



The Stapel Case: An Object Lesson in Research Integrity and Its Lapses

John M. Budd¹

1. School of Information Science & Learning Technologies, University of Missouri, Columbia, MO, Email: buddj@missouri.edu.

Abstract

In 2011 the works of Diederik Stapel came under suspicion. It was thought that he was fabricating and/or falsifying data and publishing the false results. The investigation was far-reaching and involved the three Dutch universities at which he worked. All of his published articles were examined carefully; to date (September 13, 2013), fifty-four of the articles have been retracted. The present study investigates the scope of citations to the articles he published from 1998 through 2006. These years mark the middle stage of his research and publication record, and they provide ample time for the examination of citations to his articles. The numbers and nature of the citations (positive or acknowledging problems) are reported, and the implications of the acceptance of his work are explored.

Keywords: Stapel, citations, scientific misconduct, social psychology, RCR (responsible conduct of research)

Introduction

Concern about the honesty and integrity of scholarship and publication has been increasing in recent years, and with good cause. Numerous works provide a background on the nature of scientific misconduct, its causes, and its results. D'Angelo offers a brief overview of the challenge, including several case studies of recent instances of misconduct (1). Judson examines the role of peer review in vetting the works that are eventually published, providing a historical and cultural background (2). Goodstein explains that detection of fraud and misconduct is far more easily said than done (3); the intricacies of fraud can transcend the system's ability to make ready identifications. Some tools exist to track instances of misconduct once they are identified. In addition to straightforward misconduct, the site www.retractionwatch.wordpress.com, a blog that tracks various kinds of problematic issues in

scholarly communication, reports, "Last year, an audit by the US Government Accountability Office found 'a potential for unnecessary duplication' among the billions of dollars in research grants funded by national agencies. Some researchers, it seemed, could be winning more than one grant to do the same research." (4). Given the increasing competition for external funding, the duplication is an egregious act of dishonesty. Examining the matter of research integrity from a different point of view, Harold R. Garner, Lauren J. McKiver, and Michael B. Waitkin found that there is likely to be duplication in awards for research, and the duplication could total as much as \$200 million (5). These findings illustrate that misconduct can take many forms, and can have any of various kinds of deleterious effects on research.

Another set of researchers examining the prevalence of questionable actions concentrated on a set of 2,000 psy-

chologists. John, Loewenstein, and Prelec assessed self-reporting of questionable behavior. A lengthy quotation of their findings illustrates the magnitude of the problem of questionable behavior:

We explored the response-scale effect further by comparing the distribution of responses between the two response-scale conditions across all 25 items and collapsing across the wording manipulation. Among the affirmative responses in the frequency-response-scale condition (i.e., responses of *once or twice*, *occasionally*, or *frequently*), 64% (i.e., .153/ (.151 + .062 + .023)) of the affirmative responses fell into the *once or twice* category, a nontrivial percentage fell into *occasionally* (26%), and 10% fell into *frequently*. This result suggests that the prevalence estimates from the BTS study represent a combination of single-instance and habitual engagement in the behaviors [emphasis in original] (6).

There are other difficulties noted by psychologists, and these challenges also make the identification of legitimate work a daunting task. Wicherts and colleagues attempted to obtain the data used in 141 published papers, but met with limited success (7). They contacted the corresponding authors of papers published in four major psychological journals. Their results were disheartening:

As the 141 articles included a total of 249 studies, we considered acquiring 90 to 100 data sets a realistic aim. We reasoned that adding a follow-up request after the original e-mail would take us a long way in that direction. Unfortunately, 6 months later, after writing more than 400 e-mails—and sending some corresponding authors detailed descriptions of our study aims, approvals of our ethical committee, signed assurances not to share data with others, and even our full resumes—we ended up with a meager 38 positive reactions and the actual data sets from 64 studies (25.7% of the total number of 249 data sets). This means that 73% of the authors did not share their data (7).

Among the implications of the study's results is the lack of replicability of published works. If studies cannot be repeated, or if data cannot be examined by others, the community has only the word of the researchers that protocols were followed, that appropriate methodologies were applied, and that findings can be found to be valid and reliable.

Concerns over publications that are eventually retracted are particularly troubling to the scholarly and research communities. The problem may be especially pronounced in biomedical research, as Budd, Coble, and Anderson observe (8). They point out that the numbers of retractions have been rising in recent years, and many of those retractions are due to scientific misconduct. Editors of eighteen journals which published articles by Joachim Boldt determined that eighty-eight of his papers should be retracted (9). Biomedicine is not the only field where retractions occur, though. *Retraction Watch* reports that Ulrich Lichtenthaler, of the University of Mannheim, has recently had a dozen papers retracted in management journals (10). Social psychologist Dirk Smeesters has also received publicity for problematic publications, some of which have been retracted (11). The case of Diederik Stapel is not unique, but the investigations into his publications by three universities offer a unique opportunity for investigation.

The beginnings of concern regarding Stapel's work

Beginning in approximately 1994 Diederik Stapel began a career that would propel him into stardom among social psychologists. His publication record was enviable, as was the total of citations to his work. By the end of 2011 he had published more than 100 articles. However, by late 2011, Stapel's world began to come falling down around him. Some of his publications were questioned for their validity. The initial questions led to a much more detailed scrutiny. The institutions at which he worked began to conduct a detailed examination of the entirety of his record. A trio of reports—the Levelt Committee, the Noort Committee, and the Drenth Committee—issued their lengthy final report (12). The report listed some of the specific problems that investigators mention, including:

- Generalizability of the findings from local to national and international culture,
- Verification bias and missing replications,
- Incomplete or incorrect information about the research procedures used,
- Statistical flaws,
- Failure of scientific criticism (12).

The report’s conclusion does not limit itself to the actions of Stapel; it includes severe criticism of the entire process of peer review, which did not pay sufficient attention to the above shortcomings. In particular, the authors of the report make mention of the prevalent pressure to publish and the role of editors and manuscript referees in fostering the problem of expansion of the literature. Also, since Stapel worked with a number of collaborators, the authors of the report urge journals to require all authors to state explicitly what their role was in the creation of the work.

A model for such a requirement exists and could be replicated widely. *The Journal of American Medical Association (JAMA)* insists that each author listed on a manuscript submit a detailed form stating precisely what the individual’s contribution was and that each author approves of the content (see, *JAMA* Sample Authorship Form (13)) The form requires each co-author to state explicitly that approval is given to the manuscript, that data will be provided upon request, and each listed author must indicate individual responsibilities (design, data gathering, drafting the manuscript, statistical analysis, etc.). Responsibility is thus explicitly shared.

In Stapel’s case, Wicherts has been quoted as stating that “Dr. Stapel was able to operate for so long. . . Because he was ‘lord of the data,’ the only person who saw the experimental evidence that had been gathered (or fabricated)” (see Carey, 2011 (14)). Benedict Carey points out that “In a prolific career, Dr. Stapel published papers on the effect of power on hypocrisy, on racial stereotyping and on how advertisements affect how people view themselves. Many of his findings appeared in newspapers around the world, including *The New York Times*, which reported in December on his study about advertising and identity” (14). In support of these conclusions, Jennifer Crocker and M. Lynne Cooper state, “APA began using an electronic manuscript management system in 2003. Between then and 2011 Stapel submitted 40 manuscripts; 24 were accepted and 16 were rejected. Given the number of editors and reviewers handling the manuscripts, it would be almost impossible to detect a pattern of data fabrication.” (15). Of interest is that the report found no culpability on the parts of co-authors, but the dissertation work by several students supervised by Stapel was called into question (15). To date, the doctoral degrees of twelve individuals are in jeopardy – on the other hand, several students’ dissertations were examined because they used Stapel’s data instead of gathering their own (15).

Stapel’s record

Data related to Stapel’s activities illustrate some of the concerns the scholarly community has with his work. A Scopus database search of Stapel by name yielded all publications, co-authors, and citations presented in Tables 1-3, below. The total record of Stapel’s publication and other activities (as of 24 February 2103) is reflected in Table 1:

Table 1. Selected Data On Stapel’s Publications

Publication History:	1994-present
No. of Articles (Scopus):	124
No. of Citations (Scopus):	1,756
No. of Retracted Articles:	54

Table 2 demonstrates that Stapel’s publications received numerous citations, most of them tacitly or substantively positive.

Table 2. Citations to Stapel’s Publications

Total Citations to Stapel’s Articles (1998-2006, n=39)	873
Substantive Citations (Substantive self-citations: 106)	320
Tacit Citations (Tacit self-citations: 71)	402
Not available for analysis:	151

Table 3 shows the journals in which Stapel published most frequently:

Table 3. Most Frequent Journals Stapel Published in

<i>J. of Personality and Social Psychology</i>	23
<i>European J. of Social Psychology</i>	22
<i>Personality and Social Psychology</i>	15
<i>J. of Experimental Social Psychology</i>	14
<i>Social Cognition</i>	9

The present analysis covered the years 1998 through 2006. These years mark the middle stage of his research and publication record, and they provide ample time for the examination of citations to his articles.

Table 2 makes mention of tacit citations; tacit citations to Stapel's publications are quite easy to identify. The following excerpt from is a common example, from Johnson and Stapel (so many of his citations are to his own work), *Harnessing social comparisons: When and how upward comparisons influence goal pursuit. Basic and Applied Social Psychology*, 31, 334-42.

Sample: Repeatedly, upward comparisons that threaten positive self-views have been shown to lead to performance improvements, especially when that threat arises because of extreme performance by the target (Johnson, Norton, Nelson, Stapel, & Chartrand, 2008; Johnson & Stapel, 2007a, 2007b), (p. 334).

Identifying the substantive positive citations require closer readings of the citing papers and also require some judgments. For example, the following quotation was deemed to be a substantive citation (Weber and Hertel. 2007. Motivation gains of inferior group members: A meta-analytical review. *Journal of Personality and Social Psychology*, 93, 973-93).

Sample: However, such partner information should not imply that the superior persons' performance is out of reach so that IGMs do not perceive impossibility in keeping up (Johnson & Stapel, 2007; Seta, 1982) (p. 988).

Another example of the substantive use of Stapel's work is as follows (Gibson and Poposki. 2010. How the adoption of impression management goals alters impression formation. *Personality and Social Psychology Bulletin*, 36, 1543-54):

Sample: Past research has demonstrated that activation of the self prompts social comparison processes (Stapel & Tesser, 2001), which can lead to contrast effects in judgments of the self (Stapel & Koomen, 2001a) (p. 1544).

One more example illustrates the extent to which social psychologists use Stapel's work as legitimate research (Hoffmann, Pennings, and Wies, S. 2011. Relationship marketing's role in managing the firm-investor dyad. *Journal of Business Research*, 64, 898):

Sample: A cooperative orientation activates an integration mindset that leads to assimilation in points without

congruence yet established (Stapel and Koomen, 2005), which reduces the likelihood of conflicts (p. 898).

As Table 3 indicates, numerous citing papers have accepted Stapel's work as scientifically valid and, at time, have even built their own research on Stapel's methods or findings. Few social psychologists can match the impact, as defined by citations received, that Stapel's work has garnered. The numbers, and the numbers of substantive citations point to a status as someone to whom others turned as they conducted their own work. This is not to say that citations represent a causal relationship between a scholar and her/his influence, but it is an undeniable indicator of such influence.

Additional implications

In addition to the difficulties Stapel has created for himself (as was mentioned above), ten dissertations by students Stapel supervised were found to contain fraudulent data, although those students were cleared of any wrongdoing in the inquiry. The report found that Stapel's colleagues and administrators seemed to accept his results at face value. Meanwhile, his high profile at Tilburg insulated him against initial rumblings about problems with his data. This phenomenon raises a question that probably cannot be answered through direct empirical investigation in many instances – do institutions deliberately ignore or cover up possibilities of problematic activities by faculty? If the answer is yes, there needs to be safety for those who would alert academic administrators to the presence of misconduct. Sovacool says,

institutional reforms need to be established to encourage the neutrality, anonymity, and protection of whistleblowers. Similarly, better auditing mechanisms need to be implemented to help motivate those researchers working in hostile research environments to come forward (16).

If there is to be a reduction in scientific misconduct there must be sufficient openness that whistleblowers can report what they witness with impunity. Any obstacles put in their way threaten the entirety of the scientific enterprise. As the Dutch newspaper *NRC Handelsblad* reported:

More important than the fraud of Stapel is that in the scientific world, no one has pulled the alarm about strange things in Stapel's publications," says Pim Lev-

elt today in *NRC Handelsblad*. Levelt is chairman of the committee that the work of Stapel in Tilburg investigated, such as commissions Noort and Drenth did in Groningen and Amsterdam respectively. “The whole system, from low to high, has failed. That is our shocking conclusion” (17).

Or, as another piece in the source suggests: Science has failed (12). The exercise of criticism, something that science par excellence preaches, is totally neglected. That said the chairman of the committee-Levelt afternoon during the presentation of the final report on the investigation of fraud dismissed professor Diederik Stapel (12).

The Stapel report could have farther-reaching implications relating to scientific misconduct. *NRC Handelsblad* has reported that the president of the Dutch Royal Academy of Sciences has demanded an expanded investigation into the case of Don Poldermans, a prominent cardiologist who lost his position at Erasmus University in the wake of misconduct probe (18). Poldermans has reportedly acknowledged misconduct but not fraud. Meanwhile, *Erasmus Magazine* has reported that Levelt has criticized Erasmus University’s handling of its own wayward social psychologist, Dirk Smeesters, who resigned in June amid concerns about the veracity of his data (19).

This marks the fall from grace of three of The Netherlands’ top scientists in barely a year. As this headline from NRC declares:

[The report is a] wake-up call. We don’t think that needs any translating. Updated 4:00 p.m. Eastern 11/29: Turns out Stapel will be releasing a book tomorrow. It’s called ‘Ontsporing,’ which means derailment, and, we suppose, is supposed to evoke the sense that this once-upstanding and legit researcher somehow jumped — or was knocked off? — the tracks of honesty and integrity into his career of deceit. We’re also told that the Stapel mess spawned a Dutch neologism: ‘slodderwetenschap,’ which means something like ‘sloppy science’ and refers to ineffective peer review and a culture of science that allows fraud to go undetected for so long (18).

Following Stapel

The difficulty presented to psychology (and to the social sciences in general) is well-stated by Leslie John and colleagues (2012): “[T]he prevalence of [questionable re-

search practices] raises questions about the credibility of research findings and threatens research integrity by producing unrealistically elegant results that may be difficult to match without engaging in such practices oneself” (6). As is mentioned above, their work examined responses by approximately 2,000 psychologists regarding their own activities. Many admitted to omitting confounding data, smoothing data, and even outright misconduct. Since the scientists themselves were responding the study has exception credibility and deserves to be taken extremely seriously (6).

Levelt adds a rather strident critique of scholarship in general—one that should be heeded in all disciplines:

The implications go beyond Stapel, says psycholinguist Willem Levelt, who chaired the Tilburg committee and coordinated the entire investigation. “We’re not saying all of social psychology is sloppy science,” he says. “But the fact that this could happen shows that the review process has failed from the bottom to the top.” Levelt believes the field is taking the message to heart, however. The report praises the “reproducibility project,” a large collaborative effort to replicate psychology studies set up by Brian Nosek of the University of Virginia in Charlottesville (*Science*, 30 March, p. 1558), as well as the November issue of *Perspectives on Psychological Science*, which is devoted to analyses of what ails the field and proposals to cure it (20).

In a turn of irony, Stapel, in one of the last papers he published before the evaluative investigation began (21), wrote that feedback from others is a major source of self-knowledge and esteem. One can only speculate that his own predicament influenced this paper. Battacherjee reports on the penalties suffered by Stapel: “Diederik Stapel, the former Tilburg University professor who fabricated dozens of research studies, has been spared a trial after reaching a settlement with Dutch prosecutors. He will do 120 hours of community service and forgo benefits from his former employer that would have been equivalent to 1.5 years of salary.” (22) Even though the final report on Stapel’s activities has been widely acknowledged as complete, detailed, and justified, not all social psychologists praise it. In a brief article in *Science*, the Executive Committee of the European Association of Social Psychology (EASP) calls some of the report’s conclusions “defamatory, unfounded, and false.” And social psychologist Wolfgang Stroebe of Utrecht University in

the Netherlands demanded an apology from the three investigation panels in a piece he wrote for his university's magazine (23).

It must be emphasized that this paper is in no way intended as an *ad hominem* attack on any individual, or to be derogatory to the field of social psychology. The purpose is to illustrate how an individual's position can be misused to conduct egregious publishing activities. That said, Stapel's case is also an object lesson from which social psychology and *all* disciplines can learn. *JAMA's* requirements for the reporting of responsibilities of all named authors can be a first step to ensure some measure of integrity in science. If any individual is going to sign such a voucher, that person should take the responsibility to review every component of the research. Another step that could be taken is a similar assurance to be signed by manuscript reviewers. The requirements could be, in some respects, similar to those of the authors. Reviewers must evaluate research questions thoroughly; must take a serious look at the literature review for obvious omissions that could have an effect on the paper as a whole; examine the proposed methodology very closely; pay special attention to data collection (means and outcomes); and ensure that conclusions are supported by all of the foregoing items. Many journals already have such requirements of reviewers, but the examinations of the reviews submitted must be carefully assessed so that the most informed publication decision can be made. Will these two recommendations solve all issues related to integrity? Probably not. If they reduce the instances substantially, they will have fulfilled a necessary goal.

Stapel self-published a book in Dutch entitled *Ontsporing* (or "derailment" in English translation), but this has done little to reclaim his reputation. As Battacharjee writes, "When Stapel's book came out, it got a mixed reception from critics, and it angered many in the Netherlands who thought it dishonorable of him to try to profit from his misdeeds. Within days of its release, the book appeared online in the form of PDFs, posted by those who wanted to damage his chances of making money." Battacharjee continues,

The unspooling of Stapel's career has given him what he managed to avoid for much of his life: the experience of failure. On our visit to Stapel's parents, I watched his discomfort as Rob and Dirkje tried to defend him. "I blame the system," his father said, steadfast. His argument was that Stapel's university managers and journal editors should have been watching him more closely (24).

Clearly, one of Stapel's problems was that he did not monitor his own actions very closely.

Competing interests

The author declares no competing interests.

References

1. D'Angelo J. 2012. Ethics in science: Ethical misconduct in scientific research. Boca Raton: CRC Press.
2. Judson HF. 2004. The great betrayal: Fraud in science. Boston: Houghton Mifflin Harcourt.
3. Goodstein D. 2010. On fact and fraud: Cautionary tales from the front lines of science. Princeton: Princeton University Press.
4. Retractionwatch. [Internet] [cited 2013 Nov 12]. Available from: <http://retractionwatch.wordpress.com/2013/01/30/has-double-dipping-cost-u-s-science-funding-agencies-tens-of-millions-of-dollars/>.
5. Garner HR, McKiver LJ, Waitkin MB. Research funding: Same work, twice the money? *Nature*. 2013; 493(7434): 599–601.
6. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*. 2012; 23(5): 524-532.
7. Wicherts JM, Borsboom D, Kats J, Molenaar D. The poor availability of psychological research data for reanalysis. *American Psychologist*. 2006; Oct;61(7): 726-728.
8. Budd JM, Coble ZC, Anderson KM. Retracted publications in biomedicine: Cause for concern. ACRL Conference, Philadelphia, PA: 2011. Available from: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/national/2011/papers/retracted_publicatio.pdf.
9. Editors-in-Chief Statement Regarding Published Clinical Trials Conducted without IRB Approval by Joachim Boldt. *Anesthesia & Analgesia*. [Internet] 2011 March 12 [cited 2013 Nov 12]. <http://www.aeditor.org/EIC.Joint.Statement.on.Retractions.pdf>.
10. Retractionwatch. What's new is not new again: Ulrich Lichtenthaler retracts eighth paper. [Internet] [cited 2013 Nov 12]. Available from: <http://retractionwatch.wordpress.com/2012/11/21/whats-new-is-not-new-again-ulrich-lichtenthaler-retracts-eighth-paper/>.
11. Retractionwatch. Retraction three for Dirk Smeesters. [Internet] [cited 2013 Nov 12]. Available from: <http://retractionwatch.wordpress.com/2012/12/01/retraction-three-for-dirk-smeesters/>.

12. The Levelt Committee, The Drenth Committee & The Noort Committee. Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. [Internet] 2012 Nov 28. [cited 2013 Oct. 28] Available from: <https://www.commissielevelt.nl/>.
13. Journal of the American Medical Association. Sample Authorship Form. [Internet] 2013 Oct 10 [cited 2013 Oct. 28]. Available from: http://jama.jamanetwork.com/data/ifora-forms/jama/auinst_crit.pdf.
14. Carey B. Fraud case seen as a red flag for psychology research. New York Times [Internet] 2011; 2 Nov. Available from: <http://www.sakkyndig.com/psychologi/artvit/nytimes2011.pdf>.
15. Crocker J, Cooper ML. Addressing scientific fraud. Science Dec. 2011; 334(6060): 1182. Available from: <http://www.sciencemag.org/content/334/6060/1182.short>.
16. Sovacool BK. Using criminalization and due process to reduce scientific misconduct. Am J Bioethics. 2005; 5(5): W1-W7.
17. Berkhout K. Fraudecommissies Stapel: De sociale psychologie faalde. (Trans: The Fraud Commission on Stapel: Social psychology failed.) NRC Handelsblad. [Internet] 2012. [cited 2013 Sept. 16]. Available from: <http://www.nrc.nl/nieuws/2012/11/28/fraude-commissies-stapel-de-sociale-psychologie-faalde/>.
18. Berkhout K, Rosenberg E. Nieuw geval van wetenschapsfraude – hoogleraar Erasmus MC ontslagen. (Trans: New case of scientific fraud: Professor Fired.) NRC Handelsblad. [Internet] 2011 Nov 17. [cited 2013 Sept. 16]. Available from: <http://www.nrc.nl/nieuws/2011/11/17/nieuw-geval-van-wetenschapsfraude-hoogleraar-erasmus-mc-ontslagen/>.
19. Erasmus Magazine. Fraude van Stapel is verbijsterend. (Trans: Fraud by Stapel is baffling.) [Internet] 2012. [cited 2013 Sept 16]. Available from: <http://www.erasmusmagazine.nl/nieuws/detail/article/4201-fraude-van-stapel-is-verbijsterend/>.
20. Enserink M. Final report on Stapel blames field as a whole. Science. 2012; 338(6112):1270-1271.
21. Noordewier MK, Stapel DA. When failure feels better than success: Self-salience, self-consistency, and affect. Br J Soc Psychol 2011; June; 51(2):321-30.
22. Battacharjee Y. The mind of a con man. New York Times. [Internet] 2013 April 26. [cited 2013 Sept 16]. Available from: http://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html?pagewanted=all&_r=0.
23. Social psychologists protest Stapel report “attack”. Science 2012; Dec.; 338(6113): 1401.
24. Battacharjee Y. Stapel gets community service for fabricating studies. ScienceInsider. [Internet] 2013 June 28. [cited 2013 Sept. 16]. Available from: <http://news.sciencemag.org/europe/2013/06/stapel-gets-community-service-fabricating-studies>.