that the university would have taken action, had they not investigated Chandra already in 1994 in response to information by Chandra's research nurse that a study he reported to have conducted on new-born babies had never been conducted. At that time, the university decided to take no action against Chandra. All of this became very public when the Canadian Broadcasting Corporation broadcast a three-part investigation into "The secret life of Ranjit Chandra" conducted by their reporter O'Neill-Yates (2006).

6. It is important to note that even successful replication of a study does not rule out fraud. One reason that fraud is difficult to detect is that fraudsters often invent data that support very plausible hypotheses. It is therefore possible that their guesses are correct some of the time. This is even more likely if the fraud consists of data manipulation (i.e., falsification) rather than complete fabrication.

7. When Herrmann was asked by the *Journal of Investigative Dermatology* about the possibility of fraudulent data in the manuscript of De Voss et al. (1994), he sent the following letter: "My contribution to this manuscript was (1) to be the chief of the Department where some of the work displayed in that paper took place, and (2) to read the manuscript and to correct the language when necessary. Thus, I was—as you know—not at all involved in the experimental work, which was done by, or under the advice of Prof. Dr. Marion Brach. She is the only person who can tell you about the results and experiments coming to display the fore mentioned figures. Unfortunately, I don't know how to reach Prof. Brach" (Herrmann, 2003).

8. The decision of the centre to accuse Bulfone-Paus of involvement in the fraud has been strongly criticized by an international group of 25 colleagues, who wrote: "The massive and unfair punishment that our esteemed colleague in Germany is currently being subjected to, damages science much more than it protects it from future misconduct by others. This raises serious concerns, and must not be tolerated" (cited in Oransky, 2011a). Furthermore, the Medical School of the University of Manchester, where Bulfone-Paus holds a part-time position as professor of immunology, announced in July 2011 that it did not believe the concerns about the papers merited a formal investigation (Jump, 2011). However, without further investigation it might never become clear whether the fraud allegations against Professor Bulfone-Paus are unwarranted.

9. Although the Netherlands lack institutionalized procedures for dealing with fraud, both Tilburg University and the Erasmus University of Rotterdam have acted swiftly, decisively, and with full openness in the cases of Stapel and Smeesters. The same cannot be said about Harvard University in the case of Hauser or the University of North Carolina in the case of Sanna. Although the University of North Carolina instituted an investigation committee, the committee worked in secrecy, and the findings were not disclosed. There may be legal reasons for this, but such secrecy does not prevent damage to the reputation of the researchers who are investigated. It also makes it impossible for colleagues to evaluate the validity of the charges brought against the accused.

10. Among students, for example, honor codes have been found to reduce cheating, provided that these codes are more than mere window dressing (McCabe, Trevino, & Butterfield, 2001).

11. Ironically, the effects of the attribution of fraudulence to specific subdisciplines were documented in a study conducted by Stapel, Koomen, and Spears (1999). This study, which has so far been cleared by the investigating committee and thus appears to be legitimate, showed that when confronted with a prominent case of plagiarism in psychology, social psychologists thought this scandal to be less personally relevant and less hurtful for the profession when the perpetrator was presented as a "clinical psychologist" than when the perpetrator was presented as a "psychologist."

## References


Adams, M. (2010). Big pharma researcher admits faking dozen of studies for Pfizer, Merck (opinion). *Natural News*. Retrieved from http://www.naturalnews.com/028194_Scott_Reuben_research_fraud.html

Arkes, H. R., & Harkness, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, *112*, 117–135.

Bartlett, T. (2011). Marc Hauser resigns from Harvard. *Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Marc-Hauser-Resigns-From/128296/

Bezwoda, W. R., Seymour, L., & Dansey, R. D. (1995). High-dose chemotherapy with hematopoietic rescue as primary treatment for metastatic breast cancer: A randomized trial. *Journal of Oncology*, *13*, 2483–2489.

Blake, H. (2011, March 3). Joachim Boldt profile: A glittering career built on charisma and charm. *The Telegraph*. Retrieved from http://www.telegraph.co.uk/health/healthnews/8360678/Joachim-Boldt-profile-a-glittering-career-built-on-charisma-and-charm.html

Borrell, B. (2009). A medical Madoff: Anesthesiologist faked data in 21 studies. *Scientific American*. Retrieved from http://www.scientificamerican.com/article.cfm?id=a-medical-madoff-anesthestesiologist-faked-data

Braunstein, G. D. (2012, May 3). Andrew Wakefield's vaccination and autism study: Allegations of fraud with financial motivations. *Huffington Post, Los Angeles*. Retrieved from http://www.huffingtonpost.com/glenn-d-braunstein-md/andrew-wakefields-vaccina_b_816208.html

Broad, W. J. (1983). Notorious Darsee case shakes assumptions about science. *The New York Times*. Retrieved from http://www.nytimes.com/1983/06/14/science/notorious-darsee-case-shakes-assumptions-about-science.html?pagewanted=all

Broad, W. J. (1999). Scientist faked data linking cancer to electromagnetic fields, probe finds. *The New York Times* Retrieved from http://www.nytimes.com/libray/national/science/0722499sci-fake-data.html

Broad, W. J., & Wade, N. (1982). *Betrayers of the truth*. New York, NY: Simon & Schuster.

Campbell, H. (2011). Diederik Stapel: Another world class psychology fraud. *Science 2.0*. Retrieved from http://www.science20.com/science_20/blog/diederik_stapel_another_world_class_psychology_fraud-84171

Carey, B. (2011). Fraud case seen as a red flag for psychology research. *The New York Times*. Retrieved from http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html

Chandra, R. (2001). Effect of vitamin and trace-element supplementation on cognitive function in elderly subjects. *Nutrition*, *17*, 709–712 (retracted in 2005).

Committee on Publication Ethics. (2012). *Code of conduct and best practice guidelines for journal editors*. Retrieved from http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf

Cooper, J., & Worchel, S. (1970). The role of undesired consequences in arousing cognitive dissonance. *Journal of Personality and Social Psychology*, *16*, 199–206.

Couzin, J. (2006). Truth and consequences. *Science*, *313*, 1222–1226.

Crocker, J., & Cooper L. (2011). Editorial: Addressing scientific fraud. *Science*, *334*, 1182. doi:10.1126/science.1216775. Retrieved from http://www.sciencemag.org.proxy.library.uu.nl/content/334/6060/1182.full

Crown, D. F., & Spiller, M. S. (1998). Learning from the literature on collegiate cheating: A review of empirical research. *Journal of Business Ethics*, *17*, 683–700.

Culliton, B. J. (1983). Coping with fraud: The Darsee case. *Science*, *220*, 31–35

Cyranosky, D. (2009a). Hwang verdict imminent. *Nature*, *461*, 1035. doi:10.1038/4611035a

Cyranosky, D. (2009b). Woo Suk Hwang convicted, but not of fraud. *Nature*, *461*, 1181. doi:20.1038/4611181a

Daniel, H. D. (1993). *Guardians of science*. Weinheim, Germany: VCH Verlagsgesellschaft.

Darsee, J. R., & Heymsfield, S. B. (1981). Decreased myocardial taurine levels and hypertaurinuria in a kindred with mitral-valve prolapse and congestive cardiomyopathy. *New England Journal of Medicine*, *304*, 129–135.

Deere, B. (2012). *The Lancet scandal*. Retrieved from http://briandeer.com/mmr-lancet.html

De Voss, S., Brach, M., Budnik, A., Grewe, M., Herrmann, F., & Krutmann, J. (1994). Post-transcriptional regulation of Interleukin-6 gene expression in human keratinocytes by ultraviolet B radiation. *Journal of Investigative Dermatology*, *103*, 92–96.

DFG. (2000, June 19). Pressemitteilung Nr. 26 | 19. Juni 2000. *Task Force legt Abschlußbericht vor: Unstimmigkeiten auch im Umfeld von Friedhelm Herrmann* [Task force submits final report: Irregularities also in the environment of Friedhelm Herrmann]. Retrieved from http://www.dfg.de/service/presse/pressemitteilungen/2000/pressemitteilung_nr_26/index.html

Diekmann, A. (2004). Betrug und Täuschung in der Wissenschaft: Datenfälschung, Diagnoseverfahren, Konsequenzen [Fraud and deception in science: Falsification, diagnosis, consequences]. *Jahrbuch der Deutschen Akademie der Naturforscher Leopoldina (Halle/Saale), LEOPOLDINA (R. 3)*, *49*, 1–21.

Diekmann, A. (2007). Not the first digit! Using Benford's Law to detect fraudulent scientific data. *Journal of Applied Statistics*, *34*, 321–329.

Die Zeit. (2005). Die Reputation zu verlieren ist die "härteste Strafe" [Losing your reputation is the greatest punishment]. Retrieved from http://www.zeit.de/online/2005/52/forscher

Dolgin, E. (2008). Postdoc censured for fudged images. *The Scientist*. Retrieved from http://classic.the-scientist.com/blog/display/55256/

Edsall, J. T. (1995). On the hazards of whistleblowers and on some problems of young biomedical scientists in our time. *Science and Engineering Ethics*, *1*, 329–340.

Erasmus University of Rotterdam. (2012, June 1). *Report by the committee for inquiry into scientific integrity*. Retrieved from http://www.eur.nl/fileadmin/ASSETS/press/2012/Juli/report_Committee_for_inquiry_prof._Smeesters.publicversion.28_6_2012.pdf

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, *4*, 1–11.

Fang, F. C., & Casadevall, A. (2011). Retracted science and the retraction index. *Infection and Immunity*, *79*, 3855–3859.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *59*, 203–210.

Gallup Organization. (2008). *Observing and reporting suspected misconduct in biomedical research*. Retrieved from http://ori.hhs.gov/sites/default/files/gallup_finalreport.pdf

Ghose, T. (2009). UCLA prof falsified cancer data. *The Scientist*. Retrieved from http://classic.the-scientist.com.blog/display/55408/

Goodstein, D. (2010). *On fact and fraud*. Princeton, NJ: Princeton University Press.

Grant, B. (2009). Renal researchers faked data. *The Scientist*. Retrieved from http://classic.the-scientist.com/blog/display/55823/

Herrmann, F. (2003). Letter to the editor: Potential fraud in Fig 2:JID 103:92-96, 1994. *Journal of Investigative Dermatology*, 121, 424. doi:10.1038/sj.jid.56011896

Horstkotte, H. (2004, February 25). Forschungsbetrug: Datentrickster behält Professorentitel [Data manipulator keeps professor title]. *Spiegel Online*. Retrieved from http://www.spiegel.de/wissenschaft/mensch/0,1518,287690,00.html

Jain, A. L. (2002). Influence of vitamins and trace-elements on the incidence of respiratory infection in the elderly. *Nutrition Research*, *22*, 85–87. doi:10.1016/S0271-531(01)00369-4

Jump, P. (2011). New retraction of paper by husband and wife research team. *The Times Higher Education*. Retrieved from http://www.timeshighereducation.co.uk/story.asp?storycode=418431

Katsnelson, A. (2007). Former UPenn postdoc faked images. *The Scientist*. Retrieved from http://classic.the-scientist.com/news/display/53469/

Keulemans, M. (2012a, January 21). De ontmaskering frauderende Stapel word maandenlang voorbereid [The exposure of the fraudulent Stapel was prepared for months]. *De Volkskrant*. Retrieved from http://www.volkskrant.nl/vk/nl/2672/Wetenschap-Gezondheid/article/detail/3131931/2012/01/21/Ontmaskering-frauderende-Stapel-werd-maandenlang-voorbereid.dhtml

Keulemans, M. (2012b, June 1). Diederik Stapel pleegde al fraude tijdens promotie in Amsterdam [Diederik Stapel already committed fraud during his dissertation in Amsterdam]. *De Volkskrant*. Retrieved from http://www.volkskrant.nl/vk/nl/2844/Archief/

archief/article/detail/3264084/2012/06/01/Diederik-Stapel-pleegde-al-fraude-tijdens-promotie-in-Amsterdam.dhtml

Kish-Gephart, J. J., Harrison, D. A., & Treviño, L. K. (2010). Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work. *Journal of Applied Psychology*, *95*, 1–31.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation and information in hypothesis-testing. *Psychological Review*, *94*, 211–228.

Kramer, R. M., & Brewer, M. B. (1984). Effects of group identity on resource use in a simulated commons dilemma. *Journal of Personality and Social Psychology*, *46*, 1044–1057.

Kranke, P., Apfel, C., Eberhart, L. H., Georgieff, M., & Roewer, N. (2001). The influence of a dominating centre on a quantitative systematic review of Granisetron for preventing postoperative nausea and vomiting. *Acta Anaesthesiologica Scandinavica*, *45*, 659–670.

Kranke, P., Apfel, C., & Roewer, N. (2001). Reported data on Granisetron and postoperative nausea and vomiting by Fujii et al. are incredibly nice! *Anesthesia & Analgesia*, *90*, 1004.

Kraut, A. (2011, December 3). Despite occasional scandals, science can police itself. *Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Despite-Occasional-Scandals/129997/

Kumar, M. N. (2008). A review of the types of scientific misconduct in biomedical research. *Journal of Academic Ethics*, *6*, 211–218.

LaFollette, M. C. (2000). The evolution of the "scientific misconduct issue": A historical overview. *Experimental Biology and Medicine*, *224*, 211–215.

Leake, J., & Campbell, M. (2011). Downfall of Denmark's professor sexy brains. *The Sunday Times*. Retrieved from http://vietstudies.info/NghinCuuDom_DanMach_LondonTimes.pdf

Lee, T. S. (1993, June 25). *Final findings of scientific misconduct. NIH Guide*, 22, (23). Retrieved from http://grants.nih.gov/grants/guide/notice-files/not93-177.html

Linder, D. E., Cooper, J., & Jones, E. E. (1967). Decision freedom as a determinant of the role of incentive magnitude in attitude change. *Journal of Personality and Social Psychology*, *6*, 245–254.

Loscalzo, J. (2012). Irreproducible experimental results: Causes, (mis)interpretations and consequences. *Circulation*, *125*, 1211–1214.

Lock, S. (1995). Lessons from the Pearce affair: Handling scientific fraud. *British Medical Journal*, *310*, 1547–1548.

Marcus, A. (2011a). Crystal myth: 11 more retractions from crystallography journal after 2010 fakery. *Retraction Watch*. Retrieved from http://retractionwatch.wordpress.com/2011/02/28/crystal-myth-11-more-retractions-from-crystallography-journal-after-2010-fakery/

Marcus, A. (2011b). German medical board issues sweeping findings in Boldt case: Ninety studies implicated in probe might need retraction. *Anesthesiology News*. Retrieved from http://www.anesthesiologynews.com/ViewArticle.aspx?d=Web+Exclusives&d_id=175&i=February+2011&i_id=702&a_id=16596

Marcus, A. (2012). Does anesthesiology have a problem? Final version of report suggests Fujii will take retraction record. *Retraction Watch*. Retrieved from http://retractionwatch.wordpress.com/2012/07/02/does-anesthesiology-have-a-problem-final-version-of-report-suggests-fujii-will-take-retraction-record-with-172/#more-8482

Margetts, B. (2006). Editorial: Stopping the rot in nutrition science. *Public Health Nutrition*, *9*, 169–173. doi:10.1079/PHN206947

Marsh, H. W., & Ball, S. (1989). The peer review process used to evaluate manuscripts submitted to academic journals. *Journal of Experimental Education*, *57*, 151–169

Maugh, T. H., & Mestel, R. (2001). Key breast cancer study was a fraud. *Los Angeles Times*. Retrieved from http://articles.latimes.com/2001/apr/27/news/mn-56336

McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics & Behavior*, *11*, 219–232.

Mulder, J. T. (2012). Former high-ranking scientist at upstate medical faked research, federal investigation shows. *The Post*. Retrieved from http://www.syracuse.com/news/index.ssf/2012/02/former_high-ranking_scientist.html

Mummendey, A. (2012). Scientific misconduct in social psychology: Towards a currency reform in science. *European Bulletin of Social Psychology*, *24*, 4–7.

Murray, B. (2002). Research fraud needn't happen at all. *APA Monitor*. Retrieved from http://www.apa.org/monitor/feb02/fraud.aspx

National Institute of Health [NIH]. (1993). Final findings of scientific misconduct. *NIH Guide, 22*(23). Retrieved from http://grants.nih.gov/grants/notice-files/not93-177.html

National Science Foundation. (2001). *New research misconduct policies*. Retrieved from http://www.nsf.gov/oig/session.pdf

Normile, D. (2012). A new record for retractions. *ScienceInsider*. Retrieved from http://news.sciencemag.org/scienceinsider/2012/04/a-new-record-for-retractions.html

Office of Research Integrity. (2001). *Office of Research Integrity Annual Report 2001: Findings of scientific misconduct*. Notice. 66 Fed. Reg. 64266.

O'Neill-Yates, C. (2006, January 30). The secret life of Ranjit Chandra. *The National*. Transcript published by INFACT Canada. Retrieved from http://www.infactcanada.ca/Chandra_Jan30_2006.htm

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872.

Oransky, I. (2011a). The Bulfone-Paus saga continues: Her supporters and home institution exchange sharp letters. *Retraction Watch*. Retrieved from http://retractionwatch.wordpress.com/2011/05/19/bulfone-paus-saga-continues-her-supporters-and-home-institution-exchange-sharp-letters/#comments

Oransky, I. (2011b). Bulfone-Paus retraction count grows to 13 with one in *Transplantation. Retraction Watch*. Retrieved from http://retractionwatch.wordpress.com/2011/12/13/bulfone-paus-retraction-count-grows-to-13-with-one-in-transplantation/#more-5459

Oransky, I. (2012a). Following investigation, Erasmus social psychology professor retracts two studies, resigns. *Retraction Watch*. Retrieved from http://retractionwatch.wordpress.com/2012/06/25/following-investigation-erasmus-social-psychology-professor-retracts-two-studies-resigns/#comment-15310

Oransky, I. (2012b). Retraction count for Dipak Das rises to 17. Retrieved from *Retraction Watch* http://retractionwatch.wordpress.com/category/by-author/dipak-das/

Oud, H. (2011). Fraude is te gemakkelijk in de sociale wetenschappen. *NRC Handelsblad*, p. 14. Retrieved from http://archief.nrc.nl/index.php/2011/September/13/Overig/14/Fraude+is+te+gemakkelijk+in+de+sociale+wetenschappen/check=Y

Payne, D. (2005). Researcher admits faking data. *The Scientist*. Retrieved from http://classic.the-scientist.com/news/20050321/02/

Petty, R. E., Fleming, M. A., & Fabrigar, L. R. (1999). The review process at PSPB: Correlates of interreviewer agreement and manuscript acceptance. *Personality and Social Psychology Bulletin*, *25*, 188–203.

PsychFileDrawer.org. (2012). *Archives for the attempts at replication in experimental psychology*. Available from http://psychfiledrawer.org/

Reich, E. S. (2009). *Plastic fantastic*. New York, NY: Palgrave Macmillan.

Reich, E. S. (2011). Fraud case we might have seen coming. *Nature*, *474*, 552. doi:10.1038/474552a. Retrieved from http://www.nature.com/news/2011/110728/full/news.2011.437.html

Roediger, H. L. (2012). Psychology's woes and a partial cure: The value of replication. *Observer*. Retrieved from http://www.psychologicalscience.org/index.php/publications/observer/2012/february-12/psychologys-woes-and-a-partial-cure-the-value-of-replication.html

Rosenberg, M. J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, *1*, 28–42.

Schiermeier, Q. (2010). German research centre widens misconduct probe. *Nature News*. doi:10.1038/news.2010.671. Retrieved from http://www.nature.com/news/2010/101213/full/news.2010.671.html

Schön, J. H., Berg, S., Kloc, C., & Batlogg, B. (2000). Ambipolar pentacene field-effect transistors and inverters. *Science*, *287*, 1022–1023.

Schön, J. H., Meng, H., & Bao, Z. (2001). Field-effect modulation of the conductance of single molecules. *Science*, *294*, 2138–2140.

Schulz, W. G. (2011). A puzzle named Bengü Sezen. *Science & Technology*, *89*, 40–43.

Scott, J. (1987). At UC San Diego: Unraveling a research fraud case. *Los Angeles Times*. Retrieved from http://articles.latimes.com/1987-04-30/news/mn-2837_1_uc-san-diego

Scott, J. (1988). Researcher admits faking data to get $160.000 in research funds. *Los Angeles Times*. Retrieved from http://articles.latimes.com/1988-09-20/news/mn-2318_1_research-fraud

Scott, W. A. (1974). Interreferee agreement on some characteristics of manuscripts submitted to the Journal of Personality and Social Psychology. *American Psychologist*, *29*, 698–702.

Scudellari, M. (2008). Grad student falsified data. *The Scientist*. Retrieved from http://classic.the-scientist.com/blog/print/54912/

Shephard, A. (2007). Case closed: UW researcher found guilty of falsifying AIDS data. *The Daily of the University of Washington*. Retrieved from http://dailyuw.com/news/2007/dec/05/case-closed-uw-researcher-found-guilty-of/

Simonsohn, U. (2012). *Just post it: The lesson from two cases of fabricated data detected by statistics alone* (Unpublished report). The Wharton School, University of Pennsylvania, PA.

Smit, M. (2012). De val van Dan Poldermans [The fall of Dan Poldermans]. *Medisch Contact*. Retrieved from http://medischcontact.artsennet.nl/Nieuws-26/archief-6/Tijdschriftartikel/113363/De-val-van-Don-Poldermans.htm

Smith, P. (2012). Cultural social psychology. In M. Hewstone, W. Stroebe, & K. Jonas (Eds.), *Introduction to social psychology* (5th ed., pp. 499–536). Oxford, England: Wiley-Blackwell.

Stafford, N. (2010). China: Act on scientific fraud. *Chemistry World*. Retrieved from http://www.rsc.org/chemistryworld/News/2010/January/15011003.asp

Stapel, D. (2011). Stapel betuigt openlijk 'diepe spijt' [Stapel declares publicly 'deep regret']. *Brabants Dagblad*. Retrieved from http://bd.nl/nieuws/tilburg-stad/stapel-betuigt-openlijk-diepe-spijt-*1*.121338

Stapel, D. A., Koomen, W., & Spears, R. (1999). Framed and misfortuned: Identity salience and the whiff of scandal. *European Journal of Social Psychology*, *29*, 397–402.

Stapel, D. A., & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science*, *332*, 251–252.

Stapel Investigation. (2012). Retrieved from https://www.commissielevelt.nl/levelt-committee/fraud-determined/

Sudbø, J. (2007). *ORI case summary*. Retrieved from http://ori.hhs.gov/content/case-summary-sudbo-jon

Sudbø, J., Lee, J. J., Lippman, S. M., Mork, J., Sagen, S., Flatner, N., . . . Danneberg, A. J. (2005). Non-steroidal anti-inflammatory drugs and the risk of oral cancer: A nested case-control study. *Lancet*, *365*, 1359–1366.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In S. Worchel & W. G. Austin (Eds.), *The psychology of intergroup relations* (pp. 33–47). Monterey, CA: Brooks-Cole.

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behaviour. In E. J. Lawler (Ed.), *Advances in group processes: Theory and research* (pp. 77–122). Greenwich, CT: JAI Press.

Witlox, M. (2012). Opnieuw publicaties Stapel afgekeurd. (Further Stapel publications declared fraudulent). Retrieved from http://universonline.nl/2012/09/13/opnieuw-publicaties-stapel-afgekeurd/

Visser, M. (2011). Erasmus universiteit ontslaat hoogleraar na 'fraude' [Erasmus University dismisses professor because of fraud]. *NRC Handelsblad*. Retrieved from http://www.elsevier.nl/web/Nieuws/Wetenschap/322628/Erasmus-Universiteit-ontslaat-hoogleraar-na-fraude.htm

Wade, N. (2011a). A journal's statement may aid a Harvard researcher accused of misconduct. *The New York Times*. Retrieved from http://www.nytimes.com/2011/04/26/science/26hauser.html

Wade, N. (2011b). Scientist under inquiry resigns from Harvard. *The New York Times*. Retrieved from http://www.nytimes.com/2011/07/21/science/21hauser.html

Wakefield, A. J. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., . . . Walker-Smith, J. A. (1999). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet*, *351*(9103), 637–641.

Wallace, C. (2006). Biologist charged with more fraud. *The Scientist*. Retrieved from http://classic.the-scientist.com/news/display/23624/

Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education*, *39*, 235–274.

Yong, E. (2012). *Uncertainty shrouds psychologist's resignation*. Retrieved from http://www.nature.com/news/uncertainty-shrouds-psychologist-s-resignation-1.10968

Young, M. (2011). Penkowa for dummies. *University Post*. Retrieved from http://viet-studies.info/NghinCuuDom_DanMach_London-Times.pdf