

# Flawed science: The fraudulent research practices of social psychologist Diederik Stapel

Levelt Committee

Noort Committee

Drenth Committee

This document is an English translation of the Dutch report '*Falende wetenschap: De frauduleuze onderzoekspraktijken van social-psycholoog Diederik Stapel*'. In the event of any differences between the Dutch report and the translation, the Dutch report will prevail.

28 november 2012



# Content

Foreword	5
<b>1 Introduction and background</b>	<b>7</b>
1.1 Levelt Committee: Terms of Reference	7
1.2 Noort Committee: Terms of Reference	7
1.3 Drenth Committee: Terms of Reference	8
1.4 Statistical support	9
1.5 Background to the inquiry	9
1.6 Nature of the report	10
<b>2 Working method of the Committees</b>	<b>11</b>
2.1 Committees' collaboration and distribution of work	11
2.2 Formal status of Committees	11
2.3 Procedure	12
2.3.1 <i>Information sources</i>	14
2.3.2 <i>All-embracing enquiry</i>	15
2.3.3 <i>Conclusion per publication</i>	15
2.3.4 <i>Hearing both sides of the case: general</i>	16
2.3.5 <i>Hearing both sides of the case: Mr Stapel</i>	16
2.4 Identifying fraud	17
2.4.1 <i>The definition of fraud</i>	17
2.4.2 <i>Proof and evidence of fraud</i>	17
2.5 Concrete working method: explanation and examples	18
2.5.1 <i>Organization and conduct of the experiment</i>	18
2.5.2 <i>Peculiarities in the data matrix</i>	19
2.5.3 <i>Analysis and re-analysis of data</i>	19
2.5.4 <i>Publications</i>	20
<b>3 Findings regarding the nature and extent of the fraud</b>	<b>25</b>
3.1 Extent of the fraude	25
3.2 Nature of the fraud	31
3.3 Reports of serious infringement of scientific integrity	32
3.4 Co-authors	33
3.5 Dissertations Tilburg	33
3.6 Dissertations Groningen	34
3.7 Impact of the fraud	34
<b>4 Research culture: working method and research environment</b>	<b>37</b>
4.1 Introduction	37
4.2 General	37
4.3 Working method and research environment: University of Amsterdam	38

4.4	Working method and research environment: University of Groningen	39
4.5	Working method and research environment: Tilburg University	43
5	Research culture: flawed science	47
5.1	Introduction	47
5.2	Generalizability of the findings from local to national and international culture	47
5.3	Verification bias and missing replications	48
5.4	Incomplete or incorrect information about the research procedures used	51
5.5	Statistical flaws	52
5.6	Failure of scientific criticism	53
6	Recommendations	55
Appendix 1	Lists of Committee members and statisticians	61
Appendix 2	Noort Committee questionnaire	63
Appendix 3	People interviewed by the Committees	65
Appendix 4	List of publications examined by the Levelt Committee including findings	69
Appendix 5	List of publications examined by the Noort Committee including findings	81
Appendix 6	List of publications examined by the Drenth Committee including findings	97
Appendix 7	Authority to withdraw a doctorate	101
Appendix 8	Literature references	103

# Foreword

The scientific fraud committed by Diederik Stapel, which came to light in Tilburg in early September 2011, sent shock waves across the academic world in the Netherlands and internationally. Ultimately, trust forms the basis of all scientific collaboration. If, as in the case of Mr Stapel, there is a serious breach of that trust, the very foundations of science are undermined.

Is science's purported self-cleansing ability really up to the challenge of dealing with such a serious infringement of scientific integrity? From the outset, the Rector Magnificus of Tilburg University decided to opt for full disclosure. He established a committee of enquiry, the Levelt Committee, charged with establishing the nature and extent of the fraud and examining the research culture that had allowed this fraud to persist over such a lengthy period. This picture could not be complete without involving Mr Stapel's previous places of employment within the scope of the investigation: the University of Amsterdam (UvA), where he studied from 1993 until 1999 and was awarded his doctorate in 1997, and the University of Groningen, where he held the position of full professor from 2000 until 2006. In mutual consultation between the three universities, the Drenth Committee was established in Amsterdam and the Noort Committee in Groningen. In this final report, the three Committees present their findings jointly.

When the fraud was first discovered, limiting the harm it caused for the victims was a matter of urgency. This was particularly the case for Mr Stapel's former PhD students and postdoctoral researchers, whose publications were suddenly becoming worthless. In an interim report, published in late October 2011, the Committees reported on the extent of this harm and also concluded that none of these authors or co-authors had been accessories in Mr Stapel's fraud. The report also outlined the nature of the fraud and of the research culture in which it had occurred. This interim report was made public immediately.

However, the Committees were of the opinion that the main bulk of the work had not yet even started. The case involved an extensive body of scientific work that was now tainted. In a well-functioning scientific world, the task now was to separate the wheat from the chaff. Journal publications can often leave traces that reach far into and even beyond scientific disciplines. The self-cleansing character of science calls for fraudulent publications to be withdrawn and no longer to proliferate within the literature. In addition, based on their initial impressions, the Committees believed that there were other serious issues within Mr Stapel's publications, even if they were not fraudulent. These issues are referred to in this report as 'sloppy science': a failure to meet normal standards of methodology. This brought into the spotlight a research culture in which this sloppy science, alongside out-and-out fraud, was able to remain undetected for so long. This also necessitated a scientific self-cleansing operation.

It is to the credit of the three universities involved that they gave the Committees a free hand to analyse all of Mr Stapel's publications. In this process, cooperation was requested from the co-authors. They all offered this without exception, by submitting research materials and data and in many cases by informing us in interviews of the circumstances of their research. In addition, all three universities placed statisticians at the disposal of the Committees in order to subject all of Mr Stapel's publications to a rigorous analysis. Obviously, this was a labour-intensive and expensive operation. As far as we are aware, this was the first time ever that the whole of a fraudster's body of scientific work had been scrutinized in this way.

Committees usually restrict themselves to an investigation of articles reported to be suspect or to a small random sample. In many cases, this can indeed deliver proof of fraud. However, there is always a greater interest at stake. The scientific literature must be cleansed of everything that is fraudulent, especially if it involves the work of a leading academic. Of course, this is of particular importance in medicine, where fraudulent publications can have harmful consequences for large groups of patients. But it is also the case for less applied sciences, including social psychology. No more time and money must be wasted on replications or meta-analyses of fabricated data. Researchers' and especially students' too rosy view of the discipline, caused by such publications, should be corrected. In this particular case, journals, reviewers, assessment committees and graduate schools must learn a methodological lesson: it appears that too much can go wrong in the discipline's critical functions.

Nevertheless, the most important reason for seeking completeness in cleansing the scientific record is that science itself has a particular claim to the finding of truth. This is a cumulative process, characterized in empirical science, and especially in psychology, as an empirical cycle, a continuous process of alternating between the development of theories and empirical testing. A theory has a provisional claim on truth/validity, as long as it has not been empirically disproven. Ultimately, cumulative evidence can result in consensus within the peer community on a theory's validity. This fundamental cumulative process is seriously disrupted by interference from fraudulent data and findings based on questionable methodology. The scientific researchers and institutions involved are duty bound to call a halt to this disruption.

Sadly, some of the Committees' findings are echoed in the recent open e-mail on the current situation in social psychology published by Nobel prizewinner Daniel Kahneman: 'Your field is now the poster child for doubts about the integrity of psychological research. Your problem is not with a few people who have actively challenged the validity of some priming results. It is with the much larger population of colleagues who in the past accepted your surprising results as facts when they were published. These people have now attached a question mark to the field, and it is your responsibility to remove it.' Kahneman can see 'a train wreck looming. I expect the first victims to be young people on the job market.' If this report and the wrongdoings it highlights can contribute to this cleansing process, the Stapel fraud may still have a positive impact, alongside all of the damage and personal harm it has caused.

The Committees would like to express their gratitude to the three University Rectors for their actual and moral support in their work. They would also like to thank the devoted statisticians who conducted their painstaking analyses often under extreme pressure. Without the open discussions with co-authors and with others involved, the in-depth analysis included in this report would not have been possible. Especially for the victims of this fraud, these discussions were often fraught with emotion. Therefore, their contributions to the work of our Committees deserve a special mention of gratitude. Finally, the Committees offer their thanks to their excellent executive secretaries, who have organized the work of the Committees for more than one year and are responsible for much of this report. The names of all of the people who have supported the work of the Committees in this way are listed in Appendix 1 of this final report.

This final report now brings the work of the Committees to its conclusion.

Willem J.M. Levelt, *Chair of the Coordinating Committee*

# 1 Introduction and background

In early September 2011 the Executive Board of Tilburg University suspended Prof. D.A. Stapel from his duties with immediate effect, shortly after which he was dismissed. Mr Stapel was found to have committed a serious infringement of scientific integrity by using fictitious data in his publications, while presenting the data as the output of empirical research. The Rector Magnificus of Tilburg University, Prof. P. Eijlander, formed the Levelt Committee on Friday, 9 September 2011 in order to investigate this infringement of scientific integrity. In view of Mr Stapel's earlier appointments at the University of Amsterdam and the University of Groningen and the serious suspicion that data fraud had also been committed in the course of these appointments, the inquiry was extended beyond Tilburg University. Mr Stapel had by then admitted to the Rector of Tilburg University that he had also committed fraud with data in Groningen. The Rector Magnificus of the University of Amsterdam, Prof. D.C. van den Boom, established the Drenth Committee and the Executive Board of the University of Groningen, in the person of the Rector Magnificus, Prof. E. Sterken, established the Noort Committee.

## 1.1 Levelt Committee: Terms of Reference

The Rector Magnificus of Tilburg University gave the Levelt Committee the task of investigating the extent and nature of the infringement of scientific integrity committed by Mr D.A. Stapel. There are two elements to the task.

1. The Committee should examine which publications are based on fictitious data or fictitious scientific studies, and during which period the misconduct took place.
2. The Committee should offer a view on the methods and the research culture that facilitated this infringement, and make recommendations as to how a recurrence may be prevented.

The members of the Committee are Prof. W.J.M. Levelt (chair), director emeritus of the Max Planck Institute for Psycholinguistics and former president of the Royal Netherlands Academy of Arts and Sciences; Prof. M.S. Groenhuijsen, full professor of criminal law, criminal procedure and victimology (Tilburg University) and founder-director of the interfaculty research institute Intervict; and Prof. J.A.P. Hagedaars, emeritus professor of Methods and Techniques of Social Scientific Research and former dean of the Faculty of Social Sciences (Tilburg University). The Committee's administrative secretary is Dr S.A.M. Baert. (mathematician and research policy officer at Tilburg University).

## 1.2 Noort Committee: Terms of Reference

The Executive Board of the University of Groningen gave the following official terms of reference to the Noort Committee on Wednesday, 12 October 2011.

The Committee is to investigate the extent and nature of the infringement of scientific integrity committed

by Mr D.A. Stapel. There are three elements to the task.

1. The Committee should examine which publications are based on fictitious data or fictitious scientific studies, and during which period the misconduct took place.
2. The Committee should offer a view on the possible legal and other consequences of the infringement of scientific integrity, and make recommendations for corrective measures, on the University's part, or jointly by the universities involved.
3. The Committee should offer a view on the methods and the research culture that facilitated this infringement of scientific integrity, and together with the universities involved make recommendations as to how a recurrence may be prevented.

The members of the Noort Committee are Prof. E. Noort (chair, University of Groningen), emeritus professor of Old Testament, former dean and former member of the board of the Royal Netherlands Academy of Arts and Sciences (KNAW), and chair of the Committee on Scientific Integrity of the University of Groningen;

Prof. H.E. Bröring (University of Groningen), professor of Integrative Law Studies and member of the Committee on Scientific Integrity of the University of Groningen;

Prof. J.M. Pieters (University of Twente), professor of Applied Psychology at the University of Twente, former dean and former member of the board and chair of the Dutch Association of Psychologists.

The Committee's administrative secretary is Dr Merlijne Jaspers, (Department of Research & Valorisation, University of Groningen, social psychologist and health scientist).

### **1.3 Drenth Committee: Terms of Reference**

The Rector Magnificus of the University of Amsterdam requested the Drenth Committee on Wednesday 21 September 2011 to investigate the extent and nature of the infringement of scientific integrity committed by Mr D.A. Stapel. There are three elements to the task.

1. The Committee should examine which publications are based on fictitious data or fictitious research, and during which period the misconduct (if any) took place. The publications concerned are those which Mr Stapel produced during his Amsterdam period. The Committee will also investigate whether the work for which Mr Stapel's doctorate was awarded is also among these publications. If appropriate, the scope will also include the roles of the doctoral supervisor(s) and the doctoral examination committee.
2. The Committee should offer a view on the possible legal and other consequences, and make recommendations for corrective measures, on the University's part or jointly by the universities involved.
3. The Committee should offer a view on the methods and the research culture that facilitated this infringement of scientific integrity, and make recommendations as to how a recurrence may be prevented.

The members of the Drenth Committee are Prof. P.J.D. Drenth (chair, VU University Amsterdam), former Rector Magnificus of VU University Amsterdam and former president of the Royal Netherlands Academy of Arts and Sciences; Prof. J.W. Zwemmer (University of Amsterdam), Scientific Integrity Counsellor and former Rector Magnificus of the University of Amsterdam; Prof. L.A. de Klerk (University of Amsterdam), emeritus professor of Urban Planning and former dean and director at the University of Amsterdam; and Prof. C.A.J. Klaassen (University of Amsterdam), professor of Mathematical Statistics and former vice-dean and director at the University of Amsterdam.

The substance of the terms of reference of the three Committees is mutually consistent, albeit that legal aspects are explicitly included in the scope of the Noort Committee (Groningen) and the Drenth Committee (Amsterdam).

## 1.4 Statistical support

The Committees' work was supported by teams of statisticians. The Levelt Committee was supported by Dr M.A.L.M. van Assen; the members of the Noort Committee's team were Prof. R.H. Koning, Dr W.J. Post and Dr M.A.J. van Duijn, and the members of the Drenth Committee's team were Prof. H.L.J. van der Maas and Dr D. van Ravenzwaaij.

The contribution of these statisticians was indispensable in the study, re-analysis and evaluation of all available materials. There was, from the outset, continuous consultation between the statisticians and the Committees regarding their approach. There was also ample coordination between the three Committees and the three groups of statisticians. The statisticians reported to their respective Committees about anomalies in the publications concerned, such as suspicious irregularities and improbable findings, as came to light in the analysis of available data and other research material. They also estimated the extent of evidence of fraud. Needless to say, the responsibility for all conclusions stated in this report regarding the ultimate fraudulent nature of the publications and the further consequences in a broader sense reside fully with the Committee concerned.

## 1.5 Background to the inquiry

The fraud was brought to light by three young researchers in the Department of Social Psychology at Tilburg University. They reported their suspicions of data falsification by Mr Stapel at the end of August 2011 to the head of department, Prof. M. Zeelenberg. After months of observation they had gathered sufficient details to demonstrate that something was amiss. The researchers all deserve praise for reporting these suspicions. It is noted that they were in a dependent position and had much to lose. Following the report, the head of department immediately notified the suspicion of fraud to the Rector Magnificus of Tilburg University, who in turn spoke immediately with the whistleblowers. The Levelt Committee has learned that there had been other whistleblowers in Tilburg at an earlier stage. Three young researchers had previously sounded the alarm to senior staff of the faculty regarding irregularities in the datasets delivered by Mr Stapel. The risks they ran were equally severe. Furthermore, two professors had previously observed data that were 'too good to be true'. However, none of these earlier reports were acted upon. There were a

variety of reasons for inaction, as covered in Chapters 4 and 5.

## **1.6 Nature of the report**

This report is the joint final report of the Levelt, Noort and Drenth Committees. A complete list of Mr Stapel's publications and of all dissertations for which Mr Stapel served as supervisor or co-supervisor is presented. Whenever fraud has been established (i.e. beyond reasonable doubt) or evidence of fraud was found (see also Sections 2.3 and 2.4.2) this will be reported. Furthermore a description is given of the nature of the fraud and the research culture that allowed fraud to be perpetrated.

The Committees resolved to expand and adapt the interim report on the infringement of scientific integrity committed by Mr Stapel that was presented in October 2011, resulting in a single joint final report of the three Committees.

## 2 Working method of the Committees

### 2.1 Committees' collaboration and distribution of work

In view of their remit, each of the three Committees is independent and has its own responsibilities. In pursuing this remit, the Levelt, Noort and Drenth Committees have collaborated effectively, with Tilburg University's Levelt Committee being responsible for the overall coordination.

The doctoral dissertations served as the first point of departure: in each case, these were investigated by the Committee of the university where the doctoral defence had been conducted. The same applies to the articles that are part of a doctoral dissertation (although the article may have been somewhat revised relative to the chapter of the doctoral dissertation). The majority of the doctoral defences took place in Groningen (13). This is because Mr Stapel continued to act as supervisor or co-supervisor for PhD students in Groningen even after his departure to Tilburg, because of his earlier commitments. These 13 dissertations and the related articles (20) were therefore handled by the Noort Committee. In Tilburg, five doctoral dissertations were investigated by the Levelt Committee, along with nine related articles. In Amsterdam, Mr Stapel was not involved as a doctoral supervisor or co-supervisor.

With regard to the other publications, the Drenth Committee at the University of Amsterdam investigated the (32) publications that appeared in the period from 1993 to 1999, with special attention to the research on which Mr Stapel's dissertation was based. The Noort Committee at the University of Groningen investigated the 31 other publications dating from 2000 to 2006 and Tilburg University's Levelt Committee investigated the 45 other publications from 2007 to 2011.

The Committees did not consider ASPO (Dutch Association of Researchers in Social Psychology) or other proceedings to be publications.

### 2.2 Formal status of Committees

The *Levelt Committee* is a confidential committee established by the Rector Magnificus in accordance with Tilburg University's Regulations on Scientific Integrity (*Regeling Wetenschappelijke Integriteit*). The Committee's working method is governed by these Regulations. According to Article 4 of these Regulations, the case should be handled behind closed doors. This means that no named information may be shared with third parties unless under the auspices and within the remit of the Committee. However, the Committee is able to make anonymous use of the information provided by informants in order to corroborate its findings. The Committee's deliberations strictly adhered to the principle of hearing both sides of the case (i.e. *'audi et alteram partem'*). In each case, the Committee considers and assesses the position of the parties within the frameworks and standards formulated in the KNAW Memorandum on Scientific Integrity (*Notitie Wetenschappelijke Integriteit*).

With regard to the essential status components, the same applies to the Noort and Drenth Committees. The *Noort Committee* is a special ad hoc committee of the University of Groningen's Scientific Integrity Committee (abbreviated to CWI in Dutch). Two of its members are also members of the CWI and the third, external member is an expert from the University of Twente. The *Drenth Committee* is a special, ad hoc committee established by the Rector Magnificus of the University of Amsterdam.

## 2.3 Procedure

The Committees started by collecting all Mr Stapel's publications published in the period from 1993 to 2011, including all available datasets, questionnaires, stimulus material, hypotheses and e-mail correspondence. Studying these materials with the assistance of the statisticians accounted for the majority of the time that the Committees devoted to the enquiry.

In collecting all of the relevant material, the Noort Committee made use of a questionnaire (Appendix 2), designed for the co-authors involved. This questionnaire focused particularly on the way in which data was collected and processed for the particular article or chapter and the role played by the co-author and by Mr Stapel in the publication. The Levelt Committee collected the same type of material by means of targeted, specific written and oral questions to the co-authors. The Drenth Committee did not have access to this type of material in view of the dated nature of the experiments and publications.

Mr Stapel himself was requested to provide a list of publications in which fictitious data have been used. In September 2011, he provided an initial list of his publications from 1993 until mid-2011, indicating those that did or did not involve fictitious data, to the best of his recollection.

With regard to a second list, compiled by the Levelt Committee and primarily comprising chapters in books and articles to be published, Mr Stapel initially indicated that he was not in a fit state, physically or mentally, to indicate the possibly fraudulent nature of these publications. However, in December 2011, the Levelt Committee did receive a response from Mr Stapel about this second list, accompanied by the qualification that he cannot, in view of his health, give any guarantee that the list of fraudulent publications is complete. He also indicated that he was not completely certain about the state of affairs concerning some publications.

Mr Stapel was also asked for responses concerning the fraudulent nature of the dissertations he supervised and (again) about a number of publications that he had not initially designated as being fraudulent, but which the Committee concerned – in the light of the investigations – had concluded to be probably fraudulent. These publications, along with the dissertations, were sorted into three different categories by Mr Stapel: good, in doubt, or tainted.

The Committees held many interviews with whistle blowers, present and former PhD students, co-authors, colleagues and present and former members of the faculty boards in order to build up the clearest possible picture of the course of events. Where possible these interviews were conducted face-to-face, and otherwise by Skype or by telephone. A comprehensive list of the people interviewed by the three Committees can be found in Appendix 3. The Committees attempted to conduct these interviews with concerned parties as openly as possible in order to obtain optimally useful information. This applies both to those discussions in

which more general information was provided and to those which included responses to specific questions from the Committees. Reports or memoranda on all of these interviews were compiled by the secretary of the Committee concerned for internal use. In Groningen, tape recordings were made of the interviews conducted, after permission had been obtained.

Based on all the information available, the Committees drew a distinction between the following:

- i publications in which proof of fraud was established;
- ii publications in which proof of fraud could not be established.

With regard to the publications from the Amsterdam period, particularly those dating back the furthest, and to some of the Groningen publications (17 articles), the information available was such that the publications were assessed in terms of whether indications leading to evidence of fraud were found. This issue is examined in more detail in Section 2.4.2.

With regard to the announcement of the results of the enquiry, the Rectors, in their role as commissioners of the enquiry, and the Committees both agreed from the outset that the Committees' conclusions had to be made public: science must as far as possible be a public and transparent endeavour. To avoid misunderstanding, it should also be stated that Mr Stapel has never resisted publication of the conclusions.

After careful deliberation, the Committees decided to gradually publicize their conclusions regarding each publication on a public website (<https://www.commissielevelt.nl/>). This is the fastest and most comprehensive way to meet both the need for information that existed in the world of science and beyond, and – even more importantly – the need of the various co-authors, which was explicitly expressed to the Committees, to have access to the Committees' judgment as soon as possible. This outweighed the potential disadvantage of those involved being repeatedly confronted with the issue of the Stapel fraud. The timing of the web publications was generally such as to give dissertations precedence over other publications, and recent publications over older ones, and to respond rapidly where there were urgent ad hoc reasons to do so.

Prior to the presentation on the website of the status of the publications, the following procedure was applied:

1. The Committee concerned reaches a judgment concerning the publication.
2. On the basis of contacts with the co-authors, additional questions are generally asked and the judgment is discussed further. If the publication falls under category (i) (fraudulent), or, in the case of the Drenth Committee and for some of the publications, the Noort Committee, if there are indications of fraud, the fact that the judgment is to be posted on the website is communicated with relevant parties (in particular co-authors) in order to enable them to report any factual inaccuracies.
3. The Coordinating Committee informs Mr Stapel about the intention to post the judgments on the publications on the web, providing him with the opportunity to point out any factual inaccuracies.
4. Approximately once each month, the Committees report to their Rectors and Executive Boards on the publications that are considered fraudulent. The three Committees also share this information with each other.
5. On receiving these reports, the Executive Boards of the universities notify the relevant journals and grant-awarding bodies, either immediately or after the publication of the final report, on the conclusions concerning the publications.

6. The Levelt, Noort and Drenth Committees publish their findings jointly on the website, coordinated by the Levelt Committee. For every publication in which fraud has been identified, the essential information on which the conclusion is based is presented.

The Levelt Committee has compiled a comprehensive list of the publications investigated, along with a list of publications in which the Committee was able to establish fraud. The latter list is included below in Chapter 3 and in Appendix 4. This Appendix also includes a comprehensive list of publications investigated.

The Drenth Committee has expressed its judgments in a different way because in no case did it have at its disposal the original data used for the analyses and conclusions in the publications. It was also difficult to obtain reliable information from co-authors concerning these older publications. In this report, as on the website, the Drenth Committee draws a distinction between proven fraud ('beyond reasonable doubt') and strong indications of fraud ('evidence of fraud'). The Drenth Committee reports a comprehensive list of publications investigated and a list of the publications in which such evidence of fraud has been found. These lists can be found in Chapter 3 and in Appendix 6.

For some publications, the Noort Committee applies the same working method as the Levelt Committee. For some other publications, the Noort Committee applies the same working method as the Drenth Committee, for similar reasons. These lists can be found in Chapter 3 and in Appendix 5. This latter Appendix also includes a comprehensive list of publications investigated.

### **2.3.1 Information sources**

The Committees used some combination of the following sources of information in the investigation of fraud in the publications:

- a. *The texts of all the scientific publications authored or co-authored by Mr Stapel, including all the dissertations for which he was supervisor or co-supervisor.*

This list is complete to the best of the Committees' knowledge. Almost all publications report the results of new experimental empirical research. Some publications are only theoretical in nature and do not report on new experimental research. These theoretical publications were examined for 'derivative' fraud in terms of the extent to which the conclusions were based on known fraudulent publications. Books edited by Mr Stapel were not assessed by the Committees since they solely involved compilations of chapters by various authors. Of course, the Committees did investigate any chapters written by Mr Stapel himself if they included experiments.

- b. *The experimental data*

Great pains have been taken to obtain the original datasets, fraudulent or otherwise, on which the analysis and conclusions in the publications were based. These data were not available in all cases, especially for research conducted some considerable time ago. All of the co-authors cooperated with the Committees to find the data. In some cases, it proved impossible to recover the data, either because it was not available or was incomplete, or because it was in a form that did not allow careful analysis. The Levelt Committee assessed 46 articles on experimental research. Data were available for 38 of these (in two cases including the raw data in the form of completed questionnaires).

The Noort Committee assessed 45 articles on experimental research. Data were available for 29 articles. The Drenth Committee investigated 32 publications authored or co-authored by Mr Stapel, including

six articles that were part of his dissertation. No original data were available with respect to any of these publications.

*c. The research material*

This concerns the written documents that outline exactly how the experiments were set up and conducted, including the completed questionnaires and stimulus materials as well as predictions of the results of the experiments and e-mail correspondence between Mr Stapel and the researcher. This information was also obtained from the co-authors, but was certainly not available in all cases.

In 33 of the 46 articles on experimental research, the Levelt Committee had research material from at least one study in the article at its disposal.

In 12 of the 45 articles on experimental research, the Noort Committee had research material from at least one study in the article at its disposal.

As indicated above the Drenth Committee did not have the underlying research material at its disposal for any of the 32 investigated publications.

*d. The statements by Mr Stapel on the fraudulent nature of the data.*

*e. The statements by the co-authors on the analyses and the organization of the experiment, in a number of cases in response to specific questions from the Committees.*

*f. The statisticians' reports.*

### **2.3.2 All-embracing enquiry**

Although Mr Stapel's own admission and indication of fraudulent publications are sufficient grounds for declaring a publication to be fraudulent, the Committees nevertheless decided not only to engage in discussion with various parties involved but also to study all of the available materials, all publications and all experiments in these publications. The main reason for this is the fact that with respect to a number of publications, Mr Stapel indicated that he no longer knew exactly whether the publication was based on fictitious data or not and that he had issued a proviso indicating that he could, of course, be mistaken, especially in view of his physical and mental condition at that time. It is therefore possible that publications labelled by him as fraudulent are not or that publications he described as 'good' could actually be tainted. In principle, each separate article contains an independent contribution to science. It is therefore very important that the maximum possible certainty be obtained with regard to the potentially fraudulent nature of each separate article.

Another reason for the working method chosen by the Committees was based on their commission not only to identify which publications by Mr Stapel are fraudulent but also whether other people were directly or indirectly involved (the extent of the fraud).

It was also the task of the Committees to investigate the conditions and procedures that led to a fraud of this magnitude both in terms of numbers and duration (the research culture). This requires a thorough knowledge of the actual publication practice, which must be obtained by means including the study and re-analysis of all available material.

### **2.3.3 Conclusion per publication**

In assessing the fraudulent nature of an article, all of the information sources referred to above were used together and also linked to any indications that Mr Stapel may have been involved in the collection, analysis and reporting of the data. The statisticians' reports played an important role in this. The ultimate assessment was made by the Committees after studying the reports in consultation with the statisticians and after a check

by the co-authors, who were given the opportunity to respond to the findings. In some cases, the Committees' original judgment was modified on the basis of the information obtained from the co-authors. The Committees' ultimate conclusions were also presented to Mr Stapel in order to identify any factual inaccuracies. In no case did this lead to a modification of the final assessment. This was followed by publication of the conclusions and the key arguments on the website. On the web page and in Appendices 4, 5 and 6, the key considerations of the Committees in determining the fraudulent nature are presented in brief.

#### **2.3.4. Hearing both sides of the case: general**

All three Committees have strictly adhered to the principle of hearing both sides of the case (i.e., *audi et alteram partem*). This particularly applies to information concerned with fraud in the publications. At first, the Committees gave the co-authors the opportunity to respond to their questions, which were often prompted by the statisticians' reports. Then, where applicable, they also had an opportunity to point out any factual inaccuracies in the conclusions of the three Committees.

#### **2.3.5 Hearing both sides of the case: Mr Stapel**

Mr Stapel was given an opportunity to comment on the Committees' conclusions and draft reports. He was also invited for a personal interview with the Committees. On 14 October 2011, before the publication of the interim report, he indicated to the Levelt Committee that he was not in a fit state to attend such an interview. Accordingly, a draft of this interim report was sent to him and his lawyer for their comments.

On 15 May 2012, Mr Stapel was able to accept the Committee's invitation to attend a personal interview to present his side of the case.

On 9 October 2012, he also responded to a similar invitation from the Noort Committee for an extensive personal interview.

There was no personal interview with the Drenth Committee, although its draft conclusions were presented to Mr Stapel for his comments. Mr Stapel responded via his lawyer as follows:

'My client has already admitted to having committed fraud in much of his research. He has also conceded that he has been sloppy, kept inadequate records and presented data in the best light possible. As far as my client can recall, there was, however, no fraud during his period in Amsterdam.'

A draft version of this final report entitled 'Flawed Science' has also been presented to Mr Stapel with the request to identify any factual inaccuracies. Mr Stapel and his lawyer have responded in writing to the Committees. The Committees have corrected the factual inaccuracies he has brought to the Committees' attention.

## 2.4 Identifying fraud

### 2.4.1 *The definition of fraud*

What do the Committees understand by the term ‘fraud’? In studying all the available information (publications, data, research material, the content of interviews conducted, etc.), the Committees encountered numerous violations of proper scientific standards. They not only found frequent minor irregularities, but also identified serious infringements of scientific integrity on numerous occasions. These are examined in detail in Chapters 3, 4 and 5.

However, in the light of the Committees’ assignment, the definition of fraud should be limited to the fabrication, falsification or unjustified replenishment of data, as well as the whole or partial fabrication of analysis results. It also includes the misleading presentation of crucial points as far as the organization or nature of the experiment are concerned. The term ‘data’ refers to the coded raw scores as they occur in the data matrix in which the scores, fictitious or otherwise, are recorded for each research subject for all variables.

The Committees classify a publication (empirical or otherwise) as fraudulent if, in at least one of the experimental studies in the publication, fraud has been proven beyond reasonable doubt.

The other breaches of the scientific rules are considered as, possibly serious, forms of bad science, but are not labelled ‘fraudulent’. The discussion of the research culture (Chapter 5) covers many variants of such, often serious, breaches of the rules of scientific research in all research phases.

### 2.4.2 *Proof and evidence of fraud*

In order to identify fraud in a specific publication, a combination of different indicators of fraud have been used. Some of these indicators are less compelling than others, but may constitute additional evidence for establishing fraud when combined with other observations. The website and this final report (Appendices 4, 5 and 6) present the most important grounds for the identification of fraud in a publication.

The Committees refer to proof of fraud for a publication only if it is beyond reasonable doubt.

Since the Drenth Committee had no original data or other research material at its disposal, this Committee’s analysis had to focus exclusively on suspected irregularities and statistically highly implausible results. The analysis by the Drenth Committee resulted in a judgment on the probability of fraud in each article, on the following scale: not applicable – none – negligible – slight – relatively strong – strong. The Drenth Committee has taken these last two categories as sufficient grounds for its judgment ‘evidence of fraud’. The Drenth Committee points out that this implies a high degree of probability from a statistical perspective, but that, given the unavailability of original data and Mr Stapel’s denial of fraudulent behaviour in this period, this does not constitute legal proof of fraud.

The Noort Committee likewise had no data or other research material at its disposal for a number of articles (17). For these articles, the Noort Committee adopted the same working method as the Drenth Committee, although it only classified ‘strong’ signs of fraud as ‘evidence of fraud’.

## 2.5 Concrete working method: explanation and examples

This section describes how the Committees reached their conclusions in concrete terms.

Where available, the datasets provided were investigated for any peculiarities and suspicious results. The datasets were also re-analysed and the results compared with those reported in the publication. The questionnaires and stimulus material provided were compared with the dataset provided and with the way they were reported in the publication. The publication itself was also examined for irregularities. Observations made by the co-authors with respect to the publications were also further assessed. Some of these observations were given as a response to questions from the Committee concerned. Finally, Mr Stapel's own judgment on the publication was taken into account.

A more precise picture of the working method of the Committees can be obtained by examining how the steps and working methods described in general terms above were actually applied in practice. In order to ensure a clear exposition of the variety of methods and approaches used, they are divided below in terms of the most important sources of information to which they referred. First, there is the *research material*, which is the information concerned with the design of the experiment and how it was actually carried out. The focus of the Committees' investigation here was the question as to what extent the information contained in this material corresponds with that found in the data (if available) and in the publication. Next, the *data* itself were examined for peculiarities. This inspection of the data matrix was often applied when there were other reasons to suspect fraud. In addition, *data analyses and re-analyses* were carried out, both to ascertain whether the published results could be replicated and to identify any peculiarities and inaccuracies in the analyses and datasets. Finally, there the investigations focussed on the *publication* itself, sometimes only on what was included in the article or chapter if other information was lacking and sometimes in relation to the research material and the data. In many cases, the Levelt Committee was able to use all three sources, the Noort Committee mainly used the latter two sources, and the Drenth Committee primarily based its work on the third source.

It is not feasible to provide a comprehensive overview of all the specific concrete methods applied by the Committees. However, the general working procedure and most important variations in this are reported below, partly by describing concrete examples of results that were obtained in this way.

### 2.5.1 Organization and conduct of the experiment

An important and often first step in the investigative work of the Committees was to ascertain how the experiment was organized and conducted. This involved the question of whether the information on the organization and execution of the experiment was consistent in the various sources of information (the research material, the article, the data and the interviews). For example, there was a check on whether the experimental conditions, the sequence of these conditions and the stimulus material were correctly reported in the article; in other words whether they matched the research material and the available data. There was also a check of whether the datasets included all of the variables within the stimulus material and whether all variables in the datasets also appeared in the stimulus material. During this process, frequent discrepancies were encountered.

Throughout, an examination was made, often by means of specific written or verbal questions to the co-authors, as to whether the discrepancies and peculiarities could be explained by other means than the manipulation of data. In most cases, there were either satisfactory explanations for these discrepancies, or the latter could be attributed to sloppiness or poor reporting. Otherwise, the absence of a satisfactory explanation was taken to infer strongly that data had been fabricated.

With regard to the organization of the experiments, particular attention was paid to research conducted at schools. According to Mr Stapel's statements, much of the research carried out within schools was fraudulent, or simply never took place. For certain experiments, a number of practical problems also made it difficult to accept that these experiments could actually have been conducted at the schools. This was also true for the experiments at stations and in trains or among judges. An especially critical approach was also followed in the case of experiments in which Mr Stapel, while in Tilburg, told Tilburg PhD students or co-authors that these could be taken along with other ongoing experiments in Groningen, and similarly students or co-authors in Groningen that they could join experiments in Tilburg. Needless to say, if it could be ascertained that the research had never taken place, this constituted proof of fraud. In many other cases, however, the analyses still led to suspicion of fraud, and investigation continued.

### ***2.5.2 Peculiarities in the data matrix***

For the determination of possible fraud in a publication, the original datasets were an important source of information. They enabled the Committees to compare the data with the research material and the information in the publications in order to identify discrepancies in the design and elaboration of the study (Section 2.5.1). The datasets provided could also be re-analysed in order to check the results reported in the publications (Section 2.5.3). Above all, inspection of the data matrices can reveal patterns that may at least raise suspicion of fraud. The fact that data are copied from one experiment to another is an important proof that fraud has been committed. Where there was reason for doubt, the scores or score distributions in the various datasets were scrutinized for improbable similarities, which would indicate copying. In general, copying is not immediately visible in the original datasets, but can be revealed by re-sorting them from respondent-number sequence into, for example, the sequence of scores on the dependent variable. For some experiments, simulation studies based on observed frequency distributions and correlations were used to determine the probability or otherwise of the number of identical observations.

Special attention was also paid to cases where there was no missing data reported and to the variation of the scores of the experimental subjects, within and between the research conditions (experimental and control conditions). If the measurement of the variables in an experiment takes place by means of questionnaires to be completed manually by the experimental subjects, missing data in one form or other are virtually unavoidable. The total absence of such scores therefore raises a suspicion of fraudulent data manipulation, as do results such as: absolutely all persons in the experimental group achieve the highest scores (e.g. 6 and 7), while all persons in the control group achieve the lowest scores (e.g. 1 and 2). Results of this kind are highly implausible and therefore justify further investigation.

### ***2.5.3 Analysis and re-analysis of data***

If available, the original data were used in order to check the most important analysis results in the publications, such as the reported values of the means and standard deviations, of reliabilities, test sizes, effect sizes and p-values. Particular attention was paid to discrepancies between the results of the re-analyses and the published findings that could have an impact on the conclusions in the publications.

Throughout, often by means of contact with the co-authors, possible explanations for these discrepancies other than data fraud were explored. For example, the selective omission of exceptional scores or the scores of particular research subjects sometimes occur. There may be good reasons for this, but then they must be clarified and reported.

These re-analyses not only proved useful in detecting potential fraud, they also provided the Committees with an effective insight into the research culture and the actual research practices behind the publications.

In addition to these re-analyses to check the published outcomes, supplementary analyses were often carried out for various reasons. For example, standard deviations of various variables were calculated, as well as correlations between different items of the same scale and correlations within each experimental condition. Unreported reliabilities were also calculated and effect sizes were compared across the different studies and conditions. Some results of these analyses, such as unusually large or small standard deviations, justified a more precise examination of the response patterns on the variables in question in the data matrix (as in Section 2.5.2).

Many of these additional analyses aimed at assessing the validity of the published results. These are discussed in Section 2.5.4.

#### **2.5.4 Publications**

The information in the publications themselves can also be used to demonstrate fraudulent practice. Almost every publication includes information about the organization of the experiments. This may, for example, show that a specific experiment in the described conditions could not have been carried out in that way for practical, logistic reasons. This can provide a strong indication of fraud, a supposition that can then be either weakened or confirmed by information from the researchers involved or by the information from the research material and data.

The analysis results reported in the publications can also suggest that the basic rules of proper scientific research in general have been violated, or even point to data fraud. In general terms, there are two types of data manipulation that can lead to suspect research results. The first of these involves working with genuine data, but with the selective and unreported omission of unwelcome scores for variables, the exclusion of certain exceptional subjects or of exceptional research results. The term ‘exceptional’ refers to the fact that the information omitted is not in line with what the researcher would like to see as a result. The analysis results for these types of data manipulations generally deviate from the results obtained without manipulation.

Secondly, the complete or partial ‘naive’ fabrication of data leads to results that never or seldom occur in genuine data. If, as in the case of Mr Stapel, the data fabrication occurs simply by completing or by amplifying the data matrix oneself ‘manually’ in order to obtain the desired result, the almost certain result will be data and distributions that deviate from realistic ones (in contrast to the scenario in which more statistically advanced methods of data fabrication are used).

These deviations can become particularly manifest in too low chance fluctuations, too large effects, strange multivariate correlations, mutual dependence of observations and other implausible analysis results. These points are elaborated in greater detail below, because instances of all of these consequences were found in Mr Stapel’s fraudulent publications.

## ***Chance variations that are too small***

Data from two independent studies of the same situation will tend to differ because of chance fluctuations. People generally underestimate the role played by chance. When completing the data matrix themselves, fraudulent researchers will have a tendency to make the results of two supposedly identical studies resemble each other rather too much. This lack of chance variation may also be caused by the systematic deletion of ‘exceptional’ results.

Mr Stapel’s fraudulent publications include many independent replications of essentially the same experimental (i.e. control) conditions. If the reported mean values of the dependent variables (and/or of their standard deviations) for all of the replicated conditions are exactly the same (even to 4 decimal places, as discovered in one article), it may be concluded that this is a case of fraudulent data copying. Even if the means are not exactly the same, but approximately the same, the results can suggest fraudulent manipulation. Using a statistical model, it is possible to determine how probable it is that such small differences would occur in independent replications of the same measurements, assuming that the ‘true’ values are the same. If this probability is very small (for example  $10^{-10}$ ), this is a clear sign of data manipulation.

## ***Effects and relationships that are too large***

The omission of unwelcome results can also lead to effects that are too large, in the sense that they are ‘larger than one would generally encounter in realistic, non-manipulated data’. Likewise, in the case of the naive fabrication of the data matrix by the researcher, there is a noticeable tendency to assume unrealistically strong relationships between variables and to create unrealistically large effects of the experimental manipulations. This occurred frequently in the case of Mr Stapel. For example, there are very large mean differences on the dependent variable between experimental and control groups with hardly any variations in scores within the groups. ‘Large effects’, reported in the article or as calculated by statisticians can therefore constitute an important sign of data fabrication.

It is of course necessary to apply a standard to the observation that something is ‘too large’: when is a correlation or effect size too high so that it leads to suspicion of data manipulation? Published meta-analyses show that the mean effect size in psychological research in which two groups or conditions are compared with each other, as measured with Cohen’s *d*, is approximately 0.5 (Meyer *et al.*, 2001; Richard, Bond, & Stokes-Zoota, 2003). In Cohen’s terminology, this corresponds to a medium effect.

Students from the Research Master’s programme at Tilburg University, supervised by the statistician from the Levelt Committee, investigated in late 2011 what the mean and variation of the effect sizes were in some important international journals in the field of experimental social psychology to which Mr Stapel was a frequent contributor. This was done to provide a comparison with publications authored or co-authored by Mr. Stapel. The study included an assessment of the *European Journal of Social Psychology* (EJSP), issues 1 to 3 from 2011 and the *Journal of Personality and Social Psychology* (JPSP), issues 100 (1) to 100 (3) from 2011, and issues 96 (1) from 2009, 94 (4) from 2008, and 93 (5) from 2007. For each study, Cohen’s *d* and the total explained variance of the dependent variable ( $\eta^2$ ), if available, were recorded as far as these were concerned with ‘between-subjects designs’: the design Mr Stapel used most often in his experiments.

The mean reported Cohen's  $d$  in these issues was 0.69 ( $n = 158$ ), slightly higher than in the meta-analyses referred to above. In 71% of cases, Cohen's  $d$  was lower than 0.8, which corresponds to a large effect. The mean total explained variance was equal to 22%, and was lower than 35% in 80% of the cases.

The relationships identified in various publications by Mr Stapel, with 85% or even 95% explained variance, therefore appear to be extremely rare; indeed, even 55% explained variance does not occur often. Such high values therefore call for further analysis and may indicate fraud.

The size of the effect of the experimental conditions on the dependent variable can also be evaluated in relation to the reliability of the measurement of the dependent variable. The size of the effect can never be stronger than the reliability of the dependent variable (given the usual assumptions of independent observations and measurement errors). If this is nevertheless the case, it may indicate fraud. Corrections for unreliability can also lead to improbably large effect sizes (for example 90% explained variance or more). This occurred regularly in Mr Stapel's publications. Implausibly high effect sizes and very strong relationships provided reasons for investigating whether they correspond to specific peculiarities in the data matrix.

### ***Unusual multivariate relationships***

Data fabrication, like that perpetrated by Mr Stapel, can also lead to strange multivariate results. All bivariate correlations can appear to be normal (and in the right direction), but peculiar results can occur in a multivariate analysis, which may not be directly visible but which are seldom encountered in real data. For example, a fraudster may give scores to the independent variable  $X$  and to the dependent variable  $Y$  and do so in such a way that results in  $X$  have a strong positive effect on  $Y$ . The researcher will then fabricate the scores for the covariate  $Z$  in such a way that a positive correlation arises between  $Y$  and  $Z$ . On the surface, the bivariate correlations  $X$ - $Y$  and  $Y$ - $Z$  appear fairly normal, but that does not necessarily apply to the bivariate relationship  $X$ - $Z$  or to the partial relation between  $Y$  and  $Z$ , controlled for  $X$ : in other words within the  $X$  categories.

The bivariate fabrication of data can also lead to strange results in manipulation checks. These checks investigate whether the experimental manipulation  $X$ , for example finding oneself in a situation of anxiety or not, actually leads to a higher or lower score on the anxiety variable  $Z$ . The effect of the experimental manipulation  $X$  on the dependent variable  $Y$  is then considered to run via the intervening variable  $Z$ . Theoretically, the effect of  $X$  on  $Y$  in a standard regression analysis is equal to the effect of the manipulation on the intervening variable times the direct effect of the intervening variable on the dependent variable. The manipulation check investigates whether the experimental condition actually correlates in the expected way with the scores on the intervening variable. Theoretically, the direct effect of the independent variable  $X$  on the dependent variable  $Y$  should be small, or even zero, if the intervening variable  $Z$  is kept constant. When data are fabricated pairwise, each time for each pair of two variables, multivariate relationships are not taken into account and the direct effect of  $X$  on  $Y$  may remain large, even if the effect of the intervening variables  $Z$  is controlled for. When such instances were encountered in Mr Stapel's publications, this was a reason for further investigation, especially if the effect of  $X$  on  $Y$  was larger than the effect of  $X$  on  $Z$ .

## ***Dependent observations***

In experimental research, the data from different subjects are independent of each other. In other words, knowledge of the scores on the pertinent variables of any randomly chosen experimental subject predicts nothing about the scores that any other randomly chosen subject may have obtained. In fabricated data, this is difficult to achieve because people are not good at assigning chance fluctuations. A statistical model may be used to calculate the probability of whether the reported (variance analysis) results originate from independent observations or from dependent observations and therefore from fabricated or manipulated research results. Although these kinds of statistical models are based on assumptions, they are relatively robust in terms of violations of these assumptions.

## ***Other improbable analysis results***

Even without the presence of datasets, it is possible to derive suspect data patterns from published results. For example, it was possible to use the reported value of the standard deviation of a specific variable, given the reported value of the pertinent mean, to conclude that all scores for that variable were, without exception, extreme scores at the end points of the scale. That is an unusual result that suggests fraud.

It was also sometimes possible to determine that the probability of a particular series of test outcomes is virtually impossible without data manipulation. For example, for 24 F-tests, each used to test the hypothesis that there was no difference between the experimental and control group, it was reported that the hypothesis was confirmed all 24 times with  $F < 1$ . Using the data from the article itself it is possible to calculate the probability of the occurrence of such a series of low F-values. If this probability, as in this case, is ‘impossibly’ small, there is hardly any explanation other than selective data manipulation.

It is also possible to determine whether there are implausible or impossible values for the means and standard deviations, even in a publication for which no data available. Incorrectly reported degrees of freedom (in relation to the number of respondents) and p-values that do not correspond with the values of the test statistics are also indications of irregularities.

The ultimate determination of fraud and data manipulation within an article or a concrete experiment was, as far as possible, based on a combination of the above criteria and concrete techniques.



# 3 Findings regarding the nature and extent of the fraud

## 3.1 Extent of the fraud

The Committees have arrived at the conclusion that the extent of Mr Stapel's fraud is very substantial. The Committees encountered a total of fifty-five publications in which fraud has been established.

### **Levelt Committee**

The Levelt Committee established fraud in thirty-four publications, of which, according to the acknowledgements, thirteen were funded through an NWO Pioneer grant granted to Mr Stapel, and one through a Transforum research grant. One or more chapters in three dissertations supervised by Mr Stapel were also found to be based on fictitious data. A full list of the investigated publications, any grants awarded to Mr Stapel, and a concise account of the salient grounds for fraud are given in Appendix 4 (see also the website [www.commissielevelt.nl](http://www.commissielevelt.nl)).

The Levelt Committee established fraud in the following publications:

- Avramova, Y.R., Stapel, D.A. & Lerouge, D. (2010). The influence of mood on attribution. *Personality and Social Psychology Bulletin*, 36, 1360-1371.
- Avramova, Y.R., Stapel, D.A. & Lerouge, D. (2010). Mood and context-dependence: Positive mood increases and negative mood decreases the effect of context and perception. *Journal of Personality and Social Psychology*, 99, 203-214.
- Avramova, Y.R., Stapel, D.A. (2008). Moods as spotlights: The influence of moods on accessibility effects. *Journal of Personality and Social Psychology*, 95, 542-554. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Blanton, H., & Stapel, D.A. (2008). Unconscious and spontaneous and ... complex: The three selves model of social comparison assimilation and contrast. *Journal of Personality and Social Psychology*, 94, 1018-1032.
- Häfner, M., & Stapel, D.A. (2010). Information to go: Fluency enhances the usability of primed information. *Journal of Experimental Social Psychology*, 46, 73-84. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Johnson, C.S., Norton, M.I., Nelson, L., Stapel, D. A., & Chartrand, T.L. (2008). The downside of feeling better: Self-regard repair harms performance. *Self and Identity*, 7, 262-277.
- Johnson, C.S. & Stapel, D.A. (2007). No pain, no gain: The conditions under which upward comparisons lead to better performance. *Journal of Personality and Social Psychology*, 92, 1051-1067. (Research supported by a Pioneer grant from NWO, and a research grant from the Heymans Institute of the University of Groningen both awarded to D.A. Stapel).
- Johnson, C.S., & Stapel, D.A. (2007). When different is better: Performance following upward comparison. *European Journal of Social Psychology*, 37, 258-275. (Research supported by a Pioneer

grant from NWO, and a research grant from the Heymans Institute of the University of Groningen both awarded to D.A. Stapel).

- Johnson, C.S. & Stapel, D.A. (2010). It depends on how you look at it: Being versus becoming mindsets determine responses to social comparisons. *British Journal of Social Psychology*, *49*, 703-723.
- Johnson, C.S., & Stapel, D.A. (2011). Happiness as alchemy: Positive mood and responses to social comparisons. *Motivation and Emotion*, *35*, 165-180.
- Lammers, J., & Stapel, D.A. (2011) Racist biases in legal decisions are reduced by a justice focus. *European Journal of Social Psychology*, *41*, 375-387.
- Meijers, M.H.C., & Stapel, D.A. (2011). Me tomorrow, the others later: How perspective fit increases sustainable behavior. *Journal of Environmental Psychology*, *31*, 14-20. (Research supported by Transforum Grant WP-088 awarded to D.A. Stapel).
- Nelissen, R.M.A., Van de Ven, N., & Stapel, D.A. (2011). Status concerns and financial debts in adolescents. *Social Influence*, *6*, 39-56.
- Noordewier, M.K., Van Horen, F., Ruys, K.I., & Stapel, D.A. (2010). What's in a name? 361.708 Euros: The effects of marital name change. *Basic and Applied Social Psychology*, *32*, 17-25.
- Noordewier, M.K., & Stapel, D.A. (2009). Judging the unexpected: Disconfirmation of situation-specific expectancies. *European Journal of Social Psychology*, *39*, 944-956. (Research supported by a 'Tiber' grant from the Departments of Economics and Business and Social and Behavioral Sciences of Tilburg University awarded to D.A. Stapel).
- Noordewier, M.K., & Stapel, D.A. (2010). Affects of the unexpected: When inconsistency feels good (or bad). *Personality and Social Psychology Bulletin*, *36*, 642-654.
- Noordewier, M.K., & Stapel, D.A. (2011). When failure feels better than success: Self-salience, self-consistency, and effect. *British Journal of Social Psychology*, *50*, 321-330.
- Noordewier, M.K., & Stapel, D.A. (2011). Stereotype disconfirmation affect: When sweet hooligans make you happy and honest salesmen make you sad. *Basic and Applied Social Psychology*, *33*, 1-6.
- Ruys, K.I., & Stapel, D.A. (2008). The secret life of emotions. *Psychological Science*, *19*, 385-391. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Ruys, K.I., & Stapel, D.A. (2008). How to heat up from the cold: Examining the preconditions for (unconscious) mood effects. *Journal of Personality and Social Psychology*, *94*, 777-791. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Ruys, K.I., & Stapel, D.A. (2008). Emotion elicitor or emotion messenger?: Subliminal priming reveals two faces of facial expressions. *Psychological Science*, *19*, 593-600. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Ruys, K.I., & Stapel, D.A. (2009). The unconscious unfolding of emotions. *European Review of Social Psychology*, *20*, 232-271. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Ruys, K.I., Stapel, D.A. & Aarts, H. (2011). From (unconscious) perception to emotion: A global-to-specific unfolding view of emotional responding. In I. Nyklicek, A.J.J.M. Vingerhoets, & M. Zeelenberg (Eds.), *Emotion Regulation and Well-Being* (pp.49-66). New York: Springer.
- Stapel, D.A., (2007). In the mind of the beholder: The interpretation comparison model of accessibility effects. In D.A. Stapel & J. Suls (Eds.) *Assimilation and Contrast in Social Psychology* (pp. 313-327). New York, NY, USA: Psychology Press. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Stapel, D.A. (2011). Priming as proxy: Understanding the subjectivity of social life. In K.C. Klauer, A.

Voss and C. Stahl (Eds.), *Cognitive Methods in Social Psychology* (pp. 148-183). New York Guilford Publications.

- Stapel, D.A., & Johnson, C.S. (2007). When nothing compares to me: How defensive motivations and similarity shape social comparison effects. *European Journal of Social Psychology*, 37, 824-838. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Stapel, D.A., & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science*, 332, 251-253.
- Stapel, D.A., & Marx, D.M. (2007). Distinctiveness is key: How different types of self-other similarity moderate social comparison effects. *Personality and Social Psychology Bulletin*, 33, 439-448. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Stapel, D.A., & Marx, D.M. (2007). Making sense of war: Using the Interpretation Comparison Model to understand the Iraq conflict. *European Journal of Social Psychology*, 37, 401-420. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Stapel, D.A., & Noordewier, M.K. (2011). The mental roots of system justification: System threat, need for structure, and stereotyping. *Social Cognition*, 29, 238-254.
- Stapel, D.A., & Noordewier, M.K. (2011). Stop making sense: The ultimate fear.
- Stapel, D.A., & Semin, G.R. (2007). The magic spell of language: Linguistic categories and their perceptual consequences. *Journal of Personality and Social Psychology*, 93, 34-48. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from 'Tiber', awarded to D.A. Stapel).
- Stapel, D.A. & Van der Linde, L.A.J.G. (2011). What drives self-affirmation effects?: On the importance of differentiating value affirmation and attribute affirmation. *Journal of Personality and Social Psychology*, 101, 34-45.
- Van Doorn, J. & Stapel D.A. (2011) When and How Beauty Sells: Priming, Conditioning, and Persuasion Processes, *Journal of Consumer Research*, published online June 1, 2011.

The Levelt Committee established data fraud in one or more chapters of the following dissertations:

- Avramova, Y.R. (2010), *How the mind moods* (Chapters 2, 3, 4 and 5)
- Noordewier, M.K. (2009), *Consistency and the unexpected* (Chapters 2, 3, 4 and 5)
- Van den Broek, L.M. (2009), *De ironie van gelijkheid* (Chapter 7)

## Noort Committee

The Noort Committee established fraud in twenty-one publications and seven dissertations. There was also evidence of fraud in another three publications. Of the above, according to the acknowledgements, eighteen publications were funded through an NWO Pioneer grant that was awarded to Mr Stapel, three publications through NWO subsidy T32 MH19728 that was awarded to Mr Stapel and two publications through his fellowship of the Dutch Royal Academy of Sciences (KNAW). Furthermore one or more chapters in seven dissertations supervised by Mr Stapel were found to have been based on fictitious data. A complete list of the investigated publications, any grants awarded to Mr Stapel, and a concise account of the salient grounds for fraud are given in Appendix 5.

- Gordijn, E.H., & Stapel, D.A. (2006). Behavioural effects of automatic interpersonal versus intergroup social comparison. *British Journal of Social Psychology*, *45*, 717-729.
- Joly, J.F., & Stapel, D.A. (2009). Staff, miter, book, share: How attributes of Saint Nicholas induce normative behavior. *European Journal of Social Psychology*, *39*, 145-153. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a “Breedtestrategie II” grant from the University of Groningen awarded to D.A. Stapel).
- Joly, J.F., Stapel, D.A., & Lindenberg, S.M. (2008). Silence and table manners: When environments activate norms. *Personality and Social Psychology Bulletin*, *34*, 1047-1056. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a “Breedtestrategie II” grant from the University of Groningen awarded to D.A. Stapel and S.M. Lindenberg).
- Lindenberg, S.M., Joly, J.F., & Stapel, D.A. (2011). The norm-activating power of celebrity: The dynamics of success and influence. *Social Psychology Quarterly*, *74*, 98-120. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Maringer, M., & Stapel, D.A. (2007). Unfinished business: How completeness affects the impact of emotional states and emotion concepts on social judgments. *Journal of Experimental Social Psychology*, *43*, 712-718. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel).
- Maringer, M., & Stapel, D.A. (2009). Correction or comparison? The effects of prime awareness on social judgments. *European Journal of Social Psychology*, *39*, 719-733. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).
- Marx, D.M., & Stapel, D.A. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology*, *91*, 243-254. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel).
- Marx, D.M., & Stapel, D.A. (2006). It's all in the timing: Measuring emotional reactions to stereotype threat before and after taking a test. *European Journal of Social Psychology*, *36*, 687-698. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel).
- Renkema, L.J., Stapel, D.A., Maringer, M., & Van Yperen, N.W. (2008). Terror management and stereotyping: Why do people stereotype when mortality is salient? *Personality and Social Psychology Bulletin*, *34*, 553-564.
- Schwinghammer, S.A., & Stapel, D.A. (2006). The effects of different types of self-activation on social comparison orientation. *Social Cognition*, *24*, 703-722. (Research supported by grant 007413212 and

by a Pioneer grant from NWO awarded to D.A. Stapel).

- Schwinghammer, S.A., & Stapel, D.A. (2011). Measure by measure: When implicit and explicit social comparison effects differ. *Self and Identity*, 10, 166-173. (Research supported by grant 007413212 and by a Pioneer grant from NWO awarded to D.A. Stapel).
- Stapel, D.A., & Blanton, H. (2004). From seeing to being: Subliminal social comparisons affect implicit and explicit self-evaluations. *Journal of Personality and Social Psychology*, 87, 468-481. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel).
- Stapel, D.A., & Marx, D.M. (2006). Hardly thinking about others: On target closeness and cognitive busyness in social comparison effects. *Journal of Experimental Social Psychology*, 42, 397-405. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel).
- Stapel, D.A., & Suls, J. (2004). Method matters: Effects of implicit versus explicit social comparisons on activation, behavior, and self-views. *Journal of Personality and Social Psychology*, 87, 860-875. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel).
- Stapel, D.A., & Schwinghammer, S.A. (in press). Me and the why of social comparison: how self-activation determines social comparison motives. *Self and Identity*.
- Stapel, D.A., & Van der Zee, K.I. (2006). The self salience model of other-to-self effects: Integrating self-enhancement, complementarity, and imitation principles. *Journal of Personality and Social Psychology*, 90, 258-271. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel).
- Trampe, D., Stapel, D.A., & Siero, F.W. (2007). On models and vases: Body dissatisfaction and proneness to social comparison effects. *Journal of Personality and Social Psychology*, 92, 106-118. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and research supported by grant number T32 MH19728 from NWO awarded to D.A. Stapel and F.W. Siero).
- Trampe, D., Stapel, D.A., Siero, F.W. (2011). The Self-Activation Effect of Advertisements: Ads can Affect Whether and How Consumers Think about the Self. *The Journal of Consumer Research*, 37, 1030-1045. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and research supported by grant number T32 MH19728 from NWO awarded to D.A. Stapel and F.W. Siero).
- Trampe, D., Stapel, D.A., Siero, F.W., & Mulder, M. (2010). Beauty as a tool: The effect of model attractiveness, product relevance, and elaboration likelihood on advertising effectiveness. *Psychology & Marketing*, 27, 1101-1121. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and research supported by grant number T32 MH19728 from NWO awarded to D.A. Stapel and F.W. Siero).
- Van den Bos, A., & Stapel, D.A. (2008). The impact of comprehension versus self-enhancement goals on group perception. *Social Psychology*, 39, 222-230.
- Van den Bos, A., & Stapel, D.A. (2009). Why people stereotype affects how they stereotype: The differential influence of comprehension goals and self-enhancement goals on stereotyping. *Personality and Social Psychology Bulletin*, 35, 101-113. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from 'Tiber', awarded to D.A. Stapel).

The Noort Committee established data fraud in one or more chapters of the following dissertations.

- Grob, J.D.M. (2009), *Dial E for Emotion : context and consequences of emotion regulation* (Chapter 5).
- Joly, J.F. (2008), *People on Our Minds: When Humanized Contexts Activate Social Norms* (Chapters 2, 3 and 5).
- Maringer, M. (2007), *Feeling one thing, seeing another: Emotion comparison effects in person judgments* (Chapters 2, 3, 4 and 5 ).
- Renkema, L.J. (2009), *Facing death together. Understanding the consequences of mortality threats* (Chapter 5).
- Schwinghammer, S.A. (2006), *The self in social comparison* (Chapters 2, 3 and 5).
- Trampe, D. (2007), *Social Influence: Comparison, Construal, and Persuasion Processes* (Chapters 2, 3 and 4).
- Van den Bos, A. (2008), *Why we stereotype influences how we stereotype*, (Chapters 2, 3 and 4).

The Noort Committee concluded from the statistical analyses of Mr Stapel's publications and in accordance with the decision rules given in Section 2.4.2 that there is evidence of fraud in the following articles.

- Stapel, D.A., & Koomen, W. (2001). When we wonder what it all means: Interpretation goals facilitate accessibility and stereotyping effects. *Personality and Social Psychology Bulletin*, 27, 915-929. (Research supported by a fellowship from the Royal Netherlands Academy of Arts and Sciences awarded to D.A. Stapel).
- Stapel, D.A., Koomen, W., & Ruys, K. (2002). The effects of diffuse and distinct affect. *Journal of Personality and Social Psychology*, 83, 60-74. (Research supported by a fellowship from the Royal Netherlands Academy of Arts and Sciences awarded to D.A. Stapel).
- Stapel, D.A., & Koomen, W. (2006). The flexible unconscious: Investigating the judgmental impact of varieties of unaware perception. *Journal of Experimental Social Psychology*, 42, 112-119. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel).

The following chapters in books and journal articles were based in full or in part on findings from articles in which the Committees had found evidence of fraud.

- Stapel, D.A. (2000). Moving from fads and fashions to integration: Illustrations from knowledge accessibility research. *European Bulletin of Social Psychology*, 12, 4-27.
- Stapel, D.A. (2002). Wat je ziet ben je zelf: Contexteffecten op oordelen en gedrag. *Nederlandsch tijdschrift voor psychologie*, 57, 31-41.
- Stapel, D.A., & Koomen, W. (2001). Let's not forget the past when we go to the future: On our knowledge of knowledge accessibility effects. In G.B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 229-246). Mahwah NJ: Lawrence Erlbaum.
- Stapel, D.A. (2003). Making sense of hot cognition: Why and when description influences our feelings and judgments. In J.P. Forgas, K.D. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 227-250). Cambridge: Cambridge University Press. (Sydney Symposium of Social Psychology series, 5).

## Drenth Committee

The Drenth Committee concluded from the statistical analyses of Mr Stapel's publications and in accordance with the decision rules given in Section 2.4.2 that there is evidence of fraud in the following articles.

- Stapel, D.A., & Koomen, W. (1996). Differential consequences of trait inferences: A direct test of the trait-referent hypothesis. *European Journal of Social Psychology*, 26, 827–837.
- Stapel, D.A., & Koomen, W. (1997). Using primed exemplars during impression formation: Interpretation or comparison? *European Journal of Social Psychology*, 27, 357–367.
- Stapel, D.A., & Koomen, W. (1998). Interpretation versus reference framing: Assimilation and contrast effects in the organizational domain. *Organizational behavior and human decision making processes*, 76, 132–148.
- Stapel, D.A., & Koomen, W. (1999). Correction processes in person judgments: The role of timing. *European Journal of Social Psychology*, 29, 131–138.
- Stapel, D.A., Koomen, W., & van der Pligt, J. (1996). The referents of trait inferences: The impact of trait concepts versus actor–trait links on subsequent judgments. *Journal of Personality and Social Psychology*, 70, 437–450. Dissertation, Chapter 3.
- Stapel, D.A., & Schwarz, N. (1998). Similarities and differences between the impact of traits and expectancies: What matters is whether the target stimulus is ambiguous or mixed. *Journal of Experimental Social Psychology*, 34, 227–245. Dissertation, Chapter 5.
- Stapel, D.A., & Spears, R. (1996). Event accessibility and context effects in causal inference: Judgment of a different order. *Personality and Social Psychology Bulletin*, 22, 979–992.

The Drenth Committee found strong indications of fraud in Chapters 3 and 5 of Mr Stapel's dissertation. The Drenth Committee was specifically charged with investigating possible fraud in the research on which Mr Stapel's dissertation was based, and whether that should have repercussions on the doctorate awarded to him by the University of Amsterdam. After all there was no need to answer the last question because Mr Stapel surrendered his doctor's degree certificate to the Rector Magnificus of the University of Amsterdam at his own accord. Regarding the feasibility in principle of withdrawing a doctor's degree, see the Noort Committee's recommendations and legal comments in Appendix 7.

The Committees' findings show that fabrication of data in one form or another started before the Tilburg period. The first publication in Groningen in which fraud has been proven is from 2004, and the first publication where evidence of fraud was found dates back to 2001. During the Amsterdam period the first publication where evidence of fraud was found was from 1996.

## 3.2 Nature of the fraud

The Committees have established pursuant to their findings that Mr Stapel manipulated existing datasets or fabricated his own. The following picture emerged from the information obtained in the interviews with co-authors, colleagues, and other people involved, of how the collection, analysis and reporting of data was usually manipulated.

In a first variant the research preparation would proceed normally. Stimuli and questionnaires were developed, and the number of subjects and the procedure to be followed were determined after extensive discussions with the co-authors. In the end, however, the questionnaires were not administered. In these cases Mr Stapel created his own datasets, which largely confirmed the prior expectations.

In a second variant Mr Stapel altered data in existing datasets. These data actually were collected and were usually entered by a student assistant, who then forwarded the dataset to Mr Stapel. After manipulating the data Mr Stapel passed the dataset to the PhD student, who had been required to set down his or her expectations about the outcomes in advance. The Committees are in possession of datasets entered by student assistants and the datasets ultimately given to the PhD student. The differences between the datasets clearly show the changes made.

In a third variant Mr Stapel had contacts with fellow researchers and then sent them datasets that he claimed to have had in his possession for some considerable time, asking them to analyse and write up the study. In this variant too, Mr Stapel supplied fabricated datasets of obscure origin.

Where the Drenth Committee concluded from irregularities, inaccuracies (e.g. incorrect number of degrees of freedom and p-values) and extremely improbable data or data distributions that there was evidence of fraud in several studies in the Amsterdam period, Mr Stapel himself has described his research method in Amsterdam as 'grey'. By his own account he now deems his conduct in Amsterdam involving omitting conditions and variables, or strategically and selectively repeating experiments to be scientifically inadmissible. According to Mr Stapel, his research practice in Groningen went from 'grey' to 'black', and he says in Tilburg the manner of fabrication became 'ever crazier, faster and stranger'. An extensive account of Mr Stapel's methods is given in Sections 4.2 to 4.5, inclusive.

### **3.3 Reports of serious infringement of scientific integrity**

In addition to fraudulent infringements of scientific integrity, the Committees also encountered other violations of proper scientific research procedures, which are discussed in anonymized form in Chapter 5.

The still ongoing learning process that PhD students find themselves in means that they cannot be fully blamed for improper research procedures, if they were involved in them. However, this argument does not apply in the same way to post-doctoral researchers and scientific staff. They must be expected to perform independent, critical and proper research; they are actually appointed to educate undergraduates and PhD students and are responsible for teaching them a critical attitude.

The Levelt Committee observed a sufficiently serious violation of scientific standards on the part of a few individuals to warrant passing their names in confidence to the Rector Magnificus of Tilburg University.

## 3.4 Co-authors

The Committees found no evidence that co-authors deliberately collaborated on data falsification with Mr Stapel. However the Levelt and Noort Committees concluded that a more critical attitude should have been expected from some co-authors, for example where data copying was reflected in the results in published tables. The Levelt and Noort Committees also established scientific carelessness on the part of co-authors in a number of publications. Their conduct does not comply with the fundamental rules of sound scientific research. By way of example: several similar experiments were performed, where only one or a few yielded the expected results; the latter were then offered for publication with no new independent replications and no mention of the unsuccessful cases. Also without mention, numerous subjects were omitted from the analyses on dubious grounds, or certain (test) findings were misreported or deliberately omitted. This suggests a culture in which scientific integrity is not held in high esteem. The section on culture discusses this point in more detail.

## 3.5 Dissertations Tilburg

The Levelt Committee has established with certainty that fictitious data were also used in doctoral research supervised by Mr Stapel in Tilburg. The Committee first established that there was no fraud in the thesis of Dr F. van Horen and Dr C. Wiekens. On the other hand, it is certain that all chapters of the dissertations of Dr Y.R. Avramova and Dr M.K. Noordewier do rely on fictitious data. Fraud was found in one chapter of the thesis of Dr L. van den Broek. Ms M.H.C. Meijers, who was due to defend her thesis on 7 October 2011, withdrew the thesis of her own accord when she learned of the data fraud. In her case the Committee confined itself to assessing the article that had already been published.

The Levelt Committee then asked whether these findings should have repercussions for previously awarded doctoral degrees. The Committee used the following criterion: were the PhD students concerned aware that the data were fictitious, or was there an element of culpable ignorance? In answer, the Committee has ascertained that none of the cases involved culpability or complicity. In the main all these PhD students and co-authors were misled with great subtlety, and to a significant degree were also educated within the improper research culture outlined above. Furthermore, notwithstanding the chapters where fictitious data were used, the PhD students also performed studies in which they genuinely did collect their own data. However, because the findings were considerably less interesting than the results of the donated data, these studies remained unpublished.

The Committee is therefore of the opinion that there should be no repercussions on the awarded doctoral degree in any of these cases. Nonetheless, it is proper that the fraudulent nature of the data used in these dissertations be made known, in order to prevent possibly false conclusions being drawn, and to curtail the impact of possibly false interpretations that have been offered.

## 3.6 Dissertations Groningen

Following discussions with all researchers whose PhDs were earned in Groningen about their articles and chapters published prior to, or based on, their dissertations, the Noort Committee arrived at the same conclusion as the Levelt Committee. There was no question of any knowledge or complicity in the fraud on the part of these PhD students. Neither can these junior researchers from Groningen be said to have been ‘culpable’, i.e., negligently ignorant, as defined by the Levelt Committee. The Noort Committee therefore shares the Levelt Committee’s opinion that the data fraud committed by Mr Stapel should have no repercussions on the degrees awarded to these PhDs.

Opinions are mixed regarding the dissertations, as distinct from the opinions about the people who were awarded the PhDs. The main criterion used by the Noort Committee was the extent to which data was collected, entered, analysed and interpreted by the researchers themselves. Regrettably, datasets provided or processed by Mr Stapel had actually been used in some cases. There were also cases in which data at a certain stage of interpretation were in Mr Stapel’s possession beyond the control of the PhD student, and where the data was altered by Mr Stapel.

The Noort Committee found no fraud in the above sense in the theses of Dr M.B. Braun-Ekker, Dr E. Kamans, Dr S.J. Ko, Dr J. Lammers, Dr H. Oldenhuis, and Dr S.W. van der Velde. In these cases Mr Stapel was supervisor for formal reasons only, or it has been established that the PhD student had control of all data from start to finish. The situation is different for the dissertations of Dr Janneke F. Joly, Dr J. Grob, Dr S.A. Schwinghammer, Dr D. Trampe, Dr A. van den Bos, Dr M. Maringer and Dr L. Renkema. The Noort Committee’s analyses have revealed fraudulent action on the part of Mr Stapel in one or more chapters of these dissertations: see Section 3.1 and Appendix 5.

## 3.7 Impact of the fraud

The Committees find that great harm has been done to co-authors, and PhD students in particular, as a consequence of the fraud committed by Mr Stapel. The Committees became aware in the course of the interviews conducted that the consequences for those involved can be both formal and informal in nature. In a formal sense, the people affected are hampered in their careers, such as when extending temporary contracts and applying for grants. To have had Mr Stapel as PhD supervisor once worked in a student’s favour; but the opposite is now true.

On an individual level some victims have been demoralized. In an informal sense, there is an element of stigmatization that may persist long into their further career. For some victims the consequences may be more drastic than they are yet able to foresee.

The impact is not restricted to individuals, but the standing of science, and of social psychology in particular, has been badly dented by Mr Stapel’s actions. Other victims include the universities of Tilburg, Groningen and Amsterdam, whose reputations have been damaged. Relatively small institutions are particularly sensitive to their international reputation. Psychology research in Tilburg in particular will therefore be confronted with this case of fraud long into the future. It is important in this respect that the

Executive Board of Tilburg University immediately formed a Committee to get to the bottom of this case of fraud.

Academic publishers who have been obliged to withdraw published articles and providers of research funds whose grants were used fraudulently are also victims, as are fellow researchers who may have been denied the grants and direct funding that were awarded instead to Mr Stapel. Beyond Tilburg and Groningen much time and money have also been wasted on futile attempts to replicate or build upon the results that Mr Stapel published.

The Committees have noted that this case of fraud has led to greater international awareness of problematic aspects of customary research practices. The discussions about replication, data archiving and the research culture in general are being conducted with more emphasis than ever, and are set only to become more intense in the near future. This report returns to this point in the 'Recommendations' Chapter.



## 4 Research culture: working method and research environment

### 4.1 Introduction

In addition to determining the nature and extent of the fraud, the Committees were also charged with investigating the contributory factors that helped sustain this extensive fraud for so many years. An answer to this question is relevant in making recommendations on how to prevent any recurrence. This question is all the more acute in that it should have been possible, or even imperative, to note as irregular and undesirable various conspicuous aspects of Mr Stapel's methods and results, even in the absence of any immediate suspicion of fraud. Moreover, these peculiarities, in exceptional cases, had been observed at an earlier stage, but no action was taken.

The Committees have found that PhD students and co-authors were not accessories to the fraud committed by Mr Stapel (Section 3.4). The conclusion stated in the interim report that is upheld after intensive study of the publications and interviews with the co-authors is that PhD students and co-authors were not complicit. Nonetheless the Committees are forced to the conclusion that Mr Stapel's misrepresentations of the organization and progress of research, the remarkably fraudulent data, and in general Mr Stapel's findings, were invariably accepted without criticism, even when doubt would have been reasonable. It is the co-authors who were in the best position to spot the fraud. Why didn't they? We discuss below various factors both inside and outside Mr Stapel's research environment that encouraged this failure of regular scientific criticism. This analysis is based on the numerous interviews that the Committees conducted with Mr Stapel's fellow researchers, Mr Stapel himself, faculty administrators, PhDs, and co-authors.

The first and most important group of factors must be sought in Mr Stapel's working method and research environment. He abused the trust that is the basis of all scientific collaboration. He used his position of great prestige and power to commit fraud and to stifle any possible doubt about his methods. These aspects are covered in the following sections. Although the same tendencies can be observed in all three universities, certainly in Tilburg and Groningen, these aspects are nonetheless handled separately for each. Mr Stapel held different positions in the three institutions where he worked, and his methods were related to his position.

The second group of factors resides in the general way in which the research was conducted both within and outside Mr Stapel's working environment. This aspect is discussed in Chapter 5.

### 4.2 General

Before all else it must be borne in mind that reciprocal trust is essential in all collaboration between colleagues, and science is no exception. This is pre-eminently the case with the kind of master-apprentice

relationship that occurs in science (e.g. supervisor-PhD student, or senior researcher-junior researcher). The junior must be able to trust unconditionally in the integrity of the ‘master’. It is this aspect of trust that played a crucial part throughout the fraud affair: a scientist does not commit fraud; it is almost inconceivable that this might happen.

The fraud and the attendant abuse of trust could also go unobserved because the fraud was often planned and executed with subtlety, in the sense that Mr Stapel made sure that the entire course of the research appeared realistic. He was also a solo performer to the extreme in the execution of the research and the supervision of his PhD students. There are many details of this aspect in the descriptions of the Groningen and Tilburg situations.

These more detailed local descriptions also reveal Mr Stapel’s considerably powerful position, at any rate within the University of Groningen and even more so within Tilburg University. At the University of Amsterdam he already enjoyed a reputation as a ‘golden boy’. This ‘position of power’ will certainly have hampered the identification and reporting of fraud. Mr Stapel was also seen as a charismatic leader: people who are admired are not readily suspected of fraud, and there will be a tendency to put forward ‘alternative explanations’ rather than to suspect fraud. As a supervisor Mr Stapel’s relationship with his PhD students was also usually extremely friendly. This was another impediment to a critical attitude on the part of PhD students regarding the supervisor’s possibly fraudulent actions. Finally it is apparent to the Committees from various interviews that Mr Stapel did not hesitate to use his power, prestige and charisma to prevent the detection of fraud.

These general conclusions can be substantiated more firmly on many points, illustrations of which can be found in the accounts of the local inquiries and interviews.

### **4.3 Working method and research environment: University of Amsterdam**

The significance of the Amsterdam period (1993-1999) is that this was the time that Mr Stapel was introduced to the field of social psychology as a PhD student and junior researcher. In view of the clear signs of irregularities in the published results in a number of articles, as observed by the Drenth Committee, further investigation of the ‘research culture’ and the customary research practices in the department in which Mr Stapel worked at the time seemed useful. However, it was all long ago. For a variety of reasons, data or other survey material are no longer available, and interviews with those involved can no longer provide reliable and detailed insights. The Drenth Committee could therefore arrive at only general conclusions with respect to the scientific integrity ‘culture’ within the Faculty of Psychology at the University of Amsterdam.

No explicit attention was paid to the combating or prevention of scientific misconduct within the scientific research environment in the Netherlands, and there was certainly no formal Code for research integrity, until the end of the twentieth century. Whatever measures existed were more implicit, and any action was at the discretion of individual researchers. The Faculty of Psychology at the University of Amsterdam was no

exception. This is not to say that the Faculty of Psychology failed to take the standards of scientific integrity seriously, but rather that there was no specific attention and no formal structure directed at enforcing these standards. Furthermore there was no culture of transparency, information sharing, peer review and joint responsibility in this Faculty. It was easy for researchers to go their own way. Nobody looked over their shoulder, they did not share their data with other members of the various research groups, and there was no peer review. This rather naive delegation of trust and the absence of detailed checks of raw and coded data, in particular on the part of supervisors and co-supervisors, members of doctoral examination committees, tutorial leaders and co-authors, created a working environment and work culture in which any individual researcher who ever would attempt to transgress the boundaries of integrity was very likely to have gone unnoticed. People trusted each other's scientific integrity: no feedback, no warnings, no correction. Therefore, there was no culture in which scientific integrity was given a high priority through explicit attention, detailed monitoring, an implicit exemplary function of senior staff and a collective responsibility within research groups, where the members correct each other and keep each other alert.

Unlike in the (fraudulent) school research projects in the later period, in particular in the Tilburg period, the data from the Amsterdam period were based less on research among secondary school students, although school data of this kind were used, for instance for some articles on which his doctoral thesis was based (e.g. Chapter 3). In most cases, the data came from studies involving first or junior year students (Amsterdam, Eindhoven), or undergraduate students (University of Michigan, University of Georgia), see for example Chapter 5. First-year psychology students at the University of Amsterdam fulfilled an obligation to act as experimental subjects for a number of hours. Students were sometimes also paid to take part in studies. It has been verified that during the years 1995 - 2000 research was indeed conducted with first-year students, and that Mr Stapel participated as experimental leader.

In those experiments Mr Stapel nearly always took charge of collecting, coding and processing the data in isolation and without it being checked or verified. He was seen as a most promising and talented researcher who stood out as an informal leader among his peers.

## **4.4 Working method and research environment: University of Groningen**

### ***Mr D.A. Stapel***

On 1 May 2000 Mr Stapel, who until then was a Royal Dutch Academy of Arts and Sciences (KNAW) fellow at the University of Amsterdam, was appointed professor on a Van der Leeuw chair of Cognitive Social Psychology in the then PPSW faculty, now the Faculty of Behavioural and Social Sciences at the University of Groningen. The Van der Leeuw chairs were set up to give young scientists a fast track to promotion to ordinary chairs. The chair of the advisory appointment committee for the Van der Leeuw chair was the social psychologist Prof. A.P. Buunk, who put Mr Stapel forward in a nomination with only one candidate. The expectations for candidate Stapel were extremely high because of his prolific international scientific production. He was viewed as inspiring and stimulating and could be described as a brilliant researcher, if not a genius. His CV supported this view. He passed two doctoral examinations (psychology and communication science) cum laude (University of Amsterdam), in 1997, likewise cum laude, he obtained a doctorate in social psychology (for which he received the ASPO Thesis Prize), spent some

considerable time abroad and was awarded the 1999 Jos Jaspars Early Career Award by the EAESP. His list of publications for the 1993-1999 period was impressive. On 13 March 2000 Mr Stapel announced his acceptance of the appointment 'with enthusiasm', and inaugurated on 11 September 2001 with an address about the leading role of social psychology in the social sciences, entitled 'The queen's turn'. Six years later, on 28 May 2006, he resigned from the Faculty of Behavioural and Social Sciences with effect from 1 September 2006, in order to move to Tilburg University. Although he claimed to have worked with pleasure at the University of Groningen, he said 'that at this point in my life Tilburg University is more compatible with my personal development and ambitions.'

The date of his inaugural lecture, 11 September 2001, was important to Mr Stapel, but for entirely different reasons than those for which it is engraved in the collective memory. Mr Stapel referred in his inaugural lecture to a conference forty-five years earlier, in which the 'cognitive revolution was born'. According to Mr Stapel the cognitive approach was nothing new to social psychologists. The field already viewed people as being 'natural processors and transformers of information', but the cognitive revolution brought methodological 'refinement and precision'. His programme ambitions were far reaching:

*'At the heart of the land of life sciences lives social psychology. And that makes social psychology the queen of that scientific land. With in one hand a razor (as a symbol of the precision with which they cut away theoretical and rhetorical proliferation) and in the other a camera (as a symbol of its attention to the commonplace) they are the ideal mediator between the social and behavioural sciences. Social psychology. Physics may have crowned itself king, but social psychology is the queen' (Stapel, D.A. (2001)).*

And also the following quasi-apology in his lecture for crowning social psychology queen of sciences reinforces this point rather than undermining it. There is a bipolar element in the ambition to link the everyday with the precise, with 'attention to the real, important problems and on the other hand a methodology compatible with the precise analysis of natural or cognitive sciences'. The primacy of the theory – and therefore the subordinate role of the data? – had already been expressed by him at an earlier stage:

*'The freedom we have in the design of our experiments is so enormous that when an experiment does not give us what we are looking for, we blame the experiment, not our theory. (At least, that is the way I work). Is this problematic? No.'* (Stapel, D.A. (2000)).

He made this 'accusation' of the experiment in a field with differing attitudes towards the primacy of theory or data. From a perspective in which data were wholly subordinate to theory, it is seen that the required precision concerning the data far from always materialized. 'That is so normal for us', was heard several times in the interviews, when the discussion of the boundaries to be observed in handling data turned to lack of precision regarding the data.

All interviewees, that is, all co-authors, or more broadly Mr Stapel's students and most of his colleagues – despite their anger or distress – were full of praise about his capacities: This is true for the co-promotor who took a step sideways to make way for his rising star, to the student or research assistant who was proud to be part of Mr Stapel's team. The recognition he continued to receive following his appointment as professor was considerable, as is evident from an NWO Pioneer grant (2002), funds from the 'Breedtestrategie' (Lateral Strategy) and other subsidies. There was no lack of money for assistants, PhD students, and

projects. The media knew where to find him, and sometimes vice versa. Besides the praise for his dynamism, prolific stream of ideas, accessibility and energy, ‘ambitious’ was the word most often heard. And that in the superlative.

### ***The Stapel group***

The research in the Faculty of Behavioural and Social Sciences was conducted in relatively small groups with a great deal of autonomy. Within this ‘archipelago’ the Stapel group occupied a position of its own.

In this context several interviewees, including Prof. Buunk himself, spoke of difficult relations, or alienation, between Prof. Buunk and Mr Stapel and of a consequent absence of another full professor ‘to keep an eye on things’, both regarding Stapel’s own research and that of his PhD students. In the interview with the Committee Mr Stapel said he had not personally perceived relations with Prof. Buunk as difficult. In his view any possible alienation was attributable to differences in substantive interest.

Whereas Mr Stapel’s group of PhD students had a certain separate identity and pride, it would clearly be going too far to conclude that it was a closed group: PhD students mixed not only with PhD students from their own group (Stapel), but also with those from other groups.

With one exception Mr Stapel’s PhD students always had a second supervisor. The accent in the second supervisor’s role was on the writing process and substantive arguments, with scant attention paid to the data. The second supervisors had no misgivings regarding the quality of the data used. Within the framework of his own research line and his own PhD students Mr Stapel was recognized as the undisputed leader. Within that framework they saw only a trailing and marginal role for themselves.

Mr Stapel’s group held an honours class for motivated social psychology students, with regular sessions to discuss research. This little club was an important source for recruiting PhD students. Mr Stapel relieved many of his PhD students of work by engaging a student assistant for data entry. The PhD student would frequently collect the data in person, but the student assistant would give the data directly to Mr Stapel, and it might be altered before being forwarded to the PhD student. Also, at least once, the original data set was withdrawn by Mr Stapel for more detailed analysis in response to comments from a journal. The data appeared later to have been augmented with some ten new observations, as a result of which all the hypotheses were neatly confirmed. The PhD students perceived the help they received in data entry and processing as a luxury (not necessarily to be taken for granted) because it saved much time. The practice sometimes also provoked a degree of jealousy among PhD students, which in some cases was exacerbated by the alleged quality of the data. Otherwise Mr Stapel’s practice of ‘sitting on the data’ was also appreciated as a sign of great commitment to his PhD students’ research. Mr Stapel observed in this connection in his interview with the Noort Committee that he was personally convinced that he was helping his PhD students. By his own account the collection of the data was the greatest chore in research, which PhD students must be helped through as quickly as possible. The paradox involved in ‘helping’ through falsification became apparent to him only later.

Mr Stapel was generally viewed by his research group as a charismatic, congenial research leader (other words used by respondents included: ‘brilliant’, ‘creative’, ‘diligent’, ‘single-minded’, and ‘ambitious’). It was clear from the interviews that he really was in charge. It was well known that statistics were neither

his strongest point nor held his greatest interest. If necessary PhD students therefore put any questions of a statistical nature to researchers who were expert statisticians within the Faculty. Mr Stapel had no objection. The extremely fine data that PhD students would sometimes have in their possession was usually attributed to the strong theory underlying the data gathering. Furthermore, not everything was a success story: there were also PhD students who had 'poor' data, and who were therefore unable to book fine research results.

At the time of Mr Stapel's arrival in Groningen, the university needed to make up lost ground in terms of data collection (it is generally no easy matter to acquire sufficient data). The associated catch-up operation resulted in the formation of the datalab. There were frequent lab days (test days) in which students were interviewed (compulsory participation) to guarantee the regular acquisition of new research data. Numerous people contributed to the formation of the lab in Groningen. Mr Stapel had brought the phenomenon of test days from his time in Amsterdam. It is clear that lab and test days were vital to the research of Mr Stapel and his group. Not only did Mr Stapel now have the research ideas, but also the necessary research data. Otherwise, besides his excellent reputation and wealth of ideas, it was also his ability to acquire sufficient data that made Mr Stapel an attractive research partner for foreign researchers, according to several interviewees.

The normally weekly lab meetings discussed research, including methodology. However, no clear story emerged on this point. The Stapel group had no protocols for, for example, the collection of data (including standards for questionnaires) or research reports. The PhD students in Mr Stapel's group were not familiarized with fixed and clear standards. Mr Stapel's comment about all the above points was that there was much discussion in Groningen about methodology and the handling of data, and agreements about data handling certainly were made. According to him the situation in Groningen was no different on this point from usual practice elsewhere. He underscored the lack of (fixed and clear) standards; but that too was far from a local issue, according to Mr Stapel.

Occasionally PhD students would comment informally that some data appeared to be too good to be true, according to two interviewees. They said that Mr Stapel's leadership position, his recognized ability to form hypotheses and the fact that far from all data were beautiful, helped explain the lack of serious doubt regarding data fraud. There was also an occasion on which a PhD student expressed doubt about the quality of the data in conversation with a close colleague of Mr Stapel's (but not with the faculty confidential counsellor for PhD students). Because only conjecture was involved, no action was taken. In view of the seriousness of the allegation, strong evidence would be needed in order to take further action, which is where the matter rested.

The Noort Committee finds that the observed archipelago of fairly small, independent research groups is not peculiar to Groningen. This was the time of the rise of tenure tracks (careers based on individual performance with regular reviews) and ample opportunity for talent. Consistent with this was that individuals developed their own research lines. The Noort Committee has observed that Mr Stapel received little resistance from senior-researchers in his research group. Despite the Stapel group's own identity, it cannot be concluded clearly that there was an element of avoiding researchers in other groups. Mr Stapel's unusual approach to data entry and processing is remarkable, but hardly raised doubt in the Stapel group because PhD students themselves stood to gain from it. The Noort Committee notes the significance of the specific suspicion confided by a PhD student to a close colleague of Mr Stapel's, but has been unable to

gain sufficient insight into the details and background of the allegations to conclude with confidence that this signal should have led to further investigation.

## 4.5 Working method and research environment: Tilburg University

Mr Stapel was appointed full professor at Tilburg University on 1 September 2006 for the TIBER institute in formation, a joint undertaking of the Faculty of Social Sciences and the Faculty of Economics. On 1 September 2010 Mr Stapel was appointed dean of the faculty, which then bore the name Tilburg School of Social and Behavioral Sciences. In 2010 the faculty comprised nine departments, including the department of Social Psychology.

For everyone who collaborated academically with Mr Stapel, in particular his peers and juniors in the Department of Social Psychology at Tilburg University, trust in Mr Stapel's scientific integrity was absolute. The last thing that colleagues, staff, and students would suspect is that, of all people, the department's scientific star, and faculty dean, and the man who personally taught his department's scientific ethics course, would systematically betray that trust. Many saw him as a charismatic leader whose dedication to staff and students was excellent. Only few would have had the courage or inclination to entertain the idea of data fraud.

Neither would it have been a simple matter to detect the fraud, in view of his ingenious approach. The essence of the commonest form of Mr Stapel's methods was as follows. He would develop, in intensive one-to-one interaction with another scientist (usually a research Master's student, a PhD student, or post-doctoral researcher, but sometimes a senior colleague), a theory that answered intelligently and creatively the research question, which was usually formulated together with the partner. Ingenious experimental manipulations were designed to test the theory, which were prepared and documented in detail. Mr Stapel would demonstrate substantial knowledge of the literature, and considerable skill in designing appropriate experimental manipulations. After careful discussion, the materials for the experiment, such as precise descriptions of the experimental set-up, detailed instructions for the investigators on site, piles of questionnaires, experimental supplies, such as pictures, sweets as reward, etc. would be prepared by the scientific partner, in the quantities needed. Often, at Mr Stapel's request, the scientific partner would also produce a table of expected empirical findings of all the hypotheses to be tested.

Up to this point everything was in order. However, there then followed an entirely fictitious phase. The experiments were seemingly executed under the complete supervision of Mr Stapel alone. According to his own account, Mr Stapel had excellent contacts with many educational institutions in the country. They were apparently always willing to be persuaded by him to perform studies of this kind, sometimes with the assistance of (fictitious) paid research assistants. To compensate the schools for their efforts, Mr Stapel claimed to give lectures and provide the schools concerned from time to time with computers and video projectors. It is apparent from Mr Stapel's statements to the Levelt Committee that the questionnaires and additional materials loaded into the back of his car (e.g. bars of chocolate for use in an experiment) never went to schools but were dumped in a container. Mr Stapel himself would then sit at his computer with

an empty questionnaire to create a corresponding dataset. He gave his partners (PhD students, co-authors and colleagues) the impression that data were collected at the schools by (unknown) assistants, who also processed and coded the data. The data ‘obtained’ in this way were then given directly to Mr Stapel: never to the partners. All these ‘efforts’ usually resulted several weeks later in a (fictitious) dataset being made available in its entirety to the partner for more detailed analysis, or even directly in the form of tables with the necessary averages, standard deviations, reliability checks, test results, and so on. The partner would be able to start writing the article immediately, on his own or in intensive cooperation with Mr Stapel.

This approach applies to most of the fraud encountered by the Levelt Committee. There were also variants of the above. Sometimes Mr Stapel would discuss with a colleague an interesting study that the colleague was working on, remarking that he had an old dataset, which he had once collected, but had then done nothing with so far, but that matched the colleague’s needs perfectly. Soon this fictitious dataset would be made available to the colleague, who would then be in a position to produce a nice article, obviously with Mr Stapel as co-author. He would also say in Groningen or elsewhere that he had obtained data from the Social Psychology lab in Tilburg, or could arrange for the data to come from Tilburg. In Tilburg, the same could happen in reverse with respect to Groningen. The Levelt Committee also came across a study that was actually carried out, where the raw data was sent to Mr Stapel and a couple of weeks later the PhD student received a table with the (embellished) findings. Finally, some studies were entirely invented, and ‘executed’ in solo, e.g. a study in a train station with unknown research assistants. The co-author only helped write the journal article; the dataset was never shared.

Critical questions about data collection were not encouraged in the discussions at lab meetings. Mr Stapel was always clear in his communication that someone was required to stop asking questions. Several PhD students also stated in interviews with the Levelt Committee that Mr Stapel misused his position of power in order to silence them. Mr Stapel told a research Master’s student who found suspicious patterns in the data: ‘If you want to be taken on here you will have to demonstrate that you can get something finished, so just write up the results.’ A PhD student from another department was required to tidy up her room. Mr Stapel saw boxes of completed questionnaires and wanted them to be thrown away. The PhD student said she had to keep them for five years. Mr Stapel contradicted her, adding: ‘If you want to commit fraud, you could do it anyway.’

The Levelt Committee noted another striking aspect of his style of supervising PhD students. On the one hand Mr Stapel had an intensive one-to-one working relationship with the young researchers, and many PhD students viewed him as a personal friend. They visited his home, had meals together, went to the cinema, and so on. On the other hand, however, were the threats when critical questions were asked. It would then be made clear to the PhD student concerned that such questions were seen as a lack of trust and that none should be asked. It was precisely the close relationship with Mr Stapel that made it difficult for a junior researcher to see anything in this other than well-intentioned constructive criticism from the senior partner.

The working method that the Levelt Committee reconstructed and is described above was discussed with Mr Stapel in an interview with the Committee. During the interview Mr Stapel expressed unwillingness to explain the details of his method in fabricating data because he found it personally too painful. He divulged only that there was no system in the data fraud and that he was quick to make the changes because he

himself actually wanted to avoid knowing. He did state that he did not identify with the picture sketched of him as a manipulator in order to facilitate his fraud.

Many of the peculiarities in his methods (and findings) have only emerged clearly with hindsight with the knowledge now in our possession. However, people have repeatedly seen one or the other feature of the fraud. Some discounted any misgiving as a stray thought, because they could not believe fraud was being committed, or felt powerless against Mr Stapel. However, there were others who gave the matter more attention, and yet others discussed it. Why did this discussion not lead to earlier discovery?

One possible factor in Tilburg may have contributed to the late recognition of suspicions of fraud. In the past year, six young researchers have expressed doubts about Mr Stapel's scientific integrity. Who should you approach with doubts of this kind? In accordance with the rules of the National Committee For Scientific Integrity (LOWI), the person to approach is the Confidential Counsellor appointed by the university. Contrary to LOWI guidelines ('The function of Confidential Counsellor is incompatible with that of member of the Executive Board, faculty dean, director of a research school, or educational or research institute'), the Confidential Counsellor in Tilburg was the Rector Magnificus. Certainly for an undergraduate or young PhD student, that is a very lofty and distant position. The Rector Magnificus is one of the most powerful people in the university. He is not someone you approach, it would be obvious to think, without watertight proof of your suspicions. Moreover, since his appointment, and as dean, Mr Stapel was assured of considerable support from the Executive Board. A financial audit of the Tilburg School of Social and Behavioral Sciences (TSB) shows that as full professor Mr Stapel already occupied a special position in the Faculty. The Executive Board granted him a variety of allowances and reimbursements when he was appointed that were so exceptional for TSB that they had to be channelled through the Economics faculty. Furthermore his expenses claims were not checked in the usual manner, despite good reason to do so. 'He was hardly called to account, if at all.' (TSB financial report, 9 June 2012). The situation did not improve when he himself became dean. Mr Stapel was extremely generously facilitated and treated as untouchable. Signals of this kind about Mr Stapel's special position were issued constantly, and naturally penetrated to the 'faculty shop floor'.

The Committee's terms of reference include inquiring into the prevailing culture that facilitated and sustained the fraud. This aspect is important in 'making recommendations on how to prevent any recurrence'. It was apparent to the Committee in the course of the inquiry and in the final phase in particular that even after his dismissal the Faculty did not entirely stop facilitating Mr Stapel and upheld a flow of information with extremely confidential elements. The Committee has reported this matter to the current Executive Board. The Executive Board responded immediately by setting up an inquiry. It is not within the Committee's remit to anticipate the findings of this still ongoing inquiry.

In this context a person who suspects fraud will in the first instance approach a close colleague (professor or other member of staff), in the hope that they will know what steps to take next. And even this course of action is not without problems, since the kind of relationship that the colleague might have with the alleged fraudster is unknown. It makes it much easier for whistleblowers if there is an independent university confidential counsellor for scientific integrity, who is not part of the management line.

In 2010 and 2011 three mentions of fraud were addressed to members of the academic staff in psychology.

The first two were not followed up in the first or second instances. Mr Stapel's virtually unassailable position may have played a part. The third report, to the head of department, and extraordinarily carefully prepared by three young and, certainly in their position, vulnerable whistleblowers, was immediately picked up in a professional way, with the now familiar result.

Suspicious about data provided by Mr Stapel had also arisen among fellow full professors on two occasions in the past year. These suspicions were not followed up. The Committee concludes that the three young whistleblowers showed more courage, vigilance and inquisitiveness than incumbent full professors.

## 5 Research culture: flawed science

### 5.1 Introduction

Another set of explanatory factors for the extent and duration of the fraud, alongside those set out in Chapter 4, reside in the general manner in which the research was performed both within and outside Mr Stapel's working environment. It involved a more general failure of scientific criticism in the peer community and a research culture that was excessively oriented to uncritical confirmation of one's own ideas and to finding appealing but theoretically superficial ad hoc results. This picture emerged, among other things, from the interviews with the Committees. These 'sloppy science methods' were found on a large scale in the work of Mr Stapel and his co-workers, as explained below in this chapter. The failure on all levels of the scientific review procedures is explicitly discussed in the concluding subsection

The Committees embarked on an extensive study of all research material, available data and publications, and interviewed those involved in depth in order to gain a better understanding of the extent and nature of the fraud, and of how it could possibly have happened in the research environment in which Mr Stapel, his PhD students and close colleagues worked. The statisticians initially analysed the dissertations and publications with a view to gain a deeper, clearer and more definitive picture of the fraud. The statisticians and the Committees subjected the collection, coding, analysis and reporting of the data to critical scrutiny.

However, the Committees could not be blind to evident violations of the basic rules of sound scientific research. The Committees were forced increasingly to the conclusion that, even in the absence of fraud in the strict sense, there was a general culture of careless, selective and uncritical handling of research and data. The observed flaws were not minor 'normal' imperfections in statistical processing, or experimental design and execution, but violations of fundamental rules of proper scientific research with a possibly severe impact on the research conclusions. The Committees are of the opinion that this culture partly explains why the fraud was not detected earlier.

### 5.2 Generalizability of the findings from local to national and international culture

The discovery of the methodological defects, which constitutes an unintended and unexpected finding of this inquiry, did raise the crucial question for the Committees as to whether this research culture, which is described in more technical detail below, is also rife throughout the field of social psychology, nationally and internationally. Could it be that in general some aspects of this discipline's customary methods should be deemed incorrect from the perspective of academic standards and scientific integrity?

It is relevant in this connection that the Committees' investigations covered all Mr Stapel's publications, which, while numerous, obviously do not constitute a random sample of social psychology publications.

The Committees are therefore unable to make any statement on these grounds about social psychology as a whole. It would nonetheless be simplistic to dismiss the findings given below as merely a local aberration. Mr Stapel worked in too many different places and in too many different capacities to support that view: from the start of his research career at the University of Amsterdam, through his first professorship in Groningen, to achieving celebrity status in Tilburg. Likewise he also collaborated with too many people from diverse universities in the Netherlands and abroad. Furthermore he published in nearly all the respected international journals in his field. It was extremely rare for his extraordinarily neat findings to be subjected to serious doubt, even in the doctoral boards of ‘his’ doctoral candidates, even in the international review procedures, and even where the fraud was blatant. Taken together all of the above reinforces the picture of an international research community of which Mr Stapel, his PhD students and close colleagues were part, and in which the customary research methods and associated standards and values were mutually shared.

Another clear sign is that when interviewed, several co-authors who did perform the analyses themselves, and were not all from Stapel’s ‘school’, defended the serious and less serious violations of proper scientific method with the words: that is what I have learned in practice; everyone in my research environment does the same, and so does everyone we talk to at international conferences.

It goes without saying that the Committees are not suggesting that unsound research practices are commonplace in social psychology. The Committees are unwilling or unable to make any statement about social psychology in general, although they consider the findings of this report to be sufficient reason for the field of social psychology in the Netherlands and abroad to set up a thorough internal inquiry into the state of affairs in the field. Sections 5.3 to 5.6, inclusive, might serve as input to such an inquiry, which will be one of the ‘Recommendations’. The Committees look very favourably upon various initiatives that have already been taken in this field.

## 5.3 Verification bias and missing replications

One of the most fundamental rules of scientific research is that an investigation must be designed in such a way that facts that might refute the research hypotheses are given at least an equal chance of emerging as do facts that confirm the research hypotheses. Violations of this fundamental rule, such as continuing to repeat an experiment until it works as desired, or excluding unwelcome experimental subjects or results, inevitably tend to confirm the researcher’s research hypotheses, and essentially render the hypotheses immune to the facts.

Procedures had been used in the great majority of the investigated publications that lead to what is referred to here as verification bias. There are few articles in which all the mentioned violations of proper scientific method were encountered simultaneously. On the other hand, publications were rarely found in which one or more of these strategies was not used.

Verification bias is not the same as the ‘usual’ publication bias, which is the phenomenon in which negative or weak findings that do not clearly confirm the theoretical expectations, if at all, but were obtained in soundly executed research, do not find their way into the journals, unlike ‘positive’ results. Verification bias refers to something more serious: the use of research procedures in such a way as to ‘repress’ negative results by some means.

Verification bias arises in the investigated publications in a variety of ways, the most important of which are enumerated below.

- An experiment fails to yield the expected statistically significant results. The experiment is repeated, often with minor changes in the manipulation or other conditions, and the only experiment subsequently reported is the one that did yield the expected results. It is unclear why in theory the changes made should yield the expected results. The article makes no mention of this exploratory method; the impression created is of a one-off experiment performed to check the a priori expectations. It should be clear, certainly with the usually modest numbers of experimental subjects, that using experiments in this way can easily lead to an accumulation of chance findings. It is also striking in this connection that the research materials for some studies shows the use of several questionnaire versions, but that the researchers no longer knew which version was used in the article.
- A variant of the above method is: a given experiment does not yield statistically significant differences between the experimental and control groups. The experimental group is compared with a control group from a different experiment –reasoning that ‘they are all equivalent random groups after all’ – and thus the desired significant differences are found. This fact likewise goes unmentioned in the article.
- The removal of experimental conditions. For example, the experimental manipulation in an experiment has three values. Each of these conditions (e.g. three different colours of the otherwise identical stimulus material) is intended to yield a certain specific difference in the dependent variable relative to the other two. Two of the three conditions perform in accordance with the research hypotheses, but a third does not. With no mention in the article of the omission, the third condition is left out, both in theoretical terms and in the results. Related to the above is the observed verification procedure in which the experimental conditions are expected to have certain effects on different dependent variables. The only effects on these dependent variables that are reported are those that support the hypotheses, usually with no mention of the insignificant effects on the other dependent variables and no further explanation.
- The merging of data from multiple experiments. It emerged both from various datasets and interviews with the co-authors that data from multiple experiments had been combined in a fairly selective way, and above all with benefit of hindsight, in order to increase the number of subjects to arrive at significant results.
- Research findings were based on only some of the experimental subjects, without reporting this in the article. On the one hand ‘outliers’ (extreme scores on usually the dependent variable) were removed from the analysis where no significant results were obtained. This elimination reduces the variance of the dependent variable and makes it more likely that ‘statistically significant’ findings will emerge. There may be sound reasons to eliminate outliers, certainly at an exploratory stage, but the elimination must then be clearly stated.
- Conversely, the Committees also observed that extreme scores of one or two experimental subjects were kept in the analysis where their elimination would have changed significant differences into insignificant ones; there was no mention anywhere of the fact that the significance relied on just one or a few subjects.

- Finally entire groups of respondents were omitted, in particular if the findings did not confirm the initial hypotheses, again without mention. The reasons given in the interviews were ad hoc in nature: ‘those students (subjects) had participated in similar experiments before’; ‘the students just answered whatever came into their heads’, but the same students had in the first instance simply been accepted in the experiment and the analysis. If the omitted respondents had yielded different results they would have been included in the analyses.
- The reliabilities of the measurement scales used and the reporting thereof were often handled selectively, and certainly to the experimenters’ advantage, in the sense of confirming the research hypotheses. It is impossible to ‘test’ hypotheses with unreliable measurement instruments, and therefore, for example, the reliability was estimated for a part of the research group, with the unreported omission of ‘unreliable’ subjects. Or items were selected differently for each study in such a way, which did lead to a reliable instrument, but with no awareness that that this was achieved at the expense of the mutual comparability of studies.
- Where discrepancies were found between the reliabilities as reported by the researcher (usually the alpha coefficient) and as calculated by the statisticians, the reported values were usually conspicuously higher.  
If the reliability of a dependent variable or a covariate was not reported, its value often turned out to be too low relative to the accepted standard (e.g. less than 0.60).
- Sometimes the reliability was deliberately not reported, in particular if it was extremely low. For instance, a co-author reported that the supervisors urged that the data be sold as effectively as possible, and discouraged attempts to undermine the data, which might make editors and reviewers suspicious. If they asked any questions the missing data would be provided later.
- There was also selective treatment of the measurement scales, depending on what it was required to prove. Various one or two dimensions (underlying variables) were used with the same set of items in different experiments, depending on what was most expedient in the light of the research hypotheses.
- The following situation also occurred. A known measuring instrument consists of six items. The article referred to this instrument but the dataset showed that only four items had been included; two items were omitted without mention. In yet another experiment, again with the same measuring instrument, the same happened, but now with two different items omitted, again without mention. The only explanation for this behaviour is that it is meant to obtain confirmation of the research hypotheses. It was stated in the interviews that items that were omitted did not behave as expected. Needless to say, ‘good’ ad hoc reasons were given, but none were mentioned in the publication, and the omissions could be ascertained only by systematically comparing the available survey material with the publication.
- When the re-analysis revealed differences in significance level (p values) when applying the same statistical tests, it was usual for the values reported in the articles to be ‘more favourable’ for the researcher’s expectations. Similarly, incorrect rounding was also found, for example:  $p = 0.056$  became  $p = 0.05$ .

Many of the above verification procedures have some value in a more exploratory setting, but then their use must be reported explicitly. The results of exploration of this kind would then be confirmed in a new independent replication prior to publication. It is remarkable that this hardly happened, if at all, although experiments of this type are often relatively simple to replicate.

Interviews with co-authors and with Mr Stapel revealed that many of the procedures mentioned above were mainly oriented towards the rapid positive handling of data collection, analysis and reporting, with a view to meeting the publication criteria. Co-authors often did not happen to consider this to be careless research. The lack of replication research can also be included under this heading of rapid publication. Many gave publication pressure as an important motive for the methods they used. The Committees refrain from comment about whether this explanation is sound, or is merely an excuse. It would be advisable to investigate this matter in a broader context.

## 5.4 Incomplete or incorrect information about the research procedures used

It follows from the fundamental principles of openness and controllability that research procedures must be described in a way that allows for accurate replication of a given experiment. The previous section has already stated that elements of the research procedures that were used selectively to achieve confirmation of the hypotheses by whatever means, often went unmentioned. Furthermore numerous discrepancies were found between the way the experiment was actually carried out as could be deduced from the available data and the research material, and what was stated in the article. These discrepancies were not necessarily oriented to verification, but did obscure what actually had happened and thereby rendered the experiment unverifiable for others. The information provided was moreover often too vague to support replication of the research. From interviews with co-authors it even emerged that they were sometimes unaware of the specific nature of the stimulus material before starting to write the article. The experimental design was then described ‘creatively’, or in other words as the co-author imagined it, or in general and vague terms. This is surprising in that researchers repeatedly stated in the interviews with the Committees that social psychology is often concerned with extremely precise and subtle effects and manipulations. Mr Stapel actually had a reputation as an extremely inventive and skilled experimenter, which is inconsistent with careless accounts of experimental design and execution.

Some concrete examples, a small selection from the countless cases of laxity and vagueness, are given below.

- The subjects were only identified in very broad terms, such as ‘students in the Netherlands’.
- Reference was made to an existing measuring instrument, but the fact that a nonstandard variant was actually used went unmentioned.
- Where a seven-point-scale was used, a five-point scale was reported.
- It was reported that ‘attractiveness’ was measured using the opinion of someone else present at the experiment, but in fact it was a self-assessment.
- No mention was made of the fact that the experiment was conducted in a session in which even more experimental conditions were varied and even more dependent variables measured. Measurements of this kind can influence each other.

- The reported number of experimental subjects was sometimes lower than the number actually investigated; was only a subgroup investigated?
- The datasets contained more variables than reported in the article, but the reverse also occurred: was this information obtained from other experiments with the same subjects, without mentioning the fact?
- A commonly used technique in the investigated publications was the use of fillers, which, as a diversionary tactic, experimental subjects are led to believe are relevant variables, but have no place in the analysis. This practice should be mentioned, but usually was not.
- The nature and extent of missing data were not stated, but simply omitted from the analysis without mention.
- ‘Names’ (John, Mary), which are often important elements of the experimental conditions, were invented in the article, or were different from the names actually used, or were not mentioned.

These and many other examples of carelessness in the investigated publications present a picture of a research environment surrounding Mr Stapel in which the diligent and critical handling of research and data were not held in high esteem, and were no part of the practical research education of PhD students.

## 5.5 Statistical flaws

The statisticians found countless flaws while studying and re-analysing all the survey material. This is inevitable to some extent; social psychology researchers cannot be expected to be aware of the latest, specialized techniques in the statistics field. But it is disturbing that statistical flaws frequently revealed a lack of familiarity with elementary statistics. For instance, one article had an entire series of t-test results on the difference between pairs of means, of the type:  $t = 0.90$ ,  $p < .05$  and therefore ‘significant’. The co-author was astounded to be told by the Levelt Committee that a co-author should have seen this absurdity.

Another co-author appeared to have no idea of what people actually ‘do’ in a factor analysis, or what the alpha reliability coefficient represents, let alone what impact the calculated unreliability of the covariates might have on the estimate of the experimental effect.

In one case the raw data were also replaced by ‘estimated data’, obtained from some form or other of smoothing or curve fitting. These were presented as the actual raw data. The co-author did not really see a problem in this, and also had no idea how the smoothing had happened, guessing that it was ‘by using an option in Excel’ (the statisticians were unable to reproduce the estimated data or link it with the original data in any way whatsoever).

These are just a few examples of statistical incompetence and lack of interest found. In the interviews co-authors usually did not perceive the above as a shortcoming or a problem.

## 5.6 Failure of scientific criticism

As set out above, many factors contributed to the late discovery of the fraud committed by Mr Stapel. Nonetheless the urgent question that remains is why this fraud and the widespread violations of sound scientific methodology were never discovered in the normal monitoring procedures in science.

The data and findings were in many respects too good to be true. The research hypotheses were almost always confirmed. The effects were improbably large. Missing, impossible, or out-of-range data are rare or absent. Highly conspicuous impossible findings went unnoticed.

It is almost inconceivable that co-authors who analysed the data intensively, or reviewers of the international 'leading journals', who are deemed to be experts in their field, could have failed to see that a reported experiment would have been almost infeasible in practice, did not notice the reporting of impossible statistical results, such as a series of t-values linked with clearly impossible p-values, and did not spot values identical to many decimal places in entire series of means in the published tables. Virtually nothing of all the impossibilities, peculiarities and sloppiness mentioned in this report was observed by all these local, national and international members of the field, and no suspicion of fraud whatsoever arose.

Where the immediate working environment of Mr Stapel was a factor, only PhD students and postdoctoral researchers were involved. There was no reason for other co-authors, fellow professors in the Netherlands and abroad, members of doctoral boards or reviewers of international journals to disregard their critical scientific function. The Committees can reach no conclusion other than that from the bottom to the top there was a general neglect of fundamental scientific standards and methodological requirements.

This certainly also applies to the editors and reviewers of international journals. Furthermore, many journals insist prior to publication on authors filling in forms in various variants guaranteeing correct research procedures and availability of data and survey material. Authors evidently frequently fail to comply (see among others Wicherts *et al.*, 2006). The journals perform no further monitoring of this requirement.

Co-authors also reported more than once in interviews with the Committees that reviewers encouraged irregular practices. For instance, a co-author stated that editors and reviewers would sometimes request certain variables to be omitted, because doing so would be more consistent with the reasoning and flow of the narrative, thereby also omitting unwelcome results. Reviewers have also requested that not all executed analyses be reported, for example by simply leaving unmentioned any conditions for which no effects had been found, although effects were originally expected. Sometimes reviewers insisted on retrospective pilot studies, which were then reported as having been performed in advance. In this way the experiments and choices of items are justified with the benefit of hindsight.

Not infrequently reviews were strongly in favour of telling an interesting, elegant, concise and compelling story, possibly at the expense of the necessary scientific diligence. It is clear that the priorities were wrongly placed. It is surely simple to post all the information of relevance to an article on a website and to provide an explicit reference in the article.

The preference of journals for elegant, concise, attractive findings is consistent with the assessment of various of the Committees' interviewees, that Mr Stapel has made no significant contribution to the social psychological theory. His main contributions were in inventive experimental applications and variants, and in 'interesting' findings. He had a good feeling for what was 'in', we were told. His citation score is no higher, and is rather lower than the average level of his Tilburg research group, as can be seen in various citation indices with and without Mr Stapel's publications. That many of his publications appeared in leading journals does not detract from this fact. This raises the question as to whether leading journals are sufficiently critical regarding publications that make no essential contribution to theory building in the field.

Time and again journals and experienced researchers in the domain of social psychology accepted that Mr Stapel's hypotheses had been confirmed in a single experiment, with extremely large effect sizes. Evidently he alone was in a position to achieve the precise manipulations needed to make the subtle effects visible. People accepted, if they even attempted to replicate the results for themselves, that they had failed because they lacked Mr Stapel's skill. However, there was usually no attempt to replicate, and certainly not independently. The few occasions when this did happen systematically, and failed, were never revealed, because this outcome was not publishable. In the meantime the Committees have been made aware of several cases of this kind in the Netherlands and abroad, in which much research funding and expensive research time has been wasted.

A 'byproduct' of the Committees' inquiries is the conclusion that, far more than was originally assumed, there are certain aspects of the discipline itself that should be deemed undesirable or even incorrect from the perspective of academic standards and scientific integrity.

In the case of the fraud committed by Mr Stapel, the critical function of science has failed on all levels. Fundamental principles of scientific method have been ignored, or set aside as irrelevant. In the opinion of the Committees this has contributed significantly to the delayed discovery of the fraud. It is to the credit of the whistleblowers in Tilburg that they did discover these infringements of scientific integrity and took the correct action.

However, the discussion of factors that facilitated Mr Stapel's fraud, the power derived from his celebrity-like status and the failure of the critical function of his scientific world do not alter the Committees' conclusion that the fraud committed by Mr Stapel was such that these factors and circumstances do not excuse his actions in any way whatsoever.

## 6 Recommendations

More than a year after the presentation of the interim report, it can be said that the Stapel case has served as a catalyst for further reflection and improvement. It seems that now, more than ever before, new violations of scientific integrity are being brought to light and investigated. Researchers who use statistical techniques to investigate publications for fraud are also receiving increased attention. Various general initiatives to prevent fraud have been taken and, by establishing committees and issuing reports, organizations such as the KNAW, the VSNU, the European Federation of Academies of Sciences and Humanities (ALLEA) and the Inter Academy Council (IAC) have contributed to the debate about violations of scientific integrity and their prevention.

For example, the recently presented report by the Schuyt Committee (KNAW) offers useful suggestions on the question of how research data should be handled carefully and with integrity. The Schuyt Committee concluded that it is not so much necessary to develop new rules but rather more important to impress the content of the rules that already exist on the researchers. By the same token, all four of the committees of enquiry just mentioned stress the institutional necessity of creating a research environment that promotes and facilitates research integrity. Although mutual trust remains at the basis of scientific collaboration, this cannot be taken for granted. Of course, the shortcomings identified must not result in organized distrust or overblown bureaucracy that unnecessarily impedes scientific work, but rather in the creation of a research environment in which researchers are encouraged, through coaching, training and effective controls, to take account of the rules for careful and honest academic research.

There have also been various similar initiatives in the field of psychology. The ASPO recently issued a report entitled, *Sharpening scientific policy after Stapel* (<http://www.sociale-psychologie.nl/>), with recommendations about training, data storage and replication research within social psychology. In *de Psycholoog*, the journal of the Dutch Association of Psychologists, various articles have also appeared featuring analyses of the Stapel case and suggestions for preventing academic misconduct. In November 2012, the journal *Perspectives on Psychological Science* published a special issue devoted to replication within psychological research.

In addition, a number of researchers have also launched a large-scale ‘Reproducibility Project’ (<http://openscienceframework.org/project/EZcUj/wiki/home>), in order to examine the possibility of replicating the original results of studies published in 2008 in three leading psychological journals (*Journal of Personality and Social Psychology*, *Psychological Science*, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*). The scientific integrity committee within the Psychology Faculty at the University of Amsterdam has proposed a number of precise and strict guidelines for responsible research (including a mandatory data protocol). In Tilburg, the dean has implemented the recommendations from the interim report concerning the management and storage of data and the organization of research. Similar measures have been taken in Groningen.

In the interim report, the Committees made a number of recommendations. The Committees are pleased that their recommendation to be generous to the disadvantaged PhD students has been adopted by the respective Rectors. They have now taken action and talks are ongoing with those concerned.

The recommendation report the criminal offences committed by Mr Stapel to the Public Prosecutor's Office has been followed up by Tilburg University and the University of Groningen and a preliminary criminal enquiry is now in progress.

There have also been financial investigations in Tilburg and Groningen. There is still a lack of clarity with regard to whether external and internal grants were properly spent and administered, and this situation will need to be rectified by means of further investigations.

The recommendation to investigate whether extremely unworthy scientific conduct that takes place after the award of the doctoral degree can be grounds for withdrawal is, according to the Drenth Committee, no longer applicable for Mr Stapel, since he has already surrendered his doctoral certificate of his own accord. However, this recommendation retains its general relevance for the future. Opinions vary on the legal possibilities for withdrawing a doctoral degree. Legislative changes would seem necessary if only to dispel any remaining doubts on this issue. It would appear that the Higher Education and Research Act (WHW) requires a specific addendum on the authority to withdraw degrees in cases where the withdrawal of a doctoral degree relates to conduct perpetrated after the conferral of the degree. For more details on this, see Appendix 7 by the Noort Committee.

Following the recommendations by the Levelt Committee, Tilburg University has appointed an independent confidential counsellor for integrity. Professor Cyrille Fijnaut, emeritus professor in criminal law and criminology at Tilburg Law School has agreed with the Executive Board of Tilburg University to take on this role. There have also been modifications to the regulations on scientific integrity.

In addition, Tilburg University has amended its policy on the supervision of PhD students. Every PhD student must now be supervised by two or more co-supervisors or supervisors. The recommendations from the interim report that have not yet been implemented or implemented in full are included below, supplemented by a number of additional recommendations.

### ***Further handling of the Stapel case***

The Rectors of the universities involved must ensure that the relevant journals are notified of the publications where fraud has been established or evidence of fraud has been revealed. Although it is the responsibility of the journals to withdraw the articles from their own or other databases in which these articles can be consulted or to designate them as fraudulent when the articles are requested, the universities must monitor whether this has actually happened.

With regard to the fraudulent or partially fraudulent dissertations, the universities involved must ensure that anyone requesting or given access to such a dissertation is also notified of its fraudulent or partially fraudulent nature. For example, this can be done by means of a reference to the present final report.

The universities concerned must inform the providers of research funds for Mr Stapel's research about any fraudulent publications that have flowed from this research. They must also determine how the grants were actually used at faculty and department levels.

## ***Code of conduct and supervision***

At faculty level, conditions must be created to promote honest scientific activity and the effective handling of complaints or suspicion of fraud. According to the guidelines agreed at national level in the Netherlands in the LOWI, a confidential counsellor for scientific integrity must be appointed and there must be regulations to guarantee the protection of whistle blowers in the case of reports of suspected infringements of scientific integrity. Every student and member of staff must be informed about the existence and content of the regulations on scientific integrity.

The Committees recommend that newly appointed staff members and PhD students be notified on appointment of the code of conduct that applies within the institution and that their appointment be subject to their acceptance of this code. Recommended codes of conduct are ‘The Netherlands Code of Conduct for Scientific Practice’ (VSNU, 2004, revised 2012) and the ‘European Code of Conduct for Research Integrity’ (ESF-ALLEA 2011). For those who apply statistics, the ‘ISI Declaration on Professional Ethics’ issued by the International Statistical Institute is also of relevance; see: <http://www.isi-web.org/about-isi/professional-ethics>.

In the training program for PhD students, the relevant basic principles of philosophy of science, methodology, ethics and statistics that enable the responsible practice of science must be covered. Based on these insights, research Master’s students and PhD students must receive practical training from their supervisors in the application of the rules governing proper and honest scientific research, which should include examples of such undesirable conduct as data massage. The Graduate School must explicitly ensure that this is implemented.

## ***PhD research***

The data on which an experimental psychology doctoral dissertation is based must as a rule be collected and analysed by the PhD students themselves.

Supervisors and co-supervisors must realize that the responsibility for the entire doctoral dissertation also involves checking and assessing the quality of the original data. They must be certain that the data has been collected, coded and processed honestly and responsibly.

The doctoral examination board must form a clear impression of the way in which research data has been collected. The mere fact that a chapter has already been accepted for publication in a peer-reviewed journal cannot be deemed to justify its approval.

The doctoral formulas must refer to both the rights and obligations towards science and society that are or will be attached to the doctorate by law or custom.

## ***Psychology research practice and publications***

In the recommendations, the Committees not only wish to focus on preventing or reducing fraud, but also on improving the research culture. The European Code refers to ‘minor misdemeanours’: some data massage, the omission of some unwelcome observations, ‘favourable’ rounding off or summarizing, etc. This kind of misconduct is not categorized under the ‘big three’ (fabrication, falsification, plagiarism) but it is, in principle, equally unacceptable and may, if not identified or corrected, easily lead to more serious

breaches of standards of integrity. For example, the Committees encountered many aspects of research practice that reveals sloppy science and verification bias. This provides justification for the following recommendations.

Far more than is customary in psychology research practice, replication must be made part of the basic instruments of the discipline and at least a few journals must provide space for the publication of replicated research. Conducting research replications must be included as part of the research Master's program.

Research data that underlie psychology publications must remain archived and be made available on request to other scientific practitioners. This not only applies to the dataset ultimately used for the analysis, but also the raw laboratory data and all the relevant research material, including completed questionnaires, audio and video recordings, etc. It is recommended that a system be applied whereby on completion of the experiment, the protocols and data used are stored in such a way that they can no longer be modified. It must be clear who is responsible for the storage of and access to the data. The publications must indicate where the raw data is located and how it has been made permanently accessible. It must always remain possible for the conclusions to be traced back to the original data. Journals should only accept articles if the data concerned has been made accessible in this way.

A publication must clearly indicate which co-authors have been completely responsible for the research presented and which are solely responsible for a specific part of it. If no specific indications are included in the list of authors, every author will be deemed to be fully responsible for the research presented.

The Committees recommend that the Dutch and international social psychology disciplines thoroughly reflect on and investigate the contribution to theory development in their discipline, the methodological validity of published social and psychological research as well as the review and other procedures in place for monitoring the theoretical relevance and methodological validity. This investigation can also be of benefit for related disciplines.

It is recommended that, where appropriate, the editors of journals use statistical techniques to identify probable irregularities in the reported results. Since it is likely that in the future that the identification of fraud in this way will become more difficult as fraudsters increasingly may make use of advanced simulation techniques available in standard statistical packages, the importance of complete and accessible data storage, replication research and improved review procedures must also continue to be emphasized.

### ***Assessment of research output***

Inspection committees that have assessed social psychology do not appear to have paid sufficient attention to a number of the facts highlighted in this report. They appear to have uncritically followed the assessments conducted by journals, both in terms of methodology and the contribution to theory. Another key issue in this area is the extent to which inspection and research assessment committees contribute to the continuation of a perceived pressure to publish and its associated conventions. These conventions particularly concern the number, sequence and responsibility of the co-authors and the repetitive publication of similar research results.

The Committees recommend that the VSNU pay additional attention to the aspects of the Standard Evaluation Protocol (SEP) that emphasize the quality of content.

## ***Disclosure***

The Committees have found that a confidential setting in which people can present their own account is an essential condition for finding the truth. In other words, full disclosure in public can be detrimental in attempting to identify the truth. For this reason, the Committees recommend that in the investigation of cases of scientific fraud, activities be conducted in secrecy, also in the future.

In each case, the results of these kinds of enquiries must then be made public, except in exceptional circumstances.

In the interests of trust in science, it is of great importance that clarity be achieved with regard to the prevalence and nature of fraud cases.

In this context, the Committees recommend that every scientific integrity committee should publish an annual report with information on numbers and the nature of cases. This report must be made public (except in exceptional circumstances) and be accompanied by information on the measures taken in response to the advice. This could take place on an individual university basis, but preferably also within a VSNU context, or within the context of the LOWI, in which the KNAW and NWO also participate.



# Appendix 1 Lists of Committee members and statisticians

## **Levelt Committee**

Prof. W.J.M. Levelt (chair)

[www.mpi.nl/people/levelt-pim](http://www.mpi.nl/people/levelt-pim)

Prof. M.S. Groenhuijsen

[www.tilburguniversity.edu/webwijs/show/?uid=m.s.groenhuijsen](http://www.tilburguniversity.edu/webwijs/show/?uid=m.s.groenhuijsen)

Prof. J.A.P. Hagenaars

[www.tilburguniversity.edu/webwijs/show/?uid=jacques.a.hagenaars](http://www.tilburguniversity.edu/webwijs/show/?uid=jacques.a.hagenaars)

Administrative secretary: Dr S.A.M. Baert

Statistician: Dr M.A.L.M. van Assen

## **Noort Committee**

Prof. E. Noort (chair)

[www.rug.nl/staff/e.noort](http://www.rug.nl/staff/e.noort)

Prof. H.E. Bröring

[www.rug.nl/staff/h.e.broring](http://www.rug.nl/staff/h.e.broring)

Prof. J.M. Pieters

[www.utwente.nl/gw/co/Medewerkers/Vakgroepvoorzitter/pietersnl.doc](http://www.utwente.nl/gw/co/Medewerkers/Vakgroepvoorzitter/pietersnl.doc)

Administrative secretary: Dr M. Jaspers

Statisticians: Prof. R.H. Koning

Dr W.J. Post

Dr M.A.J. van Duijn

## **Drenth Committee**

Prof. P.J.D. Drenth (chair)

[www.knaw.nl/Pages/DEF/26/294.bGFuZz1OTA.html](http://www.knaw.nl/Pages/DEF/26/294.bGFuZz1OTA.html)

Prof. J.W. Zwemmer

[www.uva.nl/over-de-uva/organisatie/medewerkers/content/z/w/j.w.zwemmer/j.w.zwemmer.html](http://www.uva.nl/over-de-uva/organisatie/medewerkers/content/z/w/j.w.zwemmer/j.w.zwemmer.html)

Prof. L.A. de Klerk

[www.uva.nl/over-de-uva/organisatie/medewerkers/content/k/l/l.a.deklerk/l.a.de-klerk.html](http://www.uva.nl/over-de-uva/organisatie/medewerkers/content/k/l/l.a.deklerk/l.a.de-klerk.html)

Prof. C.A.J. Klaassen (secretary)

[www.science.uva.nl/math/People/show\\_person.php?Person\\_id=Klaassen+C.A.J.](http://www.science.uva.nl/math/People/show_person.php?Person_id=Klaassen+C.A.J.)

Statisticians: Prof. H.L.J. van der Maas

Dr D. van Ravenzwaaij



# Appendix 2 Noort Committee questionnaire

## Noort Committee - Groningen

### *Questionnaire for co-authors of publications by Diederik Stapel*

1. Article/Publication (title, journal / book, year):
2. (Co)author:
3. In which period did you cooperate with Dr Stapel?
4. Concerning this publication: Who supervised and who conducted the research?
5. Who instigated and who funded this study?
6. Who determined the research question?
7. Who designed the research tools and setup (questionnaires etc.)?
8. Have you actually seen the experimental setup and tools and were you given the opportunity to comment on them?
9. Was the setup specially designed for this study?  
If yes, by who?  
If no, who designed them previously?  
If no, have the instruments been used previously by you or a colleague?
10. What did the raw data consist of (interviews, questionnaire, reaction times etc.)?
11. Who collected the (raw) data?
12. Have you seen the raw data?
13. Are the raw data still available?
14. Did you analyse the (raw) data?
13. Is there a logbook or history list describing the statistical analyses?
14. Who wrote the first version of the publication?
15. What was Dr Stapel's contribution to the publication?
16. Are the peer review reports (for the acceptance of the paper) still available?
17. Was an advisory committee involved in this study?  
If yes, which one?  
If yes, are there minutes of committee meetings?
18. Do you have any further remarks regarding this study?

# Noort Committee - Groningen

## *Questionnaire for authors of dissertations alongside Diederik Stapel*

1. Title of dissertation:
2. In which period did you cooperate with Dr Stapel?
3. Concerning this dissertation: Who supervised and who conducted the research?
4. Who instigated and who funded this study?
5. Who determined the research questions?
6. Who designed the research tools and setup (questionnaires etc.)?
7. Have you actually seen the experimental setup and tools and were you given the opportunity to comment on them?
8. Was the setup specially designed for this study?  
If yes, by who?  
If no, who designed them previously?  
If no, have the instruments been used previously by you or a colleague?
9. What did the raw data consist of (interviews, questionnaire, reaction times etc.)?
10. Who collected the (raw) data?
11. Have you seen the raw data?
12. Are the raw data still available?
13. Did you analyse the (raw) data?
14. Is there a logbook or history list describing the statistical analyses?
15. Who wrote the first version of the chapters in the dissertation?
16. What was Dr Stapel's contribution to the dissertation/publication?
17. Are there peer review reports available (for the acceptance of papers from the dissertation)?
18. Was an advisory committee involved in this study?  
If yes, which one?  
If yes, are there minutes of committee meetings?
19. Do you have any further remarks regarding this dissertation?

# Appendix 3 People interviewed by the Committees

Some people were inevitably interviewed by more than one Committee because of the chronological spread of the activities.

## Levelt Committee interviewees

Dr Y.R. Avramova  
Prof. I. van Beest  
Dr L.M. van den Broek  
Prof. A.P. Buunk  
Drs A.J. van Doorn  
Prof. P. Eijlander  
Drs M. Elshout  
Drs E.R.K. Evers  
Prof. G.L.M. van Heck  
Dr F. van Horen  
Dr Y. Inbar  
Dr C.S. Johnson  
Dr Janneke F. Joly  
Prof. G. Keren  
Prof. E.J. Krahmer  
Dr J. Lammers  
Dr D. Lerouge  
Drs L.A.J.G. van der Linde  
Prof. dr. S.M. Lindenberg  
Drs M.H.C. Meijers  
Drs M.H.J. Meijs  
Dr R.M.A. Nelissen  
Dr M.K. Noordewier  
Prof. F.G.M. Pieters  
Dr L. Renkema  
Dr K.I. Ruys  
Prof. C.G. Rutte  
Dr S.A. Schwinghammer  
Prof. G.R. Semin  
Mr D.A. Stapel  
Drs F.M. Terpstra  
Dr N. van de Ven  
Prof. T.M.M. Verhallen  
Prof. A.J.R. van de Vijver  
Prof. A.J.J.M. Vingerhoets

Dr C.J. Wiekens  
Drs J. van Wolferen  
Prof. M. Zeelenberg

## Noort Committee interviewees

Prof. J.M.F. ten Berge  
Prof. H. Blanton  
Drs. R. Boerrigter  
Dr J. W. Bolderdijk  
Dr A. van den Bos  
Prof. A.P. Buunk  
Prof. H.P.M. Creemers  
Prof. A. Dijksterhuis  
Dr M.A.J. van Duijn  
Prof. E.H. Gordijn  
Dr J.D.M. Grob  
Dr J.A.M. Heesink  
Dr Janneke F. Joly  
Dr E. Kamans  
Prof. H.A.L. Kiers  
Dr A. Klapwijk  
Dr S.J. Ko  
Dr W. Koomen  
Dr J. Lammers  
Prof. S.M. Lindenberg  
Dr M. Maringer  
Dr D. Marx  
Dr K. Massar  
Prof. D. Muller  
Dr H.K.E. Oldenhuis  
Prof. S. Otten  
Prof. K.I. van Oudenhoven-van der Zee  
Prof. J.L. Peschar  
Dr W.J. Post  
Prof. T.T. Postmes  
Dr L. Renkema  
Dr K.I. Ruys  
Dr S.A. Schwinghammer  
Dr F.W. Siero  
Prof. R. Spears  
Prof. J. Suls  
Prof. A. Tesser  
Dr D. Trampe

Dr S.W. van der Velde  
Prof. E. van de Vliert  
Dr C.J. Wiekens  
Prof. N. van Yperen

## **Drenth Committee interviewees**

Prof. K. van Dam  
Prof. A. Fischer  
Dr E. Jacobs  
Prof. G. Kerkhof  
Dr W. Koomen  
Dr A. Nienhuis  
Prof. J. van der Pligt  
Prof. N. Schwarz  
Dr U. Simonsohn  
Prof. R. Spears  
Dr A. Straathof  
Prof. A. Tesser  
Prof. H.K. Thierry  
Dr A.S. Velthuisen



# Appendix 4 List of publications examined by the Levelt Committee including findings

List of publications examined by the Levelt Committee

\* The Levelt Committee established fraud in this paper, see also the list ‘fraud established’

- \*Avramova, Y.R., Stapel, D.A. & Lerouge, D. (2010). The influence of mood on attribution. *Personality and Social Psychology Bulletin*, 36, 1360-1371.
- \* Avramova, Y.R., Stapel, D.A. & Lerouge, D. (2010). Mood and context-dependence: Positive mood increases and negative mood decreases the effect of context and perception. *Journal of Personality and Social Psychology*, 99, 203-214.
- \* Avramova, Y.R., & Stapel, D.A. (2008). Moods as spotlights: The influence of moods on accessibility effects. *Journal of Personality and Social Psychology*, 95, 542-554.
- \* Blanton, H., & Stapel, D.A. (2008). Unconscious and spontaneous and ... complex: The three selves model of social comparison assimilation and contrast. *Journal of Personality and Social Psychology*, 94, 1018-1032.
- Gordijn, E.H., & Stapel, D.A. (2008). When controversial leaders with charisma are effective: The influence of terror on the need for vision and impact of mixed attitudinal messages. *European Journal of Social Psychology*, 38, 389-411.
- Häfner, M., & Stapel, D.A. (2007). Attraktive Models in der Werbung: Auswirkungen sozialer Vergleiche auf den Betrachter und die beworbenen Produkte. In S. Trepte & E.H. Witte (Eds.), *Sozialpsychologie und Medien* (pp. 159-170). Lengerich: Pabst.
- Häfner, M., & Stapel, D.A. (2009). Familiarity can increase (and decrease) stereotyping: Heuristic processing or enhanced knowledge usability? *Social Cognition*, 27, 615-622.
- \* Häfner, M., & Stapel, D.A. (2010). Information to go: Fluency enhances the usability of primed information. *Journal of Experimental Social Psychology*, 46, 73-84.
- \* Johnson, C.S. & Stapel, D.A. (2007). No pain, no gain: The conditions under which upward comparisons lead to better performance. *Journal of Personality and Social Psychology*, 92, 1051-1067.
- \* Johnson, C.S. & Stapel, D.A. (2010). It depends on how you look at it: Being versus becoming mind-sets determine responses to social comparisons. *British Journal of Social Psychology*, 49, 703-723.
- \* Johnson, C.S., & Stapel, D.A. (2011). Happiness as alchemy: Positive mood and responses to social comparisons. *Motivation and Emotion*, 35, 165-180.
- \* Johnson, C.S., & Stapel, D.A. (2007). When different is better: Performance following upward comparison. *European Journal of Social Psychology*, 37, 258-275.
- \* Johnson, C.S., Norton, M.I., Nelson, L., Stapel, D.A., & Chartrand, T.L. (2008). The downside of feeling better: Self-regard repair harms performance. *Self and Identity*, 7, 262-277.
- Johnson, C.S., & Stapel, D.A. (2010). Harnessing social comparisons: When and how upward comparisons influence goal pursuit. *Basic and Applied Social Psychology*, 32, 234-242.
- Johnson, C.S., & Stapel, D.A. (2011). Reflection vs. Self-reflection: Sources of self-esteem boost determine behavioral outcomes. *Social Psychology*, 42, 144-151.
- Lammers, J., & Stapel, D.A. (2011). Power increases dehumanization. *Group Processes & Intergroup*

*Relations*, 14, 113-126.

- \* Lammers, J., & Stapel, D.A. (2011) Racist biases in legal decisions are reduced by a justice focus. *European Journal of Social Psychology*, 41, 375-387.
- Lammers, J., Stapel, D.A., & Galinsky, A.D. (2010). Power increases hypocrisy: Moralizing in reasoning, immorality in behavior. *Psychological Science*, 21, 737-744.
- Lammers, J., & Stapel, D.A. (2009). How power influences moral thinking. *Journal of Personality and Social Psychology*, 97, 279-289.
- Lammers, J., Stoker, J.I., Jordan, J., Pollmann, M.M.H., & Stapel, D.A. (2011). Power increases infidelity among men and women. *Psychological Science*, 22, 1191-1197.
- Lammers, J., Stoker, J.I., & Stapel, D.A. (2010). Power and behavioral approach orientation in existing power relations and the mediating effect of income. *European Journal of Social Psychology*, 40, 543-551.
- Lammers, J., Stoker, J.I., & Stapel, D.A. (2009). Differentiating social and personal power: Opposite effects on stereotyping, but parallel effects on behavioral approach tendencies. *Psychological Science*, 20, 1543-1549.
- \* Meijers, M.H.C., & Stapel, D.A. (2011). Me tomorrow, the others later: How perspective fit increases sustainable behavior. *Journal of Environmental Psychology*, 31, 14-20.
- \* Nelissen, R.M.A., Van de Ven, N., & Stapel, D.A. (2011). Status concerns and financial debts in adolescents. *Social Influence*, 6, 39-56.
- \* Noordewier, M.K., & Stapel, D.A. (2011). When failure feels better than success: Self-salience, self-consistency, and effect. *British Journal of Social Psychology*, 50, 321-330.
- \* Noordewier, M.K., & Stapel, D.A. (2011). Stereotype disconfirmation affect: When sweet hooligans make you happy and honest salesmen make you sad. *Basic and Applied Social Psychology*, 33, 1-6.
- \* Noordewier, M.K., & Stapel, D.A. (2010). Affects of the unexpected: When inconsistency feels good (or bad). *Personality and Social Psychology Bulletin*, 36, 642-654.
- \* Noordewier, M.K., & Stapel, D.A. (2009). Judging the unexpected: Disconfirmation of situation-specific expectancies. *European Journal of Social Psychology*, 39, 944-956.
- \* Noordewier, M.K., Van Horen, F., Ruys, K.I., & Stapel, D.A. (2010). What's in a name? 361.708 Euros: The effects of marital name change. *Basic and Applied Social Psychology*, 32, 17-25.
- Otten, S., & Stapel, D.A. (2007). Who is the Donald? How social categorization affects aggression-priming effects. *European Journal of Social Psychology*, 37, 1000-1015.
- Rosema, M., Jost, J.T., & Stapel, D.A. (2008). Social psychology and the study of politics. In L. Steg, A.P. Buunk and T. Rothengatter (Eds.), *Applied Social Psychology: Understanding and Managing Problems* (pp. 291-315). New York, NY, USA: Cambridge University Press.
- \* Ruys, K.I., & Stapel, D.A. (2008). The secret life of emotions. *Psychological Science*, 19, 385-391.
- \* Ruys, K.I., & Stapel, D.A. (2008). How to heat up from the cold: Examining the preconditions for (unconscious) mood effects. *Journal of Personality and Social Psychology*, 94, 777-791.
- \* Ruys, K.I., & Stapel, D.A. (2008). Emotion elicitor or emotion messenger?: Subliminal priming reveals two faces of facial expressions. *Psychological Science*, 19, 593-600.
- \* Ruys, K.I., & Stapel, D.A. (2009). The unconscious unfolding of emotions. *European Review of Social Psychology*, 20, 232-271.
- \* Ruys, K.I., Stapel, D.A. & Aarts, H. (2011). From (unconscious) perception to emotion: A global-to-specific unfolding view of emotional responding. In I. Nyklicek, A.J.J.M. Vingerhoets, & M. Zeelenberg (Eds.), *Emotion Regulation and Well-Being* (pp. 49-66). New York: Springer.

- Ruys, K.I., & Stapel, D.A. (2009). Learning to like or dislike by association: No need for contingency awareness. *Journal of Experimental Social Psychology*, *45*, 1277-1280.
- \* Stapel, D.A., (2007). In the mind of the beholder: The interpretation comparison model of accessibility effects. In D.A. Stapel & J. Suls (Eds.) *Assimilation and Contrast in Social Psychology* (pp. 313-327). New York, NY, USA: Psychology Press.
- \* Stapel, D.A. (2011). Priming as proxy: Understanding the subjectivity of social life. In K.C. Klauer, A. Voss and C. Stahl (Eds.), *Cognitive Methods in Social Psychology* (pp. 148-183). New York Guilford Publications.
- \* Stapel, D.A., & Johnson, C.S. (2007). When nothing compares to me: How defensive motivations and similarity shape social comparison effects. *European Journal of Social Psychology*, *37*, 824-838.
- \* Stapel, D.A., & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science*, *332*, 251-253.
- \* Stapel, D.A., & Marx, D.M. (2007). Distinctiveness is key: How different types of self-other similarity moderate social comparison effects. *Personality and Social Psychology Bulletin*, *33*, 439-448.
- \* Stapel, D.A., & Marx, D.M. (2007). Making sense of war: Using the Interpretation Comparison Model to understand the Iraq conflict. *European Journal of Social Psychology*, *37*, 401-420.
- \* Stapel, D.A., & Noordewier, M.K. (2011). The mental roots of system justification: System threat, need for structure, and stereotyping. *Social Cognition*, *29*, 238-254.
- \* Stapel, D.A., & Noordewier, M.K. (2011). Stop making sense: The ultimate fear. *Psychological Inquiry*, *27*, 615-622.
- \* Stapel, D.A., & Semin, G.R. (2007). The magic spell of language: Linguistic categories and their perceptual consequences. *Journal of Personality and Social Psychology*, *93*, 34-48.
- \* Stapel, D.A., & Van der Linde, L.A.J.G. (2011). What drives self-affirmation effects?: On the importance of differentiating value affirmation and attribute affirmation. *Journal of Personality and Social Psychology*, *101*, 34-45.
- Teunissen, P.W., Stapel, D.A., Scheele, F., Scherpbier, A.J.J.A., Boor, K., Van Diemen-Steenvoorde, J.A.A.M., & Van der Vleuten, C.P.M. (2009). The influence of context on residents' evaluations: Effects of priming on clinical judgment and affect. *Advances in Health Sciences Education*, *14*, 23-41.
- Teunissen, P.W., Stapel, D.A., Van der Vleuten, C.P.M., Scherpbier, A.J.J.A., Boor, K., & Scheele, F. (2009). Who wants feedback? An investigation of the variables influencing residents' feedback seeking. *Academic Medicine*, *84*, 910-917.
- Van den Bos, K., Martin, L.L., & Stapel, D.A., (2010). Meaning making in delayed-return cultures: The case of personal uncertainty. *Social and Personality Psychology Compass*, *4*, 559-573.
- \* Van Doorn, J., & Stapel D.A. (2011). When and How Beauty Sells: Priming, Conditioning, and Persuasion Processes, *Journal of Consumer Research*, published online June 1, 2011.
- Wiekens, C.J., & Stapel, D.A. (2010). Self-awareness and saliency of social versus individualistic behavioral standards. *Social Psychology*, *41*, 10-19.
- Wiekens, C.J., & Stapel, D.A. (2008). I versus we: The effects of self-construal level on diversity. *Social Cognition*, *26*, 368-377.
- Wiekens, C.J., & Stapel, D.A. (2008). The mirror and I: When private opinions are in conflict with public norms. *Journal of Experimental Social Psychology*, *44*, 1160-1166.

## **Dissertations:**

- \*Avramova, Y.R. (2010), *How the mind moods*
- \*Noordewier, M.K. (2009), *Consistency and the unexpected*
- Van Horen, F. (2010), *Breaking the mould on copycats: What makes product imitation strategies successful?*
- \*Van den Broek, L.M. (2009), *De ironie van gelijkheid*
- Wiekens, C.J. (2009), *Self-Awareness*

## **The Levelt Committee has established fraud in the following publications:**

- Avramova, Y.R., Stapel, D.A. & Lerouge, D. (2010). The influence of mood on attribution. *Personality and Social Psychology Bulletin*, 36, 1360-1371.
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (school data), data supplied by Mr. Stapel.
  - Very doubtful results, e.g., too strong effect sizes in particular given the reliabilities of the scales, all F's smaller than 1 for non-significant results.
- Avramova, Y.R., Stapel, D.A. & Lerouge, D. (2010). Mood and context-dependence: Positive mood increases and negative mood decreases the effect of context and perception. *Journal of personality and Social Psychology*, 99, 203-214.
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (school data), data supplied by Mr. Stapel.
  - Highly unlikely design of experiments, that is, unlikely next to impossible to realize experimental set up in the described circumstances.
  - Highly implausible results regarding effect sizes, lack of missing data.
- Avramova, Y.R., & Stapel, D.A. (2008). Moods as spotlights: The influence of moods on accessibility effects. *Journal of Personality and Social Psychology*, 95, 542-554. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (school data), data supplied by Mr. Stapel.
  - Results highly implausible, e.g., effect size measures are unreasonably high, all F's smaller than 1 when no effect is expected, main effects on separate elements of the composite score are exactly identical.
- Blanton, H., & Stapel, D.A. (2008). Unconscious and spontaneous and ... complex: The three selves model of social comparison assimilation and contrast. *Journal of Personality and Social Psychology*, 94, 1018-1032.
  - According to Mr. Stapel: fraudulent.
  - Suspiciously high effect sizes given single-item dependent variable.
- Häfner, M., & Stapel, D.A. (2010). Information to go: Fluency enhances the usability of primed information. *Journal of Experimental Social Psychology*, 46, 73-84. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
  - The reliabilities of accessibility (three items of word completion task for 'general evaluative+',

‘general evaluative-’, ‘meaning+’ and ‘meaning-’), are very low or even negative or zero, meaning that their items do not correlate at all. This is consistent with a procedure of first creating averages and then creating items that resulted in these averages without taking into account that these items should correlate.

- Several means of corresponding conditions in the first two experiments are almost identical for all conditions. This is very unlikely to impossible and highly suspicious especially since these are not pure replications, because there was a switch in stimuli.
  - The effect of the manipulation is much larger on the dependent variable than on the manipulation check which is strange as the manipulation check variable mediates the effect of the manipulation on the dependent variable. Moreover, the effect size is very large, given the reliability.
- Johnson, C.S., Norton, M.I., Nelson, L., Stapel, D. A., & Chartrand, T.L. (2008). The downside of feeling better: Self-regard repair harms performance. *Self and Identity*, 7, 262-277.
    - Conclusive evidence of copying data: For three of the items (hpab2, hpgo2, hpund2) the means are exactly the same (3.73 for N=40).
    - Three self-evaluation items have an unbelievably high reliability of .96. Surprisingly, an 8-item scale (study 2) of self-evaluation including the 3-item scale has a lower reliability.
  - Johnson, C.S. & Stapel, D.A. (2007). No pain, no gain: The conditions under which upward comparisons lead to better performance. *Journal of Personality and Social Psychology*, 92, 1051-1067. (Research supported by a Pioneer grant from NWO , and a research grant from the Heymans Institute of the University of Groningen both awarded to D.A. Stapel)
    - Conclusive evidence of copying data: In both conditions ‘no affirm’ and ‘global affirm’, the variable ‘attain’ has exactly the same mean and standard deviation (4.2143 and .5783 respectively).
    - = Effects unbelievably high on single items.
  - Johnson, C.S., & Stapel, D.A. (2007). When different is better: Performance following upward comparison. *European Journal of Social Psychology*, 37, 258-275. (Research supported by a Pioneer grant from NWO, and a research grant from the Heymans Institute of the University of Groningen both awarded to D.A. Stapel)
    - Conclusive evidence of copying data: For two items (hpabilit and hpgood the means are exactly the same (3.53 for N=70).
  - Johnson, C.S. & Stapel, D.A. (2010). It depends on how you look at it: Being versus becoming mindsets determine responses to social comparisons. *British Journal of Social Psychology*, 49,703-723.
    - According to Mr. Stapel: fraudulent.
    - Conclusive evidence of copying data: Two items have exactly the same scores for N=55.
    - Unbelievably high effects given the reliabilities of the items.
  - Johnson, C.S., & Stapel, D.A. (2011). Happiness as alchemy: Positive mood and responses to social comparisons. *Motivation and Emotion*, 35, 165-180.
    - According to Mr. Stapel: fraudulent.
    - Unbelievably high effects given the reliabilities of the items.

- Lammers, J., & Stapel, D.A. (2011). Racist biases in legal decisions are reduced by a justice focus. *European Journal of Social Psychology, 41*, 375-387.
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (data had allegedly been collected by Mr. Stapel among judges).
  - Results highly improbable, e.g., no missing data, too high effect sizes and correlations given reliability of the variables concerned.
  
- Meijers, M.H.C., & Stapel, D.A. (2011). Me tomorrow, the others later: How perspective fit increases sustainable behavior. *Journal of Environmental Psychology, 31*, 14-20. (Research supported by Transforum Grant WP-088 awarded to D.A. Stapel)
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (school data), data supplied by Mr. Stapel.
  - Implausible results, for example too high factor loadings, correlations, and too strong effect sizes, and no missing data.
  
- Nelissen, R.M.A., Van de Ven, N., & Stapel, D.A. (2011). Status concerns and financial debts in adolescents. *Social Influence, 6*, 39-56.
  - The article states explicitly that attractiveness was rated by the participant's neighbour. However, as is clear from the experimental instructions, self-rating was used instead. Self-rating and rating by others are definitely two different variables.
  - The article states that the grade point average on school performance was provided for each student by their teacher. However, given the stimulus materials, it is clear that the grade is based on a remembered and self-reported grade average. Again, these are two different variables and altogether an explicitly incorrect account of the actual research procedures.
  - Re-analysis of the data showed reliabilities for the variables attitude and subjective norm of .45 and .38 in contrast to the reported reliabilities in the article of .71 and .76. The data were collected and analysed by students, but these reliabilities and a factor analysis reported in footnote 1, were not checked by the supervisors.
  
- Noordewier, M.K., & Stapel, D.A. (2009). Judging the unexpected: Disconfirmation of situation-specific expectancies. *European Journal of Social Psychology, 39*, 944-956. (Research supported by a 'Tiber' grant from the Departments of Economics and Business and Social and Behavioral Sciences of Tilburg University awarded to D.A. Stapel).
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (school data), data supplied by Mr. Stapel.
  - Implausible results, e.g., implausibly high effect sizes for single item dependent variable, at the same time suspiciously high and low reliabilities of one scale in different conditions, and very high correlations between different variables in some conditions.
  
- Noordewier, M.K., & Stapel, D.A. (2010). Affects of the unexpected: When inconsistency feels good (or bad). *Personality and Social Psychology Bulletin, 36*, 642-654.
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (school data), data supplied by Mr. Stapel.
  - Results extremely unlikely, e.g., too high effects, too clean data (no missings).

- Noordewier, M.K., & Stapel, D.A. (2011). When failure feels better than success: Self-salience, self-consistency, and effect. *British Journal of Social Psychology*, 50, 321-330.
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (school data), data supplied by Mr. Stapel.
  - Results extremely unlikely, e.g., suspiciously high negative and high positive correlations, and also uncommonly high reliabilities of the Self-esteem Rosenberg scale.
  
- Noordewier, M.K., & Stapel, D.A. (2011). Stereotype disconfirmation affect: When sweet hooligans make you happy and honest salesmen make you sad. *Basic and Applied Social Psychology*, 33, 1-6.
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (school data), data supplied by Mr. Stapel.
  - Very implausible results, e.g., suspiciously high effect sizes given the nature of the manipulation; reported means too similar for independent replications.
  
- Noordewier, M.K., Van Horen, F., Ruys, K.I., & Stapel, D.A. (2010). What's in a name? 361.708 Euros: The effects of marital name change. *Basic and Applied Social Psychology*, 32, 17-25.
  - According to Mr. Stapel: fraudulent.
  - Results extremely unlikely, e.g., effect sizes too large for single item dependent variable, unbelievably high factor loadings, Cronbach's alpha extremely high.
  
- Ruys, K.I., & Stapel, D.A. (2008). The secret life of emotions. *Psychological Science*, 19, 385-391. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
  - According to Mr. Stapel: fraudulent.
  - Conclusive evidence of copying data: identical means and standard deviations in table 2 and table 5, belonging to different experiments. Identical means and standard deviations in table 6.
  - There is no correspondence between parts of the stimulus material and some of the information contained in the data or reported in the article.
  - Given the reliabilities ranging from -.2 to .36, the effect sizes are highly unlikely next to impossible. Strangely enough, for some of the same measurements, the reliabilities increase to .87 in the second study.
  
- Ruys, K.I., & Stapel, D.A. (2008). How to heat up from the cold: Examining the preconditions for (unconscious) mood effects. *Journal of Personality and Social Psychology*, 94, 777-791. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
  - According to Mr. Stapel: fraudulent.
  - Evidence of copying data: one mean and standard deviation in table 3 and table 4 are exactly the same.
  
- Ruys, K.I., & Stapel, D.A. (2008). Emotion elicitor or emotion messenger?: Subliminal priming reveals two faces of facial expressions. *Psychological Science*, 19, 593-600. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
  - According to Mr. Stapel: fraudulent.
  - Conclusive evidence of copying data: many cells in table 1 report the same mean and standard deviation, involving 15 out of 32 cells. Moreover, table 1 of article (37) contains 8 cells with exactly

the same mean and standard deviation as table 1 in article (39).

- Given the reliabilities ranging from -.07 to .23, the effect sizes are highly unlikely, next to impossible.
- Ruys, K.I., & Stapel, D.A. (2009). The unconscious unfolding of emotions. *European Review of Social Psychology*, 20, 232-271. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
  - According to Mr. Stapel: fraudulent.
  - Theoretical article, in important respect based on fraudulent articles.
- Ruys, K.I., Stapel, D.A. & Aarts, H. (2011). From (unconscious) perception to emotion: A global-to-specific unfolding view of emotional responding. In I. Nyklicek, A.J.J.M. Vingerhoets, & M. Zeelenberg (Eds.), *Emotion Regulation and Well-Being* (pp.49-66). New York: Springer.
  - According to Mr. Stapel: fraudulent.
  - Theoretical article, in important respect based on fraudulent articles.
- Stapel, D.A., (2007). In the mind of the beholder: The interpretation comparison model of accessibility effects. In D.A. Stapel & J.Suls (Eds.) *Assimilation and contrast in social psychology* (p 313-327). New York, NY, USA: Psychology Press. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
  - According to Mr. Stapel: fraudulent.
  - Theoretical article, in important respect based on fraudulent articles.
- Stapel, D.A. (2011). Priming as proxy: Understanding the subjectivity of social life. In K.C. Klauer, A. Voss and C. Stahl (Eds.), *Cognitive methods in social psychology* (pp. 148-183). New York Guilford Publications.
  - According to Mr. Stapel: fraudulent.
  - Theoretical article, in important respect based on fraudulent articles.
- Stapel, D.A., & Johnson, C.S. (2007). When nothing compares to me: How defensive motivations and similarity shape social comparison effects. *European Journal of Social Psychology*, 37(5), 824-838. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
  - According to Mr. Stapel: fraudulent.
  - Conclusive evidence of copying data: the reported mean and standard deviation of ‘psychology upward’ in Table 2, page 833 is exactly the same as the reported mean and standard deviation of ‘psychology upward’ in Table 3, page 836.
- Stapel, D.A., & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science*, 332, 251-253.
  - According to Mr. Stapel: fraudulent.
  - Fake data collection (data had allegedly been collected by Mr. Stapel at a station).
- Stapel, D.A., & Marx, D.M. (2007). Distinctiveness is key: How different types of self-other similarity moderate social comparison effects. *Personality and Social Psychology Bulletin*, 33, 439-448. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)

- Highly implausible test outcomes, especially series of obtained F-values  $< 1$  and high p-values.
  - Suspiciously homogeneous scores within a group.
  - Reported means in table 1 are three times the same for independent replications.
  - The coauthor could not offer any other explanation of these findings, than data fraud by Mr. Stapel.
- Stapel, D.A., & Marx, D.M. (2007). Making sense of war: Using the Interpretation Comparison Model to understand the Iraq conflict. *European Journal of Social Psychology*, 37, 401-420. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
    - Highly unlikely that the data collection in the train (study 1) took place. The students thanked in the acknowledgements for their help conducting the research declared not to have participated in this data gathering. Both the second author and the students had no recollection of how the data were gathered.
    - Suspiciously high effect sizes.
    - Suspicious answering patterns in datasets.
  - Stapel, D.A., & Noordewier, M.K. (2011). The mental roots of system justification: System threat, need for structure, and stereotyping. *Social Cognition*, 29, 238-254.
    - According to Mr. Stapel: fraudulent.
    - Unlikely gender composition in samples of subjects.
    - Highly unbelievable results, e.g., seven out of seven F-values  $< 1$  for research hypotheses where no effect is expected (and later: six out of six  $F < 1$ ); very high correlations between scales of need for structure and stereotyping, higher than can be expected from these scales' reliabilities.
  - Stapel, D.A., & Noordewier, M.K. (2011). Stop making sense: The ultimate fear. *Psychological Inquiry*, 27, 615-622.
    - According to Mr. Stapel: fraudulent.
    - Theoretical article, in important respect based on fraudulent articles.
  - Stapel, D.A., & Semin, G.R. (2007). The magic spell of language: Linguistic categories and their perceptual consequences. *Journal of Personality and Social Psychology*, 93, 34 – 48. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from 'Tiber', awarded to D.A. Stapel)
    - According to Mr. Stapel: fraudulent.
    - Suspiciously high correlations and effect sizes, given assumed reliability of single-item measure.
    - Suspicious differences in standard deviations for the same variable in different experiments in combination with almost impossibly high standard deviation in one experiment.
    - Several independent attempts to replicate the results in Amsterdam, Hong Kong and Michigan failed.
  - Stapel, D.A., & Van der Linde, L.A.J.G. (2011). What drives self-affirmation effects?: On the importance of differentiating value affirmation and attribute affirmation. *Journal of Personality and Social Psychology*, 101, 34-45.
    - According to Mr. Stapel: fraudulent.
    - Fake data collection (school data), data supplied by Mr. Stapel.

- Extremely unlikely correlations given the reliabilities of the items.
- Van Doorn, J., & Stapel D.A. (2011). When and How Beauty Sells: Priming, Conditioning, and Persuasion Processes, *Journal of Consumer Research*, published online June 1, 2011.
- According to Mr. Stapel: fraudulent.
  - Conclusive evidence of copying data in the datasets that were originally provided by Mr. Stapel. The findings in the article are based on datasets that were later provided by Mr. Stapel replacing the previous dataset.
  - Suspiciously high reliabilities for some items and highly unbelievably large effect sizes given the low reliabilities.

## **Dissertations:**

- Avramova, Y.R. (2010), *How the mind moods*

Chapter 2, see article:

Avramova, Y.R., & Stapel, D.A. (2008). Moods as spotlights: The influence of moods on accessibility effects. *Journal of Personality and Social Psychology*, 95, 542-554.

- According to Mr. Stapel: fraudulent.
- Fake data collection (school data), data supplied by Mr. Stapel.
- Results highly implausible, e.g., effect size measures are unreasonably high, all F's smaller than 1 when no effect is expected, main effects on separate elements of the composite score are exactly identical.

Chapter 3 *Rumination and reflection effects on social perception*

- Fake data collection (school data), data supplied by Mr. Stapel.
- Results highly implausible, e.g., suspiciously high effect of weak manipulation on single item, suspiciously high reliabilities, absence of correlation between two identical items.

Chapter 4, see article:

Avramova, Y.R., Stapel, D.A. & Lerouge, D. (2010). Mood and context-dependence: Positive mood increases and negative mood decreases the effect of context and perception. *Journal of Personality and Social Psychology*, 99, 203-214.

- According to Mr. Stapel: fraudulent.
- Fake data collection (school data), data supplied by Mr. Stapel.
- Highly unlikely design of experiments, that is, unlikely next to impossible to realize experimental set up in the described circumstances.
- Highly implausible results regarding effect sizes, lack of missing data.

Chapter 5, see article:

Avramova, Y.R., Stapel, D.A. & Lerouge, D. (2010). The influence of mood on attribution. *Personality and Social Psychology Bulletin*, 36, 1360-1371.

- According to Mr. Stapel: fraudulent.
- Fake data collection (school data), data supplied by Mr. Stapel.
- Very doubtful results, e.g., too strong effect sizes in particular given the reliabilities of the scales,

all F's smaller than 1 for non-significant results.

- Noordewier, M.K. (2009), *Consistency and the unexpected*

Chapter 2, see article

Noordewier, M.K., & Stapel, D.A. (2010). Affects of the unexpected: When inconsistency feels good (or bad). *Personality and Social Psychology Bulletin*, 36, 642-654.

- According to Mr. Stapel: fraudulent.
- Fake data collection (school data), data supplied by Mr. Stapel.
- Results extremely unlikely, e.g., too high effects, too clean data (no missings).

Chapter 3, see article

Noordewier, M.K., & Stapel, D.A. (2011). Stereotype disconfirmation affect: When sweet hooligans make you happy and honest salesmen make you sad. *Basic and Applied Social Psychology*, 33, 1-6.

- According to Mr. Stapel: fraudulent.
- Fake data collection (school data), data supplied by Mr. Stapel.
- Very implausible results, e.g., suspiciously high effect sizes given the nature of the manipulation; reported means too similar for independent replications.

Chapter 4 *Disconfirmation of price expectancies: When money can't buy you happiness*

- Fake data collection (school data), data supplied by Mr. Stapel.
- Unlikely results, e.g., suspiciously strong effect on single item.

Chapter 5, see article

Noordewier, M.K., & Stapel, D.A. (2011). When failure feels better than success: Self-salience, self-consistency, and effect. *British Journal of Social Psychology*, 50, 321-330.

- According to Mr. Stapel: fraudulent.
- Fake data collection (school data), data supplied by Mr. Stapel.
- Results extremely unlikely, e.g., suspiciously high negative and high positive correlations, and also uncommonly high reliabilities of the Self-esteem Rosenberg scale.

- Van den Broek, L.M. (2009), *De ironie van gelijkheid*

Chapter 7

- Raw data is different from data on which the analyses were performed.
- Results highly implausible, e.g., extremely unlikely outcomes of F-tests (all F's smaller than 1 when no effect is expected, i.e., 16 out of 16 times for study 1 and 17 out of 17 for study 2).



# Appendix 5 List of publications examined by the Noort Committee including findings

List of publications examined by the Noort Committee

\* *The Noort Committee determined fraud in this paper, see also the list 'fraud established'*

# *The Noort Committee found evidence of fraud in this paper, see also the list 'evidence of fraud'*

^ *The Noort Committee has found the following in this chapter, it is (partly) based on articles of which the Committees have found evidence of fraud*

- Gordijn, E., Koomen, W., & Stapel, D.A. (2001). Level of prejudice in relation to knowledge of cultural stereotypes. *Journal of Experimental Social Psychology, 37*, 150-157.
- \*Gordijn, E.H., & Stapel, D.A. (2006). Behavioural effects of automatic interpersonal versus intergroup social comparison. *British Journal of Social Psychology, 45*, 717-729.
- \*Joly, J.F. & Stapel, D.A. (2009). Staff, miter, book, share: How attributes of Saint Nicholas induce normative behavior. *European Journal of Social Psychology, 39*, 145-153.
- \*Joly, J.F., Stapel, D.A., & Lindenberg, S.M. (2008). Silence and table manners: When environments activate norms. *Personality and Social Psychology Bulletin, 34*, 1047-1056.
- Ko, S.J., Judd, C.M., & Stapel, D.A. (2009). Stereotyping based on voice in the presence of individuating information: Vocal femininity affects perceived competence but not warmth. *Personality and Social Psychology Bulletin, 35*, 198-211.
- Ko, S.J., Muller, D., Judd, C.M., & Stapel, D.A. (2008). Sneaking in through the back door: How category-based stereotype suppression leads to rebound in feature-based effects. *Journal of Experimental Social Psychology, 44*, 822-839.
- Koomen, W., Visser, M., & Stapel, D.A. (2000). The credibility of newspapers and fear of crime. *Journal of Applied Social Psychology, 30*, 921-934.
- \*Lindenberg, S.M., Joly, J.F., & Stapel, D.A. (2011). The norm-activating power of celebrity: The dynamics of success and influence. *Social Psychology Quarterly, 74*, 98-120.
- \*Maringer, M., & Stapel, D.A. (2007). Unfinished business: How completeness affects the impact of emotional states and emotion concepts on social judgments. *Journal of Experimental Social Psychology, 43*, 712-718.
- \*Maringer, M., & Stapel, D.A. (2009). Correction or comparison? The effects of prime awareness on social judgments. *European Journal of Social Psychology, 39*, 719-733.
- ^Martin, L.L., Strack, F., & Stapel, D.A. (2001). What's on our mind and how did it get there? The role of knowledge accessibility in social psychological phenomena. In A. Tesser & N. Schwarz (Eds.), *Intraindividual processes* (pp. 236-256). Malden, MA: Blackwell Publishers. (Blackwell Handbook of Social Psychology).
- \*Marx, D.M., & Stapel, D.A. (2006). It's all in the timing: Measuring emotional reactions to stereotype threat before and after taking a test. *European Journal of Social Psychology, 36*, 687-698.
- Marx, D.M., & Stapel, D.A. (2006). Understanding stereotype lift: The role of the social self. *Social Cognition, 24*, 776-792.
- \*Marx, D.M., & Stapel, D.A. (2006). Distinguishing stereotype threat from priming effects: On the role

- of the social self and threat-based concerns. *Journal of Personality and Social Psychology*, 91, 243-254.
- Marx, D.M., & Stapel, D.A. (2006). It depends on your perspective: The role of self-relevance in stereotype-based underperformance. *Journal of Experimental Social Psychology*, 42, 768-775.
  - Marx, D.M., Stapel, D.A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparison under threat. *Journal of Personality and Social Psychology*, 88, 432-446.
  - Renkema, L.J., Stapel, D.A., & Van Yperen, N.W. (2008). Go with the flow: Confirming to others in the face of existential threat. *European Journal of Social Psychology*, 38, 747-756.
  - Renkema, L.J., Stapel, D.A., & Van Yperen, N.W. (2009). Quantity and proximity: The terror-managing function of abstract figures. *Social Cognition*, 27, 929-938.
  - \*Renkema, L.J., Stapel, D.A., Maringer, M., & Van Yperen, N.W. (2008). Terror management and stereotyping: Why do people stereotype when mortality is salient? *Personality and Social Psychology Bulletin*, 34, 553-564.
  - \*Schwinghammer, S.A., & Stapel, D.A. (2006). The effects of different types of self-activation on social comparison orientation. *Social Cognition*, 24, 703-722.
  - \*Schwinghammer, S.A., & Stapel, D.A. (2011). Measure by measure: When implicit and explicit social comparison effects differ. *Self and Identity*, 10, 166-173.
  - Spears, R., Gordijn, E., Dijksterhuis, A.P., & Stapel, D.A. (2004). Reaction in action: Intergroup contrast in automatic behavior. *Personality and Social Psychology Bulletin*, 30, 605-616.
  - Stapel, D.A. (2000). Moving from fads and fashions to integration: Illustrations from knowledge accessibility research. *European Bulletin of Social Psychology*, 12, 4-27.
  - Stapel, D.A. (2001). De koningin is aan zet: Waarom de alledaagse precisie van de sociale psychologie haar majesteitelijk maakt. *Nederlandsch tijdschrift voor psychologie*, 56, 284-292.
  - ^Stapel, D.A. (2002). Wat je ziet ben je zelf: Contexteffecten op oordelen en gedrag. *Nederlandsch tijdschrift voor psychologie*, 57, 31-41.
  - ^Stapel, D.A. (2003). Making sense of hot cognition: Why and when description influences our feelings and judgments. In J.P. Forgas, K.D. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 227-250). Cambridge: Cambridge University Press. (Sydney Symposium of Social Psychology series, 5).
  - \*Stapel, D.A., & Blanton, H. (2004). From seeing to being: Subliminal social comparisons affect implicit and explicit self-evaluations. *Journal of Personality and Social Psychology*, 87, 468-481.
  - Stapel, D.A., Joly, J.F., & Lindenberg, S.M. (2010). Being there with others: How people make environments norm-relevant. *British Journal of Social Psychology*, 49, 175-187.
  - Stapel, D.A., & Koomen, W. (2000). Distinctness of others, mutability of selves: Their impact on self-evaluations. *Journal of Personality and Social Psychology*, 79, 1068-1087.
  - Stapel, D.A., & Koomen, W. (2000). The impact of opposites: Implications of trait inferences and their antonyms for person judgment. *Journal of Experimental Social Psychology*, 36, 439-464.
  - Stapel, D.A., & Koomen, W. (2001). I, we, and the effects of others on me: How self-construal moderates social comparison effects. *Journal of Personality and Social Psychology*, 80, 766-781.
  - Stapel, D.A., & Koomen, W. (2001). Let's not forget the past when we go to the future: On our knowledge of knowledge accessibility effects. In G.B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 229-246). Mahwah NJ: Lawrence Erlbaum.
  - Stapel, D.A., & Koomen, W. (2001). The impact of interpretation versus comparison goals on knowledge accessibility effects. *Journal of Experimental Social Psychology*, 37, 134-149.

- #Stapel, D.A., & Koomen, W. (2001). When we wonder what it all means: Interpretation goals facilitate accessibility and stereotyping effects. *Personality and Social Psychology Bulletin*, 27, 915-929.
- Stapel, D.A., & Koomen, W. (2005). Competition, cooperation, and the effects of others on me. *Journal of Personality and Social Psychology*, 88, 1029-1038.
- Stapel, D.A., & Koomen, W. (2005). Multiple ambiguity and an impression formation focus make inapplicable primes useful. *Social Cognition*, 23, 324-335.
- Stapel, D.A., & Koomen, W. (2005). When less is more: The consequences of affective primacy for subliminal priming effects. *Personality and Social Psychology Bulletin*, 31, 1286-1295.
- #Stapel, D.A., & Koomen, W. (2006). The flexible unconscious: Investigating the judgmental impact of varieties of unaware perception. *Journal of Experimental Social Psychology*, 42, 112-119.
- #Stapel, D.A., Koomen, W., & Ruys, K. (2002). The effects of diffuse and distinct affect. *Journal of Personality and Social Psychology*, 83, 60-74.
- \*Stapel, D.A., & Marx, D.M. (2006). Hardly thinking about others: On target closeness and cognitive busyness in social comparison effects. *Journal of Experimental Social Psychology*, 42, 397-405.
- Stapel, D.A., & Schwinghammer, S.A. (2004). Defensive social comparison and the constraints of reality. *Social Cognition*, 22, 147-167.
- \*Stapel, D.A., & Schwinghammer, S.A. (in press). Me and the why of social comparison: how self-activation determines social comparison motives. *Self and Identity*.
- \*Stapel, D.A., & Suls, J. (2004). Method matters: Effects of implicit versus explicit social comparisons on activation, behavior, and self-views. *Journal of Personality and Social Psychology*, 87, 860-875.
- Stapel, D.A., & Tesser, A. (2001). Self-activation increases social comparison. *Journal of Personality and Social Psychology*, 81, 742-750.
- \*Stapel, D.A., & Van der Zee, K.I. (2006). The self salience model of other-to-self effects: Integrating self-enhancement, complementarity, and imitation principles. *Journal of Personality and Social Psychology*, 90, 258-271.
- \*Trampe, D., Stapel, D.A., & Siero, F.W. (2007). On models and vases: Body dissatisfaction and proneness to social comparison effects. *Journal of Personality and Social Psychology*, 92, 106-118.
- \*Trampe, D., Stapel, D.A., & Siero, F.W. (2011). The Self-Activation Effect of Advertisements: Ads can Affect Whether and How Consumers Think about the Self. *The Journal of Consumer Research*, 37, 1030-1045.
- \*Trampe, D., Stapel, D.A., Siero, F.W., & Mulder, M. (2010). Beauty as a tool: The effect of model attractiveness, product relevance, and elaboration likelihood on advertising effectiveness. *Psychology & Marketing*, 27, 1101-1121.
- \*Van den Bos, A., & Stapel, D.A. (2008). The impact of comprehension versus self-enhancement goals on group perception. *Social Psychology*, 39, 222-230.
- \*Van den Bos, A., & Stapel, D.A. (2009). Why people stereotype affects how they stereotype: The differential influence of comprehension goals and self-enhancement goals on stereotyping. *Personality and Social Psychology Bulletin*, 35, 101-113.
- Van der Velde, S.W., Stapel, D.A., & Gordijn, E.H. (2010). Imitation of emotion: When meaning leads to aversion. *European Journal of Social Psychology*, 40, 536-542.

## Dissertations

- Braun-Ekker, B.M. (2010), *Dealing with a deviant group member*
- \*Grob, J.D.M. (2009), *Dial E for Emotion : context and consequences of emotion regulation*
- \*Joly, J.F. (2008), *People on Our Minds: When Humanized Contexts Activate Social Norms*
- Kamans, E. (2010), *When the weak hit back: Studies on the role of power in intergroup conflict*
- Ko, S.J. (2007), *What the Voice Reveals*
- Lammers, J. (2008), *Toward a more Social Social Psychology of Power*
- \*Maringer, M. (2007), *Feeling one thing, seeing another: Emotion comparison effects in person judgments*
- Oldenhuis, H.K.E. (2007), *I know what they think about us: Metaperceptions and intergroup relations*
- \*Renkema, L.J. (2009), *Facing death together. Understanding the consequences of mortality threats*
- \*Schwinghammer, S.A. (2006), *The self in social comparison*
- \*Trampe, D. (2007), *Social Influence: Comparison, Construal, and Persuasion Processes \**
- Van den Bos, A. (2008), *Why we stereotype influences how we stereotype*
- Van der Velde, S.W. (2009), *Imitation of Emotion: How meaning affects the link between imitation and liking*

## The Noort Committee has established fraud in the following publications:

- Gordijn, E.H., & Stapel, D.A. (2006). Behavioural effects of automatic interpersonal versus intergroup social comparison. *British Journal of Social Psychology*, 45, 717-729.
  - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
  - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
  - Strong indications that the dataset in studies 2 and 3 were manipulated and faked due to duplicating cases.
  - Mr. Stapel did declare that this article is fraudulent.
- Joly, J.F., Stapel, D.A., & Lindenberg, S.M. (2008). Silence and table manners: When environments activate norms. *Personality and Social Psychology Bulletin*, 34, 1047-1056. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a “Breedtestrategie II” grant from the University of Groningen awarded to D. A Stapel)
  - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
  - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author who collected a small part of the data herself.
  - Strong indications from studies in this article that the datasets were manipulated and faked: many duplicated cases.
  - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Joly, J.F., & Stapel, D.A. (2009). Staff, miter, book, share: How attributes of Saint Nicholas induce normative behavior. *European Journal of Social Psychology*, 39, 145-153. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a “Breedtestrategie II” grant from the University of Groningen awarded to D. A Stapel and S.M. Lindenberg)
  - Datasets and additional relevant information on procedure, analysis, and results were provided by

the first author.

- Data collection was partly managed by Mr. Stapel and results were handed over for further analysis to the first author who collected the other part of the data herself.
  - Strong indications from studies in this article that the datasets were manipulated and faked.
  - Mr. Stapel did declare that this article is fraudulent.
- Lindenberg, S.M., Joly, J.F., & Stapel, D.A. (2011). The norm-activating power of celebrity: The dynamics of success and influence. *Social Psychology Quarterly*, 74, 98-120. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
    - Datasets and additional relevant information on procedure, analysis, and results were provided by J.F. Joly.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to J.F. Joly.
    - Strong indications from studies in this article that the datasets were manipulated and faked in order to construct more datasets.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Maringer, M., & Stapel, D.A. (2007). Unfinished business: How completeness affects the impact of emotional states and emotion concepts on social judgments. *Journal of Experimental Social Psychology*, 43, 712-718. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel)
    - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
    - Strong indications that the datasets were manipulated and faked due to duplicating cases.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Maringer, M., & Stapel, D.A. (2009). Correction or comparison? The effects of prime awareness on social judgments. *European Journal of Social Psychology*, 39, 719-733. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)
    - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
    - Strong indications that the datasets were manipulated and faked due to duplicating cases.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Marx, D.M., & Stapel, D.A. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology*, 91, 243-254. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel)
    - Datasets and additional relevant information on procedure, analysis, and results were provided by

the first author.

- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
  - Strong indications that the dataset in studies 1 and 3 were manipulated and faked due to duplicating cases. Replications of part of the analyses did not provide identical results as reported in the article.
  - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Marx, D.M., & Stapel, D.A. (2006). It's all in the timing: Measuring emotional reactions to stereotype threat before and after taking a test. *European Journal of Social Psychology*, 36, 687-698. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel)
    - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
    - Strong indications that the dataset was manipulated and faked due to duplicating cases for specific cells of the design. Replications of part of the analyses did not provide identical results as reported in the article.
    - = Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Renkema, L.J., Stapel, D.A., Maringer, M., & Van Yperen, N.W. (2008). Terror management and stereotyping: Why do people stereotype when mortality is salient? *Personality and Social Psychology Bulletin*, 34, 553-564.
    - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
    - Strong indications that the datasets in studies 1 and 2 were manipulated and faked due to duplicating cases.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Schwinghammer, S.A., & Stapel, D.A. (2006). The effects of different types of self-activation on social comparison orientation. *Social Cognition*, 24, 703-722. (Research supported by grant 007413212 and by a Pioneer grant from NWO awarded to D.A. Stapel)
    - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
    - The first author managed the data collection, but Mr. Stapel had access to the collected data at a certain point in time after which he handed over the data for further analysis to the first author.
    - Inconsistencies were observed in data and not all results could be replicated.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Schwinghammer, S.A., & Stapel, D.A. (2011). Measure by measure: When implicit and explicit social comparison effects differ. *Self and Identity*, 10, 166-173. (Research supported by grant 007413212 and by a Pioneer grant from NWO awarded to D.A. Stapel)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
  - The first author managed the data collection, but Mr. Stapel had access to the collected data at a certain point in time after which he handed over the data for further analysis to the first author.
  - Strong doubts about the origin of the data. Two similar datasets available (95% of the cases are the same). One dataset has suspicious duplicate cases but provides the reported results in the paper. However, it is not clear which dataset is the result of actual data collection.
- Stapel, D.A., & Blanton, H. (2004). From seeing to being: Subliminal social comparisons affect implicit and explicit self-evaluations. *Journal of Personality and Social Psychology*, 87, 468-481. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel)
    - Dataset 6 of this study was provided by W.J. Post
    - Data collection and analysis were managed by Mr. Stapel.
    - Incorrectly reported p-value.
    - Inconsistencies could be observed in data and analysis, and replications of analysis did not provide identical results as reported in the article.
    - Mr. Stapel did declare that this article is fraudulent.
- Stapel, D.A., & Marx, D.M. (2006). Hardly thinking about others: On target closeness and cognitive busyness in social comparison effects. *Journal of Experimental Social Psychology*, 42, 397-405. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel)
    - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
    - Strong indications that the dataset in studies 1 and 2 were manipulated and faked due to duplicating cases.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Stapel, D.A., & Suls, J. (2004). Method matters: Effects of implicit versus explicit social comparisons on activation, behavior, and self-views. *Journal of Personality and Social Psychology*, 87, 860-875. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel)
    - Data collection and analysis were managed by Mr. Stapel.
    - Incorrectly reported number of participants or degrees of freedom for *F*-statistics.
    - Incorrectly reported p-values.
    - High evidential value in favour of the hypothesis that the underlying data are correlated, versus the hypothesis that they are independent, as they should be.
    - Mr. Stapel did declare that this article is fraudulent.
- Stapel, D.A., & Schwinghammer, S.A. (in press). Me and the why of social comparison: how self-activation determines social comparison motives. *Self and Identity*.

- Data collection was partly managed by Mr. Stapel; the data were not handed over to the second author.
  - The second author has confirmed that this paper is partly based on the data and analysis of Chapter 3 of the thesis. The dataset used in the paper differs from the one in the thesis, and the results could not be replicated using the datasets available.
  - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Stapel, D.A., & Van der Zee, K.I. (2006). The self salience model of other-to-self effects: Integrating self-enhancement, complementarity, and imitation principles. *Journal of Personality and Social Psychology*, 90, 258-271. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel)
    - Data collection and analysis were managed by Mr. Stapel.
    - Incorrectly reported p-values.
    - Quite high effect sizes for Study 1.
    - High evidential value in favour of the hypothesis that the underlying data are correlated, versus the hypothesis that they are independent, as they should be.
    - Mr. Stapel did declare that this article is fraudulent.
- Trampe, D., Stapel, D.A., & Siero, F.W. (2007). On models and vases: Body dissatisfaction and proneness to social comparison effects. *Journal of Personality and Social Psychology*, 92, 106-118. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and research supported by grant number T32 MH19728 from NWO awarded to D.A. Stapel and F.W. Siero)
    - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
    - Suspicious data handling: many duplicated cases.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Trampe, D., Stapel, D.A., Siero, F.W. (2011). The Self-Activation Effect of Advertisements: Ads can Affect Whether and How Consumers Think about the Self. *The Journal of Consumer Research*, 37, 1030-1045. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and research supported by grant number T32 MH19728 from NWO awarded to D.A. Stapel and F.W. Siero)
    - Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
    - Inconsistencies could be observed in data and analysis, and replications of analysis did not provide identical results as reported in the article.
    - Mr. Stapel did declare that this article is fraudulent.
- Trampe, D., Stapel, D.A., Siero, F.W., & Mulder, M. (2010). Beauty as a tool: The effect of model attractiveness, product relevance, and elaboration likelihood on advertising effectiveness. *Psychology &*

*Marketing*, 27, 1101-1121. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and research supported by grant number T32 MH19728 from NWO awarded to D.A. Stapel and F.W. Siero)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
  - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
  - Suspicious data handling and analyses: many duplicated cases.
  - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Van den Bos, A., & Stapel, D.A. (2008). The impact of comprehension versus self-enhancement goals on group perception. *Social Psychology*, 39, 222-230.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
    - There are strong indications that the data in this paper were manipulated and faked: duplicate cases, and strange (inconsistent) correlation structures between variables within different conditions.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Van den Bos, A., & Stapel, D.A. (2009). Why people stereotype affects how they stereotype: The differential influence of comprehension goals and self-enhancement goals on stereotyping. *Personality and Social Psychology Bulletin*, 35, 101-113. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from 'Tiber'),
    - The Committee did not receive the data of studies #2 and #3.
    - Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
    - There are strong indications that the available data-set for Study 4 was manipulated and faked: duplicated cases, highly implausible and unlikely skewed distributions in combination with large effect sizes.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

## **Dissertations**

- Grob, J.D.M. (2009). *Dial E for Emotion : context and consequences of emotion regulation*. Datasets were provided from the empirical chapters of the dissertation. Also, some additional relevant information, like detailed descriptions of procedure and data analysis, was available.

Chapter 5 *Emotions under pressure: Facial suppression of disgust causes rebound and leakage effects*

- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
  - There are strong indications that the available dataset for Study 5.1 was based on data of chapter 4, since there are many cases in the data of chapter 5 (study 5.1) which are identical to the cases of the data in chapter 4.
  - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this dissertation.
- Joly, J.F. (2008). *People on Our Minds: When Humanized Contexts Activate Social Norms*. Datasets were provided from the empirical chapters of the dissertation.

Chapter 2, see article

Lindenberg, S.M., Joly, J.F., & Stapel, D.A. (2011). The norm-activating power of celebrity: The dynamics of success and influence. *Social Psychology Quarterly*, 74, 98-120. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)

- Datasets and additional relevant information on procedure, analysis, and results were provided by J.F. Joly.
- Data collection was managed by Mr. Stapel and results were handed over for further analysis to J.F. Joly.
- Strong indications from studies in this article that the datasets were manipulated and faked in order to construct more datasets.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

Chapter 3, see article:

Joly, J.F., & Stapel, D.A. (2009). Staff, miter, book, share: How attributes of Saint Nicholas induce normative behavior. *European Journal of Social Psychology*, 39(1), 145-153. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a "Breedtestrategie II" grant from the University of Groningen awarded to D.A. Stapel and S.M. Lindenberg)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
- Data collection was partly managed by Mr. Stapel and results were handed over for further analysis to the first author who collected the other part of the data herself.
- Strong indications from studies in this article that the datasets were manipulated and faked.
- Mr. Stapel did declare that this article is fraudulent.

Chapter 5, see article:

Joly, J.F., Stapel, D.A., & Lindenberg, S.M. (2008). Silence and table manners: When environments activate norms. *Personality and Social Psychology Bulletin*, 34, 1047-1056. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a "Breedtestrategie II" grant from the University of Groningen awarded to D.A. Stapel)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author who collected a small part of the data herself.
- Strong indications from studies in this article that the datasets were manipulated and faked: many duplicated cases.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

- Maringer, M. (2007). *Feeling one thing, seeing another: Emotion comparison effects in person judgments*.

Datasets were provided from the empirical chapters of the dissertation. Also, some additional relevant information, like detailed descriptions of procedure and data analysis, was available.

Chapter 2, see article:

Maringer, M., & Stapel, D.A. (2007). Unfinished business: How completeness affects the impact of

emotional states and emotion concepts on social judgments. *Journal of Experimental Social Psychology*, 43, 712-718. This article is partly based on Chapter 2. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from the Heymans Institute of the University of Groningen awarded to D.A. Stapel)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
- Strong indications that the datasets were manipulated and faked due to duplicating cases.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

Chapter 3 *The question determines the answer: Affective information effects on judgments of likeability and judgments of happiness*

- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
- The first author has strong doubts about the origin of the data.
- Indications that the datasets are faked.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this dissertation.

Chapter 4, see article

Maringer, M., & Stapel, D.A. (2009). Correction or comparison? The effects of prime awareness on social judgments. *European Journal of Social Psychology*, 39, 719-733. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
- Strong indications that the datasets were manipulated and faked due to duplicating cases.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

Chapter 5 *Timing of priming: Affective information effects on feeling-based and knowledge-based social judgments*

- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
- The first author has strong doubts about the origin of the data.
- Indications that the dataset of this chapter is faked.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this dissertation.

- Renkema, L.J. (2009). *Facing death together. Understanding the consequences of mortality threats*. Datasets were provided from the empirical chapters of the dissertation. Also, some additional relevant information, like detailed descriptions of procedure and data analysis, was available.

Chapter 5, see article:

Renkema, L.J., Stapel, D.A., Maringer, M., & Van Yperen, N.W. (2008). Terror management and stereotyping: Why do people stereotype when mortality is salient? *Personality and Social Psychology Bulletin*, 34, 553-564.

- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
  - Strong indications that the datasets in studies 1 and 2 were manipulated and faked due to duplicating cases.
  - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this dissertation.
- Schwinghammer S.A. (2006). *The self in social comparison*. Datasets were provided from the empirical chapters of the dissertation. Also, some additional relevant information, like detailed descriptions of procedure and data analysis, was available.

Chapter 2, see article:

Schwinghammer, S.A., & Stapel, D.A. (2006). The effects of different types of self-activation on social comparison orientation. *Social Cognition*, 24, 703-722. (Research supported by grant 007413212 and by a Pioneer grant from NWO awarded to D.A. Stapel)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
- The first author managed the data collection, but Mr. Stapel had access to the collected data at a certain point in time after which he handed over the data for further analysis to the first author.
- Inconsistencies were observed in data and not all results could be replicated.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

Chapter 3 *Self-Activation and Social Comparison Motives*, also see article:

Stapel, D.A., & Schwinghammer, S.A. (in press). Me and the why of social comparison: how self-activation determines social comparison motives. *Self and Identity*. Which is partly based on chapter 3.

- Data collection was partly managed by Mr. Stapel; the data were not handed over to the second author.
- The second author has confirmed that this paper is partly based on the data and analysis of - Chapter 3 of the thesis. The dataset used in the paper differs from the one in the thesis, and the results could not be replicated using the datasets available.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

Chapter 5, see article:

Schwinghammer, S.A., & Stapel, D.A. (2011). Measure by measure: When implicit and explicit social comparison effects differ. *Self and Identity*, 10, 166-173. (Research supported by grant 007413212 and by a Pioneer grant from NWO awarded to D.A. Stapel)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
  - The first author managed the data collection, but Stapel had access to the collected data at a certain point in time after which he handed over the data for further analysis to the first author.
  - Strong doubts about the origin of the data. Two similar datasets available (95% of the cases are the same). One dataset has suspicious duplicate cases but provides the reported results in the paper. However, it is not clear which dataset is the result of actual data collection.
- Trampe, D. (2007). *Social Influence: Comparison, Construal, and Persuasion Processes*. Datasets were provided from the empirical chapters of the dissertation.

Chapter 2, see article:

Trampe, D., Stapel, D.A., & Siero, F.W. (2007). On models and vases: Body dissatisfaction and proneness to social comparison effects. *Journal of Personality and Social Psychology*, 92, 106-118. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and research supported by grant number T32 MH19728 from NWO awarded to D.A. Stapel and F.W. Siero)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
- Suspicious data handling: many duplicated cases.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

Chapter 3, see article:

Trampe, D., Stapel, D.A., Siero, F.W. (2011). The Self-Activation Effect of Advertisements: Ads can Affect Whether and How Consumers Think about the Self. *The Journal of Consumer Research*, 37, 1030-1045. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and research supported by grant number T32 MH19728 from NWO awarded to D.A. Stapel and F.W. Siero)

- Datasets and additional relevant information on procedure, analysis, and results were provided by the first author.
- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
- Inconsistencies could be observed in data and analysis, and replications of analysis did not provide identical results as reported in the article.
- Mr. Stapel did declare that this article is fraudulent.

Chapter 4, see article:

Trampe, D., Stapel, D.A., Siero, F.W., & Mulder, M. (2010). Beauty as a tool: The effect of model attractiveness, product relevance, and elaboration likelihood on advertising effectiveness. *Psychology & Marketing*, 27, 1101-1121. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and research supported by grant number T32 MH19728 from NWO awarded to D.A. Stapel and F.W. Siero)

- Datasets and additional relevant information on procedure, analysis, and results were provided

by the first author.

- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
- Suspicious data handling and analyses: many duplicated cases.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

- Van den Bos, A. (2008). *Why we stereotype influences how we stereotype*.

Chapter 2, see article:

Van den Bos, A., & Stapel, D.A. (2009). Why people stereotype affects how they stereotype: The differential influence of comprehension goals and self-enhancement goals on stereotyping. *Personality and Social Psychology Bulletin*, 35, 101-113. (Research supported by a Pioneer grant from NWO awarded to D.A. Stapel and a research grant from 'Tiber'),

- The Committee did not receive the data of studies #2 and #3.
- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
- There are strong indications that the available data-set for Study 4 was manipulated and faked: duplicated cases, highly implausible and unlikely skewed distributions in combination with large effect sizes.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

Chapter 3, see article:

Van den Bos, A., & Stapel, D.A. (2008). The impact of comprehension versus self-enhancement goals on group perception. *Social Psychology*, 39, 222-230

- Data collection was managed by Mr. Stapel and results were handed over for further analysis to the first author.
- There are strong indications that the data in this paper were manipulated and faked: duplicate cases, and strange (inconsistent) correlation structures between variables within different conditions.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

Chapter 4 *Comprehension and stereotyping: goal or mode*

- Strong indications from studies in chapter 4 that the available datasets were manipulated and faked: many duplicated cases, many variables have the same data, whereas their meanings are quite different, and highly implausible and unlikely skewed distributions.
- Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this dissertation.

### ***The Noort Committee has found evidence of fraud in the following publications:***

- Stapel, D.A., & Koomen, W. (2001). When we wonder what it all means: Interpretation goals facilitate accessibility and stereotyping effects. *Personality and Social Psychology Bulletin*, 27, 915-929. Data collection and analysis were managed by Mr. Stapel.
  - Incorrectly reported p-values.

- Incorrectly reported number of participants or degrees of freedom for  $F$ -statistics.
  - High evidential value in favour of the hypothesis that the underlying data are correlated versus the hypothesis that they are independent as they should be.
  - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Stapel, D.A., Koomen, W., & Ruys, K. (2002). The effects of diffuse and distinct affect. *Journal of Personality and Social Psychology*, 83, 60-74.  
Data collection and analysis were managed by Mr. Stapel.
    - Cell averages are impossible given the number of participants and a reasonably balanced design.
    - Incorrectly reported number of participants or degrees of freedom for  $F$ -statistics.
    - Incorrectly reported error variances.
    - High evidential value in favour of the hypothesis that the underlying data are correlated versus the hypothesis that they are independent as they should be.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.
- Stapel, D.A., & Koomen, W. (2006). The flexible unconscious: Investigating the judgmental impact of varieties of unaware perception. *Journal of Experimental Social Psychology*, 42, 112-119.  
Data collection and analysis were managed by Mr. Stapel.
    - Cell averages are impossible given the number of participants and a reasonably balanced design.
    - Incorrectly reported error variances.
    - High evidential value in favour of the hypothesis that the underlying data are correlated versus the hypothesis that they are independent as they should be.
    - Mr. Stapel did declare that he is not sure about the character (fraudulent or not) of this article.

***The following book chapters and articles are (partly) based on findings of articles, in which the Committees have found evidence of fraud:***

- Stapel, D.A. (2000). Moving from fads and fashions to integration: Illustrations from knowledge accessibility research. *European Bulletin of Social Psychology*, 12, 4-27.
- Stapel, D.A. (2002). Wat je ziet ben je zelf: Contexteffecten op oordelen en gedrag. *Nederlandsch tijdschrift voor psychologie*, 57, 31-41.
- Stapel, D.A., & Koomen, W. (2001). Let's not forget the past when we go to the future: On our knowledge of knowledge accessibility effects. In G.B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 229-246). Mahwah NJ: Lawrence Erlbaum.
- Stapel, D.A. (2003). Making sense of hot cognition: Why and when description influences our feelings and judgments. In J.P. Forgas, K.D. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 227-250). Cambridge: Cambridge University Press. (Sydney Symposium of Social Psychology series, 5).



# Appendix 6 List of publications examined by the Drenth Committee including findings

## List of publications examined by the Drenth Committee

# The Drenth Committee has found evidence of fraud in this paper, see also the list ‘evidence of fraud’

- Dijksterhuis, A., Spears, R., Postmes, T., Stapel, D.A., Koomen, W., Van Knippenberg, A., & Scheepers, D. (1998). Seeing one thing and doing another: Contrast effects in automatic behavior. *Journal of Personality and Social Psychology*, **75**, 862–871.
- Koomen, W., Stapel, D.A., Jansen, S.S., & in 't Veld, K.H.R. (1998). Priming and timing: A test of two perspectives. *European Journal of Social Psychology*, **28**, 681–686.
- Martin, L.L., & Stapel, D.A. (1998). Metacognition and correction processes in social judgment: Are people naive dogmatists or naive empiricists? In V. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and Social Dimensions* (pp. 228-247). London: Sage.
- Stapel, D.A. (1993). Informatie door introspectie: Het gebruik van cognitieve responsanalyse in marktonderzoek. *Onderzoek*, **11**, 20–25.
- Stapel, D.A. (1994). Marketing en mentaal boekhouden. *Tijdschrift voor Marketing*, **28**, 12–17.
- Stapel, D.A. (1995). Het verschil tussen zeggen en doen: De discrepantie tussen attitude en gedrag in onderzoek naar brievenbusreclame. *Onderzoek*, **13**, 11–14.
- Stapel, D.A. (1997). The construction of social judgments: A constructive perspective for behavioral decision research? *Journal of Behavioral Decision Making*, **10**, 355–357.
- Stapel, D.A., & Hendrickx, M.T.M. (1994). Marketingbeleid universiteiten te eenzijdig. *Tijdschrift voor Marketing*, **28**, 25–29.
- # Stapel, D.A., & Koomen, W. (1996). Differential consequences of trait inferences: A direct test of the trait–referent hypothesis. *European Journal of Social Psychology*, **26**, 827–837.
- # Stapel, D.A., & Koomen, W. (1997). Using primed exemplars during impression formation: Interpretation or comparison? *European Journal of Social Psychology*, **27**, 357–367.
- Stapel, D.A., & Koomen, W. (1997). Social categorization and perceptual judgment of size: When perception is social. *Journal of Personality and Social Psychology*, **73**, 1177–1190.
- Stapel, D.A., & Koomen, W. (1998). When stereotype activation results in (counter)stereotypical judgments: Priming stereotype–relevant traits and exemplars. *Journal of Experimental Social Psychology*, **34**, 136–163.  
Dissertation Chapter 4.
- # Stapel, D.A., & Koomen, W. (1998). Interpretation versus reference framing: Assimilation and contrast effects in the organizational domain. *Organizational Behavior and Human Decision Making Processes*, **76**, 132–148.
- # Stapel, D.A., & Koomen, W. (1999). Correction processes in person judgments: The role of timing. *European Journal of Social Psychology*, **29**, 131–138.
- Stapel, D.A., & Koomen, W. (2000). How far do we go beyond the information given? The impact of knowledge activation on interpretation and inference. *Journal of Personality and Social Psychology*, **78**, 19–37.  
Dissertation Chapter 1.

- Stapel, D.A., Koomen, W., & Spears, R. (1999). Framed and misfortuned: Self-categorization and the whiff of scandal. *European Journal of Social Psychology*, **29**, 397–402.
- # Stapel, D.A., Koomen, W., & Van der Pligt, J. (1996). The referents of trait inferences: The impact of trait concepts versus actor–trait links on subsequent judgments. *Journal of Personality and Social Psychology*, **70**, 437–450.  
Dissertation Chapter 3.
- Stapel, D.A., Koomen, W., & Van der Pligt, J. (1997). Categories of category accessibility: The impact of trait versus exemplar priming on person judgments. *Journal of Experimental Social Psychology*, **33**, 47-76.  
Dissertation Chapter 2.
- Stapel, D.A., Koomen, W., & Velthuisen, A.S. (1998). Assimilation or contrast?: Comparison relevance, distinctness, and the impact of accessible information on consumer judgments. *Journal of Consumer Psychology*, **7**, 1–24.
- Stapel, D.A., Koomen, W., & Velthuisen, A.S. (1997). The context of advertising: An analysis of assimilation and contrast effects. In G. Antonides, W.F. van Raay, & S. Maital (Eds.), *Advances in economic psychology* (pp. 81-96). Chichester: Wiley.
- Stapel, D.A., Koomen, W., & Zeelenberg, M. (1998). The impact of accuracy motivation on interpretation, comparison and correction processes: Accuracy motivation x knowledge accessibility effects. *Journal of Personality and Social Psychology*, **74**, 878–893.  
Dissertation Chapter 6.
- Stapel, D.A., Martin, L.L., & Schwarz, N. (1998). The smell of bias: What instigates correction processes in social judgments? *Personality and social psychology bulletin*, **24**, 797–806.
- Stapel, D.A., Reicher, S.D., & Spears, R. (1994). Social identity, availability and the perception of risk. *Social Cognition*, **12**, 1–17.
- Stapel, D.A., Reicher, S.D., & Spears, R. (1995). Contextual determinants of strategic choice: Some moderators of the availability bias. *European Journal of Social Psychology*, **25**, 141–158.
- Stapel, D.A., & Schwarz, N. (1998). The Republican who did not want to become President: Colin Powell’s impact on evaluations of the Republican party and Bob Dole. *Personality and social psychology bulletin*, **24**, 690–699.
- # Stapel, D.A., & Schwarz, N. (1998). Similarities and differences between the impact of traits and expectancies: What matters is whether the target stimulus is ambiguous or mixed. *Journal of Experimental Social Psychology*, **34**, 227-245.  
Dissertation Chapter 5.
- Stapel, D.A., & Spears, R. (1995). Reflected or deflected glory? The impact of analogies on the perception of politicians and their policies. *Politics, Groups and the Individual*, **5**, 21–30.
- # Stapel, D.A., & Spears, R. (1996). Event accessibility and context effects in causal inference: Judgment of a different order. *Personality and Social Psychology Bulletin*, **22**, 979–992.
- Stapel, D.A., & Spears, R. (1996). Guilty by disassociation (and innocent by association): The impact of relevant and irrelevant analogies on political judgments. *Political Behavior*, **18**, 289–309.
- Stapel, D.A., & Velthuisen, A.S. (1996). “Just as if it happened to me”: The impact of vivid and self-relevant information on risk judgments. *Journal of Social & Clinical Psychology*, **15**, 102–119.
- Stapel, D.A., & Velthuisen, A.S. (1994). Maar wat heeft dat met mij te maken? Persoonlijke en onpersoonlijke impact van krant artikelen op risico-oordelen. *Massacommunicatie*, **23**, 214–227.
- Stapel, D.A., & Winkielman, P. (1998). Assimilation and contrast as a function of context–target

similarity, distinctness, and dimensional relevance. *Personality and Social Psychology Bulletin*, **24**, 634–646.

### **The Drenth Committee has found evidence of fraud in the publications mentioned below.**

*In Mr. Stapel's memory there is no question of fraudulent behaviour during his Amsterdam period and he does not agree with these findings of the Committee.*

- Stapel, D.A., & Koomen, W. (1996). Differential consequences of trait inferences: A direct test of the trait-referent hypothesis. *European Journal of Social Psychology*, **26**, 827–837.
  - Moderately high evidential value in favour of the hypothesis that the underlying data are correlated, versus the hypothesis that they are independent, as they should be.
- Stapel, D.A., & Koomen, W. (1997). Using primed exemplars during impression formation: Interpretation or comparison? *European Journal of Social Psychology*, **27**, 357–367.
  - Incorrectly reported p-values.
  - Moderately high evidential value in favour of the hypothesis that the underlying data are correlated, versus the hypothesis that they are independent, as they should be.
- Stapel, D.A., & Koomen, W. (1998). Interpretation versus reference framing: Assimilation and contrast effects in the organizational domain. *Organizational Behavior and Human Decision Making Processes*, **76**, 132–148.
  - Impossible mean age of undergraduates (18 years).
  - Cell averages are impossible, given the number of participants and a balanced design.
  - Incorrectly reported number of participants or degrees of freedom for *F*- and *t*-statistics.
  - Incorrectly reported p-values.
- Stapel, D.A., & Koomen, W. (1999). Correction processes in person judgments: The role of timing. *European Journal of Social Psychology*, **29**, 131–138.
  - \_ Incorrectly reported number of participants or degrees of freedom for *F*-statistics.
  - \_ Incorrectly reported p-values.
  - \_ High evidential value in favour of the hypothesis that the underlying data are correlated, versus the hypothesis that they are independent, as they should be.
- Stapel, D.A., Koomen, W., & Van der Pligt, J. (1996). The referents of trait inferences: The impact of trait concepts versus actor-trait links on subsequent judgments. *Journal of Personality and Social Psychology*, **70**, 437–450.

Dissertation Chapter 3.

  - Incorrectly reported error variances.
  - Quite high effect sizes for Study 2.
  - High evidential value in favour of the hypothesis that the underlying data are correlated, versus the hypothesis that they are independent, as they should be.

- Stapel, D.A., & Schwarz, N. (1998). Similarities and differences between the impact of traits and expectancies: What matters is whether the target stimulus is ambiguous or mixed. *Journal of Experimental Social Psychology*, 34, 227–245.  
Dissertation Chapter 5.
  - Incorrectly reported p–value.
  - High evidential value in favour of the hypothesis that the underlying data are correlated, versus the hypothesis that they are independent, as they should be.
  
- Stapel, D.A., & Spears, R. (1996). Event accessibility and context effects in causal inference: Judgment of a different order. *Personality and Social Psychology Bulletin*, 22, 979–992.
  - Moderately high effect sizes for Study 2.
  - High evidential value in favour of the hypothesis that the underlying data are correlated, versus the hypothesis that they are independent, as they should be.

## Appendix 7 Authority to withdraw a doctorate

Under its terms of reference regarding legal aspects, the Noort Committee has deliberated on the legality of withdrawing an awarded doctorate. Decisions to award an academic degree quickly obtain formal legal force, which is to say that they can no longer be challenged through judicial process. Nonetheless it is plausible under certain conditions for a competent administrative body to have the authority to withdraw an awarded academic degree, if the award relied on fraud on the part of the person receiving the doctorate. This constitutes impairment of material legal force. The administrative body must then provide compelling evidence for fraud.

The Higher Education and Research Act (WHW) provides no explicit authority to withdraw awarded degrees. A sustainable argument is that the authority to award implies an authority to withdraw. Assuming the existence of an implicit authority, it may be argued that the grounds for withdrawal are the same as those on which a degree was awarded. The award of a doctorate relies on the PhD student demonstrating competence to perform independent scientific research (to be assessed with reference to the problem definition, method, response to the problem definition, etc.). This is the background to the conclusion given on page 7 of the interim report, that the fraud committed by Mr Stapel should have no repercussions on the awarded degree of any of his PhD students.

Where fraud was committed by a PhD student in person, that student will not have demonstrated the competence to perform independent scientific research. In cases of this kind, it may be argued, the implicit authority may come into play in withdrawing the degree from the doctor. It is relevant that an assessment committee was engaged, in that it has a bearing on the onus of proof. Also relevant is the educational nature of the PhD track. The final aspects are legal certainty and lapse of time. These aspects create additional obstacles for the possible withdrawal of a doctorate.

The above is concerned with the possibility of withdrawing an awarded doctorate because of fraud in the dissertation itself. A different question is whether conduct *after* a doctorate has been awarded, which is therefore distinct from the grounds for awarding the doctorate based on the dissertation and its defence, could give rise to the authority to withdraw an awarded degree and, if so, the kind of conduct that might have this effect. The question is whether only (self-evidently serious) infringements of scientific integrity would qualify, or also other (serious forms of) reprehensible behaviour. The most plausible position is that the current Higher Education and Research Act, unlike the German (Hochschul-) law, provides no authority, implicit or otherwise, to withdraw a doctorate on any grounds other than those on which it was awarded.

In the light of the above the Committee considers that the options for withdrawing an awarded doctorate require further attention, primarily in an academic connection and additionally on ministerial level. Opinions differ on the legal possibility of withdrawing a doctorate. Amendment of the legislation would appear appropriate if only to eliminate the existing lack of clarity, and would therefore be in the interest of legal certainty. Regarding the withdrawal of a doctorate in connection with conduct *after* the degree has been awarded, the Higher Education and Research Act should be augmented with a specific authority to withdraw (where the Committee's preference would be to limit the authority to (self-evidently serious)

infringements of scientific integrity). In all the above the Committee finds it desirable for more emphatic and precise attention than hitherto to be given to both the rights and the duties attached to a doctorate.

## Appendix 8 Literature references

European Code of Conduct for Research Integrity (ESF-ALLEA 2011)

ISI Declaration on Professional Ethics <http://www.isi-web.org/about-isi/professional-ethics>

Klaassen, C.A.J. (2012). Evidential Value in ANOVA Results in Favor of Fabrication, preprint

Meyer, G., Finn, S.E., Eyde, L.D., Kay, G.G., Moreland, K.L., Dies, R.R., Eisman, E.J., Kubiszyn, T.W. & Reed, G.M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128-165

Nederlandse Gedragscode Wetenschapsbeoefening (VSNU 2004, herziening 2012)

Notitie Wetenschappelijke Integriteit, Over normen van wetenschappelijk onderzoek en een Landelijk Orgaan voor Wetenschappelijke Integriteit (LOWI) (2001), KNAW

Reproducibility project. <http://openscienceframework.org/project/EZcUj/wiki/home>

Richard, F.D., Bond Jr., C.F., and Stokes-Zoota, J.J. (2003). One Hundred Years of Social Psychology Quantitatively Described. *Review of General Psychology*, 7, 331-63

Simonsohn, U., (2012) Just Post It: The Lesson from Two Cases of Fabricated Data Detected by Statistics Alone , posted on <http://www.ssrn.com/>  
*Date posted:* July 22, 2012  
*Last revised:* November 13, 2012

Special Section on Replicability in Psychological Science: A Crisis of Confidence? (2012), *Perspectives of Psychological Science*, 7

Stapel, D.A. (2000). Moving from fads and fashions to integration: Illustrations from knowledge accessibility research. *European Bulletin of Social Psychology*, 12, 4-27

Stapel, D.A. (2001). De koningin is aan zet: waarom de alledaagse precisie van de sociale psychologie haar majesteitelijk maakt, *Nederlands Tijdschrift voor de Psychologie* 56, 286b

Van Lange, P.A.M., Buunk, A.P., Ellemers, N., & Wigboldus, D.H.J. (2012), Sharpening scientific policy after Stapel. <http://www.sociale-psychologie.nl>

Wicherts, J.M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61, 726-728

Zorgvuldig en integer omgaan met wetenschappelijke onderzoeksgegevens, Advies van de KNAW Commissie Onderzoeksgegevens, Amsterdam 2012

