Risk factors for fraud and academic misconduct in the social sciences

René Bekkers

Center for Philanthropic Studies, Faculty of Social Sciences, VU University Amsterdam

E-mail: r.bekkers@vu.nl

November 29, 2012

This note aims to feed the discussion about how to deal with fraud and other forms of academic misconduct in the wake of the Stapel and Smeesters affair and the publication of the report by the Schuyt Commission of the Royal Dutch Academy of Sciences (KNAW). The recent fraud cases in psychology do not only call the credibility of that particular field of science into question, but also reduce the reputation of social science research generally. The KNAW report urges universities to educate employees and students in academic honesty but does not suggest to implement a specific policy to detect fraud and other forms of academic misconduct. The diversity in research practices between disciplines makes it difficult to impose a general policy to detect and deter misconduct.

However, skeptics may view the reluctance of the KNAW to increase scrutiny as a way to cover up fraud and misconduct. Universities and science in general run a serious risk of losing their credibility in society if they do not deal with misconduct. With every new case that comes to light the public will ask: how is it possible that this case was not detected and prevented? Anticipating a large scale national investigation universities screen their employees using a list of risk factors for fraud and misconduct. This screening exercise may give a rough sense of how prevalent and serious academic misconduct is at their institution is. Below I give some suggestions for such risk factors, relying on research on academic misconduct.

At present it is unclear how prevalent and serious academic misconduct is at universities. It is difficult to obtain complete, valid and reliable estimates of the prevalence and severity of academic misconduct. Just as in crime outside the walls of academia, it is likely that there is a dark number for academic misconduct that does not come to light because there are no victims or because the victims or other witnesses have no incentive to report misconduct or an incentive not to report it. Relying on a survey among 435 European economists (a 17% response rate), Feld, Necker & Frey (2012) report that less than a quarter of all forms of academic misconduct is reported. There is no official registration of cases of academic misconduct. Cases of misconduct are sometimes covered by news media or by academics on blogs like Retraction Watch (http://retractionwatch.wordpress.com/). Using surveys, researchers have tried to estimate misconduct relying on self-reports and peer reports. In a Gallup 2008 survey among NIH grantees, 7.4% of the respondents reported suspected misconduct (Wells, 2008). Other surveys suggest a much higher incidence of misconduct. John, Loewenstein and Prelec (2012) conducted a study
among psychologists with incentives for truth-telling and found that 36% admitted to having engaged in at least one 'questionable research practice, a much higher incidence than the 9.5% reported by Fanelli (2009).

The research available shows that fraud is certainly not unique to experimental (social) psychology, as the high-profile cases of Stapel, Smeesters and Sanna from the University of Michigan seem to suggest. Fraud occurs in many fields of science. *Retraction Watch* profiles the cases of Jan Hendrik Schön in nanotechnology, Marc Hauser in biology, Hwang Woo-suk in stem cell research, Jon Sudbø and Dipak Das in medicine, and many other researchers working in the medical and natural sciences.

What forms of misconduct should be distinguished? Below is a list of behaviors that are mentioned in discussions on academic dishonesty and the code of conduct of the Association of Universities in the Netherlands (VSNU).

- *Fabrication of data.* Stapel fabricated 'data': he claimed data were collected in experiments while in fact no experiment was conducted and no data were collected. In less severe cases researchers data points are fabricated and added to a real dataset. List et al. (2001) report that 4% of economists admit having fabricated data. A similar estimate emerges from the more recent survey by Feld, Necker & Frey (2012). John, Loewenstein & Prelec report that 1.7% of psychologists admit fabrication of data. However, from this number, they estimate the true prevalence to be around 9%.
- *Omission of data points.* Smeesters admitted to have worked datasets such that the hypotheses were confirmed, e.g. by fabricating and adding 'data points' that increase the and by omitting those that reduced the p-value. John, Loewenstein & Prelec report that 43.4% of psychologists admit this.
- *Invalid procedures for data handling.* Errors in recoding, reporting or interpreting, inspired by and leading to support for the hypotheses. Research by Bakker & Wicherts (2011) shows this is quite common in psychology: 18% of statistical results in 2008 are incorrectly reported, commonly in the direction of the hypothesis favored by the author.
- *'Data snooping':* ending data collection before a target sample is achieved when a significant result is realized. This increases the likelihood of false positives or Type I errors (Strube, 2006). John, Loewenstein & Prelec report that 58% of psychologists admit this.
- *Cherry picking:* not reporting on data that were collected because the results did not support the hypothesis. John, Loewenstein & Prelec report that 50% of psychologists admit this. Cherry picking results in the *file drawer problem:* the 'unexpected' results disappear into a drawer.
- *'Harking':* Hypothesizing After Results are Known (Kerr, 1998). In a paper, reporting an unexpected finding as having been predicted from the start. John, Loewenstein & Prelec report that 35% of psychologists admit this.

All of the above forms of misconduct lead to artificially strong positive results that are difficult to replicate (Jha, 2012; Simmons, Nelson & Simonsohn, 2011). The *positive publication bias is*
enhanced by high-impact journals that want ‘novel findings’ and refuse to report (failed) replications. In addition to forms of misconduct that lead to positive publication bias, there are several other forms of misconduct:

- **Plagiarism.** Cut & paste of text without quotation marks and/or proper references.
- **Double publication.** Sending essentially the same manuscript to different journals without informing them and accepting simultaneous publication without cross-references.¹
- **Undeserved authorship.** Putting the name of a co-author on a paper who did not contribute to the paper. List et al. lumped undeserved authorship and sending manuscripts simultaneously to two journals and report that 7 to 10% of economists have engaged in this behavior.
- **Not disclosing conflicts of interest** (e.g., reviewing your own paper, a paper to which you contributed or a paper by a close colleague; sponsorship of the research by a party with interests in a certain outcome).
- **Not observing professional codes of conduct.** Each academic discipline has its own code of conduct. The content of these codes vary widely. Being aware of the code is phase 1; knowledge of its content is phase 2; observing it is phase 3.

**Trends in misconduct**

The recent Stapel and Smeesters cases suggest that misconduct is increasing. While Giner-Sorolla (2012) argues that the problems so vividly put on the agenda in this ‘year of horrors’ (Wagenmakers, 2012) are not new at all, Steen (2011) shows that the number of retractions of papers from academic journals covered in PubMed has increased sharply in the past years. These are the cases that form the tip of the iceberg because journal editors considered the evidence for misconduct so convincing. Whether the iceberg has in fact grown is not clear.

Fanelli (2012) shows that negative results are disappearing from most disciplines and countries published in ISI journal articles. Most troubling is that the proportion of positive results in journal articles from the Netherlands is stronger than in many other countries (OR: 1.16, reference category: US). Also the proportion is higher in the Social Sciences (OR: 2.14; reference category: Space Science) than in other disciplines – though less strong than in Neuroscience (3.16), Psychology (OR: 2.99) and Economics (2.65).

¹ Ironically, Bruno Frey, the third author of the Feld, Necker & Frey (2012) paper cited above, has engaged in this form of misconduct on several occasions. The Frey case is documented extensively by Olaf Storbeck on his Economics Intelligence blog (http://economicsintelligence.com/2012/03/19/self-plagiarism-bruno-frey-gets-away-with-a-slap-on-the-wrist/).
Characteristics of those who engage in misconduct

Little is known about the characteristics of those who engage in misconduct. List et al. (2001) find virtually no significant associations between self-reported misconduct and characteristics of economists. Stroebe, Postmes and Spears (2012) compared cases of academics caught for fraud and identified a set of common characteristics of these cases: the fraudsters were highly respected as researchers, published journal articles proficiently, were very quick in making their career, and had perfect datasets. Nosek, Spies & Motyl (2012) vividly illustrate the social dilemma for young researchers trying to build a career with novel findings that they cannot replicate. Pretty much the same sketch emerges from an analysis of retracted publications in PubMed (Steen, 2011). While Stapel and Smeeesters seem to have been isolated fraudsters, Steen (2011) find that a fraudster whose PubMed publication has been retracted “more frequently publishes with at least one co-author who also has fraudulent publications”.

What can be done to reduce misconduct?

Nosek, Spies & Motyl (2012) and Stroebe (Hamel, 2012, Witlox, 2012) are skeptical about self-correction in science. At present, the benefits of misconduct are too high and the risk of getting caught is simply too low. The fraudsters lined up by Stroebe et al. were able to pass peer review procedures because the procedures were not stringent enough. Reviewers should be more aware of the possibility of fraud (Matias-Guiu & García-Ramos, 2010).

Audits in which random samples of journal articles are drawn and the authors would be a solution because they increase the dection risk, Stroebe et al. argue. Food scientist Katan proposed such an audit at a KNAW conference on data sharing in 2011 (KNAW, 2012, p.47). However, audits are costly procedures.

Another recommendation is that replications should be made public. This has also been the dominant response in academic psychology to the Stapel case (Wicherts, 2011). Researchers are often unwilling or reluctant to share data (Wichterts, Borsboom, Kats & Molenaar, 2006). At present the incentives discourage researchers to share data. Researchers save time by not making their data available (Firebaugh, 2007; Giner-Sorolla, 2012). The costs required to make data available are often not budgeted. If research funders such as the Netherlands Organization for Scientific Research (NWO) impose a data sharing requirement this will create an additional cost for researchers. This makes it improbable that scientists will adopt the solution without force. At present, reluctance to share data indicates lower quality of research (Wicherts, Bakker & Molenaar, 2011).

While data-sharing is desirable for replication purposes, it is not something that universities can impose and only works in the long run. Journal editors and reviewers could insist on data-sharing, however. This also goes for the idea to require a power analysis for experiments (Ioannidis, 2005; Ioannidis & Trikalinos, 2007; Simmons, Nelson, & Simonsohn, 2011), the proposal that reviews are
published along the articles (Mooneyham, Franklin, Mrazek, & Schooler, 2012) and various other ideas proposed by Nosek & Bar-Anan (2012) such as a completely open access data repository. An even longer term proposal is that researchers pre-register their studies and indicate in advance the analyses they intend to conduct (Wagenmakers, Wetzels, Borsboom, Van der Maas, & Kievit, 2012).

Generally speaking, academic misconduct is likely to be more prevalent and more severe as the benefits of misconduct are higher, the costs are lower, and the detection risk is lower. Stroebe makes this point in two recent interviews (Hamel, 2012; Witlox, 2012). The increasing publication pressure in many sciences increases the benefit of misconduct (John, Loewenstein & Prelec, 2012). The lack of attention to details from overburdened reviewers, co-authors who are happy to score an additional publication and from dissertation supervisors loaded with work reduces the detection risk. The rat race increases the likelihood that the isolated cases of Stapel and Smeesters will be the ones who were stupid enough not to organize into a pack.

Lifelong anesthesia researcher Mutch (2011: 784) advises the following remedies against misconduct: “good mentoring, appropriately trained and diligent support staff, blinded assessment of data, data review by all investigators, a consensus agreement on data interpretation, a vigorous and independent Research Office, effective internal and external committees to assess adherence to protocols, and strong departmental leadership.”

Conversely, it is likely that in the absence of these conditions, there are more opportunities for misconduct. Given all of the above, I propose the following list of conditions that increase the potential for fraud and academic misconduct. The list can be used as a checklist or screening device for academic publications. Obviously an article with a high score is not necessarily fraudulent and requires more detailed attention. I encourage universities, journal editors, and reviewers to use this list, and to suggest additions or modifications. It is by no means intended to be a definitive list, and replication is necessary.
<table>
<thead>
<tr>
<th>Condition</th>
<th>Potential misconduct</th>
<th>Detection method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The researcher worked alone. Nobody else had or has access to the 'data'. Co-authors were not involved in the ‘data collection’ and/or ‘data analysis’.</td>
<td>Data fabrication as well as less serious forms of misconduct.</td>
<td>Ask co-authors.</td>
</tr>
<tr>
<td>2. The ‘data’ were not collected by others, but by the researcher.</td>
<td>Data fabrication.</td>
<td>Ask co-authors and co-workers.</td>
</tr>
<tr>
<td>3. There are no witnesses of the ‘data’ collection.</td>
<td>Data fabrication.</td>
<td>Ask co-authors and co-workers.</td>
</tr>
<tr>
<td>4. The raw ‘data’ (documents, fieldwork notes, questionnaires, videos, electronic data files) are not available (anymore). They are reported confidential, missing, lost, or located on a previous computer.</td>
<td>Data fabrication.</td>
<td>Ask author and co-authors, check data archive.</td>
</tr>
<tr>
<td>5. The statistics are ‘too good to be true’. The p-values of statistical tests are more often just below .050 than would be expected based on chance (Krawczyk, 2008; Simonsohn, 2012).</td>
<td>Data fabrication, selective omission of data points, cherry picking, and harking.</td>
<td>Compare p-values to expected distribution.</td>
</tr>
<tr>
<td>6. The research only finds support for the hypotheses (Fanelli, 2012).</td>
<td>Cherry picking and harking.</td>
<td>Count the proportion of hypotheses supported.</td>
</tr>
<tr>
<td>7. There is no fieldwork report or lab log entry available.</td>
<td>Data fabrication.</td>
<td>Check data archive, lab log, ask author.</td>
</tr>
<tr>
<td>8. Data are provided but original code or description of procedures followed by the author is not available or it is unclear for others how to replicate the research.</td>
<td>Data fabrication, cherry picking, harking.</td>
<td>Ask author.</td>
</tr>
<tr>
<td>9. Replication of the research is impossible with the available raw data and procedures and analyses described by the author.</td>
<td>Cherry picking, harking.</td>
<td>Ask author.</td>
</tr>
<tr>
<td>10. Replication of the research is possible but yields no support for the original findings.</td>
<td>Cherry picking, harking.</td>
<td>Try to replicate the findings.</td>
</tr>
<tr>
<td>11. The research appeared in high impact journals.</td>
<td>Misconduct with higher benefits.</td>
<td>Check impact factor.</td>
</tr>
<tr>
<td>12. The author is early in his/her career.</td>
<td>Misconduct with higher benefits.</td>
<td>Check career stage.</td>
</tr>
</tbody>
</table>
References


Ioannidis, J. P. A. (2005). Why most published research findings are false. Plos Medicine, 2(8), 696-701. doi:10.1371/journal.pmed.0020124


