

Sharpening Scientific Policy After Stapel

Paul A.M. Van Lange

VU University, Amsterdam, The Netherlands

Abraham P. Buunk

University of Groningen, The Netherlands

and Royal Netherlands Academy of Arts and Sciences (KNAW)

Naomi Ellemers

Leiden University, The Netherlands

and

Daniel H.J. Wigboldus

Radboud University Nijmegen, The Netherlands

Internal Report, September 2012

Abstract

The fraud by Diederik Stapel, uncovered in September 2011, ranks among the most severe cases of scientific misconduct in the history of science. While acknowledging the comprehensive nature of research ethics and integrity, this report discusses the specific Stapel case with the aim of identifying concrete steps that can help reduce the risk of fraud or other forms of scientific misconduct in psychological science and other disciplines. These pertain to managerial aspects of science (the accessibility of an ethics officer, provision of facilities for data storage), communicative aspects of science (public acknowledgement of contribution to an article, guidelines for communication with media), and issues relevant to peer review and journal policy.

Sharpening Scientific Policy After Stapel

The scientific fraud conducted by Diederik Stapel (Tilburg University) has shocked many people around the world. It has been intensely discussed in various academic communities, and received coverage from numerous public media around the globe. All of this is understandable: By virtue of the sheer number of published articles that contain fictitious data (at least 30 articles), and the time span during which this fraud occurred (at least 7 years), Stapel is potentially one of the most severe fraudsters in recent decades. The initial report by the Committee Levelt concluded that the fraud included not only his work during his professorship in Tilburg (2006-2011), but also his work during his professorship in Groningen (2000-2006), and that the fraud has been happening at least since 2004 (Report Committee Levelt, 2011).

How was the fraud committed? The Report Committee Levelt refers to his methods as “cunning” (p. 7). The report notes that in the initial stages of the fraud, his methods were somewhat more covert. But it also appears that his fraud became more and more brazen over the years. One of his methods was as follows: after some research meetings, Stapel and his collaborators would design questionnaires or experiments to examine a new research question. Research assistants or PhD-students would then prepare all necessary research materials, such as questionnaires and other materials needed for the research. Then, Stapel singlehandedly collected datasets at schools. He also collaborated with many colleagues for whom it was difficult to monitor his handling of the data, because they worked in a different department at a different university, or in a different country. Some weeks later, Stapel shared a data-set (which in several cases was fabricated, as we know now), which often provided good or excellent support for their hypotheses. On other occasions, Stapel approached people to collaborate with him on “old” datasets (collected at a previous university). Finally, Stapel sometimes took over (parts of) the data

collection by claiming that a research assistant would collect the data for him. In each of the fraudulent form of data fabrication, Stapel operated singlehandedly. This practice affected parts of several PhD-theses, as well as some joint publications with senior colleagues, but there were also PhD-theses (7 of the 19) that were cleared in the initial report (see Report Committee Levelt).

In some of the PhD-projects, Stapel discouraged or prohibited participation by PhD-students in the data collection. In retrospect, this might be interpreted as a strong signal of fraud. Further suspicion could be raised by the fact that, results from data collected by Stapel were often in strong support of the hypotheses, revealing effect sizes for predicted effects (and negligible effects of many non-predicted effects) and (virtually) no missing data (see Report Committee Levelt). Indeed, towards the end, there was a growing suspicion among junior research collaborators, who after months of careful examination (Keulemans, 2012), reported it to the Head of Department during an international conference, who reported it to the Rector of Tilburg University one day after his return from the conference (26 August, 2011). Stapel was suspended from Tilburg University shortly thereafter on September 7, 2011, only a few days after he admitted to the fraud.

It is not easy to understand how the fraudulent data collection by Stapel might have started and how it could have continued for so long. We suggest the following reasons. First, in terms of scale, magnitude, and level of organization, the fraudulent data collection by Stapel may be assumed to be truly exceptional. Also, peer review in social psychology is time-consuming (sometimes it takes years) as part of editorial processes involving rejections and “revise-and-resubmit” decisions before eventual publication - and before any effort toward replication (methodological or conceptual) can be undertaken by the broader field of colleagues. Second,

Stapel had a strong scientific reputation, in terms of quality, productivity, as well as scientific leadership and judgment, and he held various influential positions (such as Scientific Director of Tilburg Institute for Behavioral Economics Research, Dean of the Faculty of Social and Behavioral Science at Tilburg University, and serving as an Associate Editor for *British Journal of Social Psychology*, *Personality and Social Psychology Bulletin*, and *Psychological Science*). Moreover, such positions enhance power, and Stapel could be intimidating when PhD-students asked critical questions about the data or the data collection (Committee Levelt report, 2011). Stapel was also seen as an excellent organizer and time-manager (and so people might have assumed that perhaps help in data collection was not needed), he taught a course on research ethics, he collaborated with scholars who themselves had excellent reputations, and he was well-liked and admired by many of his colleagues. Such qualities likely undermine the suspicion of fraud. In the absence of concrete evidence, it is not easy to accuse a colleague or supervisor of fraud, and it may take time to find the evidence that is needed to do so. It is perhaps not coincidentally that a group of junior scientists (Research Assistant, PhD students) detected it, and prepared it for several months to build the case and then report it (Keulemans, 2012).

Recommendations for sharpening scientific policy

Before discussing the concrete recommendations, three general comments that put the recommendations in perspective are in order. First, there is a strong commitment to research ethics and integrity, as they are recurring themes in teaching, research meetings, and informal consultation, as well as various scientific associations around the world. Indeed, many associations such as the American Psychological Association (APA) have advanced codes for scientific conduct in psychological science, and journals of the Association of Psychological Science, such as *Psychological Science* and *Perspectives on Psychological Science*, are member

of the Committee of Publication Ethics (COPE). Moreover, the topic of research ethics has been captured in many articles and books. It even happened that in September 2011, the month in which the fraud by Stapel was detected, the American Psychological Association published a *Handbook of Ethics in Psychology*, devoting more than 1000 pages on the topic (Knapp, 2011). And in the same year, the network of All European Academies (ALLEA), with the European Science Foundation, published a Code of Conduct for Scientific Integrity (ESF/ALLEA, 2011) that is relevant to all scientific disciplines. Thus, there is a longstanding and ongoing debate about research ethics and integrity, during which scientific policy has been subject to continuous revision, updating, and refinement.

Second, like most academic fields, social psychology does not rely on a single research paradigm or method. While the focus of mainstream social psychology is on the experimental method (mostly in the laboratory), a range of dependent measures is used to assess behavior, thought or judgment processes, affective processes, psycho-physiological measures, neurological processes. This multi-method approach is often used to provide comprehensive tests of a theory, or a hypothesis, in areas such as social cognition, attitudes, decision-making, social aspects of emotion and motivation, interpersonal relations, social identity, human cooperation, and group processes. Moreover, like most academic fields, social psychology is an internationally oriented field, and has become increasingly interdisciplinary, with growing links in theory and methods to disciplines such as cognitive science, communication science, consumer science, decision-making, economics, evolutionary science, health science, management science, and neuroscience.

Third, as the title of this report clearly conveys, this report is written in response to the fraud conducted by Diederik Stapel. Since Stapel, there have been two other, independent

accusations of fraud involving Lawrence Sanna (Department of Psychology, University of Michigan) and Dirk Smeesters (Rotterdam School of Management, Erasmus University Rotterdam) who have published in journals of social psychology. Also, this report is finalized before the final report by Level appears. Hence, our recommendations are primarily “inspired” by the Stapel case, and not directly influenced by the two later suspects of fraud or any specific information in the final report by Levelt, which is expected to be released in November 2012. That said, we do realize that fraudulent behavior in science, even only in psychological science, is far more complex - and more multifaceted - than the Stapel case as such. But sometimes, it is useful to focus on a particular case to explore the ways in which a particular field, and neighboring fields and disciplines, can improve to help reduce the risks of fraud and other forms of scientific misconduct.

Thus, our aim is to identify concrete steps that can help reduce the risk of fraud or other forms of scientific misconduct in psychological science and other disciplines. However, taking into account the interdisciplinary, international orientation of social psychology and other fields of psychology, where universities have different norms, structures, and laws, we realize that there are several practical issues or obstacles that might to some degree constrain the “implementability” of our recommendations. And taking into account the longstanding nature of discussions about research ethics and integrity, we hope that our recommendations make a contribution by inspiring further scientific debate about research ethics and scientific integrity, and, when feasible, sharpening scientific policy in psychological science and other disciplines.

1. Increase the accessibility and visibility of an ethics officer in a department or faculty, who can be consulted in complete confidence after observation or suspicion of scientific misconduct. We have seen that in Stapel’s case, it is ultimately the communication of research

assistants and PhD students among each other and to higher authorities that uncovers the fraud. Indeed, in cases where fraud is detected, it is not uncommon that very junior researchers have taken the first step, alerting authorities with their suspicions. This officer therefore needs to be someone who is easily accessible to graduate students and senior scholars, and ideally approachable by scholars from other institutions or by scholars from other countries (such as the numerous international contacts of Stapel who were defrauded). Finally, the ethics officers' actions should be guided by a transparent protocol and/or code of conduct, in order to assure complainants that due process will be observed.

2. Provide central facilities at the department for the storage of data. This facilitates sharing of data from experiments among members of a well-defined research team, which can be stored at the department of the university where the funding is located. Moreover, we recommend that it becomes common practice for all authors of a research publication to share among each other not only the experimental materials, but also the raw data on which the article is based, as well as details about the statistical procedures and analyses applied. In working together, it could become normative (and reinforced by management) that at least two different members of a research team are always involved at each major stage of the research process (such as data collection, data entry, data analysis, data management, the write-up) so as to reduce the risk scientific misconduct and errors.

3. Explicitly acknowledge and recognize the contributions in an article - that is, how the work for a scientific article was distributed among the authors (e.g., conducting of studies, data-analysis, and writing of manuscript). Recently, this policy has been implemented by a few journals, and we hope other journals follow. Erroneous or inappropriate actions may occur at

different stages of the data-collection, the data-analyses, as well as the writing, and for these well-defined task domains the contributions can be explicated in an author note.

4. Adhere to norms of integrity in communication with the media. We recommend that until manuscripts involving empirical research have been peer-reviewed and been accepted for publication in a scientific journal, researchers should not contact the media or communicate their findings to the general public. When impossible (e.g. when a journalist reports on a paper presentation at a conference), we recommend communicating explicitly that the research is as yet unpublished. These recommendations are inspired by the fact that one study in which Stapel collaborated (which supposedly demonstrated that reminders of meat enhance selfish motives) was presented in the (Dutch) media before it was accepted for publication, which not only caused misunderstandings but also undermined public confidence in scientific research in psychology more generally. Peer review provides checks and balances, even though we acknowledge that a paper accepted for publication is not an absolute guarantee for scientific quality (and not all research aspects can be checked through peer-review).

5. Reviewers of manuscripts submitted for publication should pay more critical attention to the procedural and descriptive aspects of research, such as details of the sample, recruitment of participants, and treatment of missing data. For example, Stapel faced very few challenges when publishing data that were collected at high schools, without describing the name and location of the high school, and without providing specific information about the data collection processes (such as who collected the data). We are aware that sometimes such information cannot be provided, for example, because of confidentiality, as may be true for research in organizations, or because of privacy of the sample. If confidentiality is an issue, the authors

could explicitly note this in the article. Also, the authors might ask permission from those concerned if they may confidentially share such information with the editor of the journal.

6. Replicability is a necessary condition for scientific progress. Repeated failure of replication should find its way into the scientific literature. Journal editors might show greater openness to publish articles that report repeated failures of replication, as well as papers that "simply replicate" results from prior publications, often as a first step in addressing further issues. The possibility to create or foster outlets that aim to publish statistically nonsignificant (but meaningful) findings might be further explored. Through meta-analytic approaches or alternative methods, greater recognition of demonstrated failures of replication might help not only in challenging the validity of earlier published research but also in detecting errors or forms of scientific misconduct.

References

ESF/ALLEA (2011). *The European Code of Conduct for Research Integrity*. Strasbourg, ESF/Amsterdam.

Keulemans, M. (2012). Ontmaskering frauderende Stapel werd maandenlang voorbereid. *De Volkskrant* (January 21). (Unmasking fraud by Stapel was prepared months in advance).

Knapp, S. (2011, Eds). *APA Handbook of ethics in psychology*. New York, American Psychological Association.

Report Committee Levelt (2011). *Interim rapportage inzake door Prof.dr. D.A. Stapel gemaakte inbreuk op wetenschappelijke integriteit*. (Interim report regarding the breach of scientific integrity by Prof.dr. D.A. Stapel).

Author Note

The authors decided to write this paper after informal and formal discussions at the scientific associations to which they are linked. This report is written by the four authors in their personal capacity, and not on behalf of any department, scientific association, or academy to which they are affiliated. For readers less familiar with social psychology, we provided an Appendix that describes the field of social psychology. We are grateful to Marcel Zeelenberg, Head of Department at Tilburg University, Philip Eijlander, Rector Tilburg University, and the three students who detected the fraud, for sharing or confirming information about the reporting of the fraud, and to Monica Biernat, Fabrizio Butera, Pieter Drenth, David Dunning, Agnetta Fischer, Jens Förster, Michael Hogg, Robert Kail, Daniel Lakens, Tom Postmes, Kees Schuyt, Iris Schneider, Jeffry Simpson, Eliot Smith, Linda Steg, Robert Sternberg, Kees van den Bos, Eric van Dijk, Jos van Lange, Joop van der Pligt, and Nico van Yperen for providing comments on an earlier version of this paper.

Appendix: The field of social psychology

Social psychology is a field that is aimed at understanding how individuals' behavior, motivations, thoughts, and feelings are influenced by other people, real or imagined. Common themes that are examined by social psychology include, for example, the workings of stereotyping and impression formation (e.g., how do people form impressions and judgments about categories, about personality), implicit influences (e.g., how can social behavior be influenced unconsciously), attitudes and persuasion (e.g., how can one change attitudes), attribution and social judgment (e.g., how do people understand the causes of social events), interpersonal relations (e.g., what makes relationships healthy and stable), human cooperation (e.g., what are the determinants of selfish versus other-regarding behavior), social comparison (e.g., why do we want to compare ourselves to others), need for approval and affiliation (e.g., how do people deal with social exclusion), social identity and diversity (e.g., why is the self so closely linked to success or failure of the group), intergroup processes (e.g., what are the causes of intergroup conflict), and social performance (e.g., how can others promote or undermine our task motivation and behavior). Social psychology shares interests for specific themes and methods with other disciplines and fields, such as cognitive science, experimental economics, consumer science, evolutionary science, neuroscience, organizational psychology, management science, marketing, sociology, and sport psychology and movement science.

Social psychology is a field in which numerous research techniques are used. The most commonly used methodology is experimental research in the laboratory, especially in the last two decades. Participants are invited to the laboratory that consists of a number of cubicles (roughly, from six to fifteen) each equipped with a personal computer. The researcher has often written a computer program that includes the experiment. Typically, an MA or PhD student

writes the computer program on the basis of ideas and materials provided by the student, the primary supervisor, or both. The supervisor(s) discusses the computer program with the student. An MA student typically has one primary supervisor with whom the student discusses the experimental material. A PhD student (at least in the Netherlands) is most often supervised by two supervisors, although it does happen that they sometimes are supervised by one supervisor. The situations that are studied can differ in manifold ways, from assessing responses to subtle cues on the computer screen to judgments of different faces, to experimental situations in which people need to make a choice between cooperation and selfishness. Also, the responses can vary in terms of assessing emotions, thoughts, judgments, response latencies, physiological measures (such as heart rate variability, blood pressure, or hormonal changes) as well as fMRI research. These methods are also used by scientists working in other fields and disciplines, such as different fields of psychology (such as cognitive psychology, psychophysiology, economics (such as experimental economics, marketing, consumer behavior), and evolutionary science (such as evolutionary biology).

Although perhaps less often than in the past, social psychology also uses questionnaires in the field. For example, social psychologists interested in organizational processes might study the predictors of organizational commitment, or how feelings of justice might promote trust in management. Social psychologists interested in close relationships may use questionnaires to establish the factors that predict and help explain relationship satisfaction, breakup, and psychological well-being. In addition, questionnaires are often used in cross-cultural studies to examine differences and similarities in social psychological processes across cultures. Sometimes, some specific samples might be studied, such as professional soccer players to understand cohesion and individual and team performance in sport teams, or children

of varying ages at elementary schools, to understand developmental patterns in social behavior.

It is notable that it is not common practice for social psychologists to collect data at high schools, even though it does happen (and the field knows that some researchers go to schools to collect data). However, it is notably difficult to get access to schools to collect data at schools.

Questionnaires are most often administered once, even though social psychology increasingly focuses on longitudinal projects with repeated sessions over time (“waves) are designed. These “field” approaches seem more common in some other fields of psychology, medicine, sociology, political science, and economics. As noted earlier, the large majority of publications in social psychology involve experiments conducted in the laboratory at the University.