

Falende wetenschap: De frauduleuze onderzoekspraktijken van sociaal-psycholoog Diederik Stapel

Commissie Levelt

Commissie Noort

Commissie Drenth

28 november 2012

Inhoud

Voorwoord	5
1 Inleiding en aanleiding	7
1.1 Opdracht Commissie Levelt	7
1.2 Opdracht Commissie Noort	7
1.3 Opdracht Commissie Drenth	8
1.4 Statistische Ondersteuning	9
1.5 Aanleiding tot het onderzoek	9
1.6 Karakter rapportage	10
2 Werkwijze Commissies	11
2.1 Samenwerking Commissies en verdeling werkzaamheden	11
2.2 Formele status commissies	11
2.3 Procedure	12
2.3.1 <i>Informatiebronnen</i>	14
2.3.2 <i>Allesomvattend onderzoek</i>	15
2.3.3 <i>Conclusie per publicatie</i>	15
2.3.4 <i>Hoor en wederhoor: algemeen</i>	16
2.3.5 <i>Hoor en wederhoor: de heer Stapel</i>	16
2.4 Vaststelling fraude	17
2.4.1 <i>Afbakening begrip fraude</i>	17
2.4.2 <i>Bewijs van en aanwijzingen voor fraude</i>	17
2.5 Concrete werkwijze: toelichting en voorbeelden	18
2.5.1 <i>Inrichting en uitvoering experiment</i>	18
2.5.2 <i>Merkwaardigheden in de datamatrix</i>	19
2.5.3 <i>(Her)Analyse van de data</i>	20
2.5.4 <i>Publicaties</i>	20
3 Bevindingen over de aard en omvang van de fraude	25
3.1 Omvang van de fraude	25
3.2 Aard van de fraude	31
3.3 Meldingen van ernstige inbreuk op wetenschappelijke integriteit	32
3.4 Coauteurs	33
3.5 Dissertaties Tilburg	33
3.6 Dissertaties Groningen	34
3.7 Gevolgen van de fraude	34
4 Onderzoekscultuur: werkwijze en onderzoeksomgeving	37
4.1 Inleiding	37
4.2 Algemeen	38
4.3 Werkwijze en Onderzoeksomgeving: Universiteit van Amsterdam	38

4.4	Werkwijze en Onderzoeksomgeving: Rijksuniversiteit Groningen	39
4.5	Werkwijze en Onderzoeksomgeving: Tilburg University	43
5	De Onderzoekscultuur: gebrekkige wetenschapsbeoefening	47
5.1	Inleiding	47
5.2	Generaliseerbaarheid van de bevindingen van lokale naar nationale en internationale cultuur	47
5.3	Verificatiebias en ontbrekende replicaties	48
5.4	Onvolledige of foutieve informatie over de gevolgde onderzoeksprocedures	51
5.5	Statistische onvolkomenheden	52
5.6	Falende wetenschappelijke kritiek	53
6	Aanbevelingen	55
Bijlage 1	Overzicht Commissieleden en Statistici	61
Bijlage 2	Vragenlijst Commissie Noort	63
Bijlage 3	Gesprekspartners Commissies	67
Bijlage 4	Onderzochte publicaties Commissie Levelt inclusief bevindingen	71
Bijlage 5	Onderzochte publicaties Commissie Noort inclusief bevindingen	83
Bijlage 6	Onderzochte publicaties Commissie Drenth inclusief bevindingen	99
Bijlage 7	Bevoegdheid tot intrekking van de doctorsgraad	103
Bijlage 8	Literatuurverwijzingen	105

Voorwoord

De wetenschappelijke fraude van Diederik Stapel, die begin september 2011 in Tilburg aan het licht kwam, heeft in binnen- en buitenland een schok doen gaan door de academische wereld. Immers, de basis voor alle wetenschappelijke samenwerking is onderling vertrouwen. Wanneer dat vertrouwen ernstig wordt geschonden, zoals in het geval van de heer Stapel, knaagt dat aan de fundamenten van het wetenschappelijk bedrijf.

Is het spreekwoordelijke zelfreinigende karakter van de wetenschap ook opgewassen tegen zo'n ernstige schending van wetenschappelijke integriteit? De Rector Magnificus van Tilburg University besloot direct volstrekte openheid van zaken te geven. stelde een onderzoekscommissie in, de Commissie Levelt, die aard en omvang van de fraude moest vaststellen en de onderzoekscultuur in kaart moest brengen die deze langdurige fraude had mogelijk gemaakt. Dat beeld zou niet volledig worden zonder ook de eerdere werkplekken van de heer Stapel in het onderzoek te betrekken, de Universiteit van Amsterdam, waar hij van 1993 tot 1999 verbleef en in 1997 promoveerde, en de Rijksuniversiteit Groningen, waar hij van 2000 tot 2006 hoogleraar was. In goed overleg tussen de universiteiten werden ook daar onderzoekscommissies ingesteld, de Commissies Drenth en Noort. In dit eindrapport doen de drie commissies gezamenlijk verslag van hun bevindingen.

Het was, direct na de ontdekking, urgent om de schade voor de slachtoffers van de fraude zoveel mogelijk te beperken. Dat gold met name voor de promoti en postdocs van de heer Stapel, die hun publicaties in rook zagen opgaan. In een interim-rapport, dat eind oktober 2011 verscheen, hebben de Commissies over de omvang van die schade gerapporteerd en ook geconcludeerd dat geen van deze (co)auteurs medeplichtig was aan de fraude van de heer Stapel. Het rapport gaf ook een summiere schets van de aard van de fraude en van de onderzoekscultuur waarbinnen deze zich had afgespeeld. Dit interim-rapport is direct publiek gemaakt.

Maar het grote werk moest nog beginnen, althans naar het oordeel van de Commissies. Hier was sprake van een omvangrijk, maar besmet wetenschappelijk oeuvre. In een ideale wetenschappelijke wereld diende hier het kaf van het koren te worden gescheiden. Tijdschriftpublicaties trekken soms lange sporen binnen en tussen disciplines. Het zelfreinigend karakter van de wetenschap vereist dat frauduleuze publicaties worden teruggetrokken en niet verder in de literatuur prolifereren. Bovendien meenden de Commissies op grond van eerste indrukken dat er meer mis was in de publicaties van de heer Stapel, ook als die niet frauduleus waren, iets wat in dit rapport wordt aangeduid als *sloppy science*: het niet voldoen aan normale methodologische standaards. Daarmee kwam een onderzoekscultuur in beeld, waarbinnen zulke *sloppy science*, maar ook regelrechte fraude, lang ongedetecteerd kon blijven. Ook op dit punt was zelfreiniging in de wetenschap geboden.

Het siert de drie betrokken universiteiten dat zij de Commissies de vrije hand hebben gegeven in het analyseren van alle publicaties van de heer Stapel. Daarbij werd medewerking van de coauteurs verzocht, die dat zonder uitzondering hebben verleend door onderzoeksmaterialen en data aan te leveren en in veel gevallen door ons in gesprekken te informeren over de gang van zaken bij het onderzoek. Tevens werden aan de drie universiteiten statistici ter beschikking gesteld van de Commissies om alle publicaties van de heer Stapel aan een grondige analyse te onderwerpen. Het spreekt vanzelf dat het hier om een arbeidsintensieve en kostbare operatie ging. Het is bij ons weten nooit eerder gebeurd dat het hele wetenschappelijk

oeuvre van een fraudeur op de korrel is genomen. Commissies beperken zich tot het onderzoek van artikelen die als verdacht zijn gemeld, of tot een kleine steekproef. Daarmee kan in veel gevallen inderdaad bewijs voor fraude worden geleverd. Maar er is altijd een groter belang. De wetenschappelijke literatuur moet worden gezuiverd van alles wat frauduleus is, in het bijzonder wanneer het werk van een coryfee betreft. Dat is natuurlijk in sterke mate geboden in de geneeskunde, waar frauduleuze publicaties schadelijke gevolgen kunnen hebben voor grote groepen patiënten. Maar het geldt ook voor minder toegepaste wetenschappen, zoals de sociale psychologie. Er moet geen tijd en geld meer worden verknoeid aan replicaties of meta-analyses van gefabriceerde data. Bij onderzoekers en met name studenten moet een beeld van het vak worden gecorrigeerd dat ‘te mooi is om waar te zijn’. In het onderhavige geval moet bovendien door tijdschriften, reviewers, beoordelingscommissies en graduate schools een methodologische les worden getrokken; er kan blijkbaar teveel mis gaan in het kritisch functioneren van de discipline.

De belangrijkste reden echter om volledigheid na te streven in het zuiveren van de *scientific record* is dat de wetenschap zelf een bijzondere aanspraak heeft op waarheidsvinding. Dat is een cumulatief proces, dat in de empirische wetenschap, en met name in de psychologie, wordt gekenmerkt door een ‘empirische cyclus’, een voortdurende afwisseling tussen theorievorming en empirische toetsing. Een theorie heeft een voorlopige aanspraak op waarheid/geldigheid, zolang zij niet empirisch is weerlegd. Cumulatieve evidentie kan tenslotte leiden tot consensus in de *peer community* over de geldigheid van een theorie. Dit fundamentele cumulatieve proces wordt ernstig verstoord door het ‘rondzingen’ van frauduleuze data en van methodologisch dubieuze bevindingen. Betrokken wetenschappelijke onderzoekers en instellingen hebben de plicht deze verstoring een halt toe te roepen.

Helaas worden een aantal bevindingen van de Commissies weerspiegeld in de recente open email die Nobelprijswinnaar Daniel Kahneman over de situatie in de sociale psychologie publiceerde: ‘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 ziet ‘a train wreck looming. I expect the first victims to be young people on the job market.’ Als dit rapport, en de daarin gesignaleerde misstanden, kan bijdragen aan die saneringsoperatie, dan heeft de Stapel-fraude naast alle aangerichte, vooral persoonlijke schade, wellicht nog een positief effect.

De Commissies spreken hun dank uit aan de drie Rectores Magnifici voor hun daadwerkelijke en morele ondersteuning bij hun werk. Zij danken de toegewijde statistici, die onder vaak grote druk hun geacheveerde analyses uitvoerden. Zonder de open gesprekken met coauteurs en andere betrokkenen zou de diepteanalyse in dit rapport onmogelijk zijn geweest. Die gesprekken waren, met name voor de slachtoffers van deze fraude, vaak beladen met emoties. Hun bijdragen aan ons commissiewerk verdienen speciale, dankbare vermelding. Tenslotte danken de Commissies hun voortreffelijke uitvoerend secretarissen, die gedurende meer dan een jaar het commissiewerk hebben georganiseerd en een belangrijk deel van de rapportage hebben verzorgd. De namen van allen die het werk van de Commissies in deze hoedanigheden hebben ondersteund zijn in bijlage 1 van dit eindrapport opgetekend.

Met dit eindrapport sluiten de Commissies hun werkzaamheden af.

Willem J.M. Levelt, *voorzitter coördinerende Commissie*

1 Inleiding en aanleiding

Begin september 2011 besloot het College van Bestuur van Tilburg University om prof.dr. D.A. Stapel met onmiddellijke ingang op non-actief te stellen en kort daarna te ontslaan. De heer Stapel bleek een ernstige inbreuk op de wetenschappelijke integriteit gemaakt te hebben door in zijn publicaties gebruik te maken van gefingeerde data die echter gepresenteerd werden als verkregen in empirisch onderzoek. Op vrijdag 9 september 2011 is door de Rector Magnificus van Tilburg University, prof.dr. Ph. Eijlander, de Commissie Levelt ingesteld om deze inbreuk op de wetenschappelijke integriteit te onderzoeken. Aangezien de heer Stapel eerder aangesteld is geweest bij de Universiteit van Amsterdam en de Rijksuniversiteit Groningen en het ernstige vermoeden bestond dat ook tijdens deze aanstellingen sprake is geweest van datafraude, heeft het onderzoek zich verder uitgebreid dan alleen Tilburg University. De heer Stapel bekende reeds aan de Rector van Tilburg University ook in Groningen fraude met data te hebben gepleegd. Door de Rector Magnificus van de Universiteit van Amsterdam, prof.dr. D.C. van den Boom, is de Commissie Drenth ingesteld en door het College van Bestuur van de Rijksuniversiteit Groningen bij monde van de Rector Magnificus prof.dr. E. Sterken, de Commissie Noort.

1.1 Opdracht Commissie Levelt

De Rector Magnificus van Tilburg University heeft de Commissie Levelt opdracht gegeven onderzoek te doen naar de omvang en de aard van de inbreuk op de wetenschappelijke integriteit door de heer D.A. Stapel. De opdracht is tweeledig:

1. De Commissie dient in kaart te brengen welke publicaties gebaseerd zijn op gefingeerde data of gefingeerd onderzoek en over welke periode deze fraude heeft plaatsgevonden.
2. De Commissie dient zich een idee te vormen van de werkwijze en de onderzoekscultuur die deze inbreuk mogelijk hebben gefaciliteerd, en aanbevelingen te formuleren om herhaling te voorkomen.

De Commissie bestaat uit prof.dr. W.J.M. Levelt (voorzitter), emeritus directeur van het Max Planck Instituut voor Psycholinguïstiek en voormalig president van de Koninklijke Nederlandse Akademie van Wetenschappen; prof.mr. M.S. Groenhuijsen, hoogleraar straf-, strafprocesrecht en victimologie (Tilburg University) en oprichter/directeur van het interfacultair onderzoeksinstituut Intervict; en prof.dr. J.A.P. Hagenaars, emeritus hoogleraar Methoden en Technieken van Sociaal Wetenschappelijk Onderzoek en voormalig decaan van de Faculteit Sociale Wetenschappen (Tilburg University). De Commissie wordt ondersteund door ambtelijk secretaris mw. dr.ir. S.A.M. Baert (wiskundige en beleidsmedewerker onderzoek Tilburg University).

1.2 Opdracht Commissie Noort

De officiële opdracht van het College van Bestuur van de RUG werd op woensdag 12 oktober 2011 aan de Commissie Noort verstrekt en luidde als volgt:

De Commissie doet onderzoek naar de omvang en de aard van de inbreuk op de wetenschappelijke integriteit door de heer D.A. Stapel. De opdracht is driedelig:

1. De Commissie dient in kaart te brengen welke publicaties gebaseerd zijn op gefingeerde data of gefingeerd onderzoek en over welke periode deze fraude heeft plaatsgevonden.
2. De Commissie dient zich een idee te vormen van de mogelijke (juridische) gevolgen van de inbreuk op de wetenschappelijke integriteit en aanbevelingen te formuleren om hierop te reageren, hetzij als individuele universiteit, hetzij als betrokken universiteiten gezamenlijk.
3. De Commissie dient zich een idee te vormen van de werkwijze en de onderzoekscultuur die deze inbreuk op de wetenschappelijke integriteit mogelijk hebben gefaciliteerd, en in samenwerking met de betrokken universiteiten aanbevelingen te formuleren om herhaling te voorkomen.

De Commissie Noort bestaat uit prof.dr. E. Noort (voorzitter, Rijksuniversiteit Groningen), emeritus hoogleraar Oude Testament, oud-decaan en voormalig bestuurslid van de Koninklijke Nederlandse Akademie van Wetenschappen, voorzitter Commissie Wetenschappelijke Integriteit van de RUG; prof.dr. H.E. Bröring (Rijksuniversiteit Groningen), hoogleraar Integrale Rechtsbeoefening en lid van de Commissie Wetenschappelijke Integriteit (RUG); prof.dr. J.M. Pieters (Universiteit Twente), hoogleraar Toegepaste Psychologie aan de Universiteit Twente, oud-decaan en voormalig bestuurslid en voorzitter van het Nederlands Instituut van Psychologen. De Commissie wordt ondersteund door ambtelijk secretaris mw. dr. Merlijne Jaspers, (Afd. Research & Valorisation RUG, sociaal psychologe en gezondheidswetenschapper).

1.3 Opdracht Commissie Drenth

Op woensdag 21 september 2011 heeft de Rector Magnificus van de Universiteit van Amsterdam de Commissie Drenth gevraagd onderzoek te doen naar de omvang en de aard van de inbreuk op de wetenschappelijke integriteit door de heer D.A. Stapel. De opdracht is driedelig:

1. De Commissie dient in kaart te brengen welke publicaties gebaseerd zijn op gefingeerde data of gefingeerd onderzoek en over welke periode deze fraude heeft plaatsgevonden. Het betreft de publicaties die de heer Stapel gedurende zijn Amsterdamse periode heeft gedaan. De Commissie dient eveneens in kaart te brengen of het ook het wetenschappelijk werk betreft op basis waarvan het doctoraat aan de heer Stapel is verleend. De rol van de promotor en de promotiecommissie dient, indien nodig, eveneens aan de orde gesteld te worden.
2. De Commissie dient zich een idee te vormen van de mogelijke (juridische) gevolgen en aanbevelingen te formuleren om hierop te reageren, hetzij als individuele universiteit, hetzij als betrokken universiteiten gezamenlijk.
3. De Commissie dient zich een idee te vormen van de werkwijze en de onderzoekscultuur die deze inbreuk op wetenschappelijke integriteit mogelijk hebben gefaciliteerd, en aanbevelingen te formuleren

om herhaling te voorkomen.

De Commissie Drenth bestaat uit prof.dr. P.J.D. Drenth (voorzitter, VU University Amsterdam), voormalig Rector Magnificus van de Vrije Universiteit Amsterdam en voormalig president van de Koninklijke Nederlandse Akademie van Wetenschappen; prof.dr. J.W. Zwemmer (Universiteit van Amsterdam), Raadsman Wetenschappelijke Integriteit en voormalig Rector Magnificus van de Universiteit van Amsterdam; prof.dr. L.A. de Klerk (Universiteit van Amsterdam), emeritus hoogleraar Algemene Planologie en voormalig decaan en directeur aan de Universiteit van Amsterdam; en prof.dr. C.A.J. Klaassen (Universiteit van Amsterdam), hoogleraar Mathematische Statistiek en voormalig vice-decaan en directeur aan de Universiteit van Amsterdam.

De opdrachten aan de drie Commissies stemmen inhoudelijk overeen. Wel is aan de Commissies Noort (Groningen) en Drenth (Amsterdam) expliciet de opdracht gegeven ook juridische aspecten te bekijken.

1.4 Statistische Ondersteuning

Het werk van de Commissies is ondersteund door teams van statistici. Voor de Commissie Levelt was dit dr. M.A.L.M. van Assen; voor de Commissie Noort bestond het team uit prof.dr. R.H. Koning, dr. W.J. Post en dr. M.A.J. van Duijn en voor de Commissie Drenth uit prof.dr. H.L.J. van der Maas en dr. D. van Ravenzwaaij.

Bij de bestudering, heranalyse en evaluatie van al het materiaal is de inbreng van deze statistici onmisbaar geweest. De aanpak van de statistici is bij aanvang van de werkzaamheden afgestemd met de Commissies en is gecoördineerd tussen de drie Commissies en de drie groepen statistici. Ook tijdens de werkzaamheden is er overleg geweest tussen de statistici onderling en tussen de statistici en de Commissies. De statistici rapporteerden in een bericht aan hun Commissie over opvallende zaken in de betreffende publicatie zoals verdachte onregelmatigheden en onwaarschijnlijke uitkomsten, in het licht van beschikbare data en overig onderzoeksmateriaal. Tevens gaven zij aan in welke mate zij aanwijzingen voor fraude aanwezig achtten. Uiteraard zijn alle in dit rapport weergegeven conclusies, met betrekking tot het uiteindelijke frauduleuze karakter van de publicaties en de verdere consequenties die daaraan in ruimere zin verbonden worden, geheel voor de verantwoordelijkheid van de betreffende Commissie.

1.5 Aanleiding tot het onderzoek

De fraudezaak is in gang gezet door drie jonge onderzoekers uit het Departement Sociale Psychologie van de Tilburg University. Zij hebben eind augustus 2011 hun verdenkingen met betrekking tot de datavervalsing van de heer Stapel aan de departementsvoorzitter prof.dr. M. Zeelenberg gemeld. Na maanden van observatie hadden zij voldoende details verzameld om aan te tonen dat er iets niet klopte. De onderzoekers verdienen alle lof voor het melden van deze verdenkingen. Zij zaten immers in een afhankelijke positie en hadden veel te verliezen. Na deze melding heeft de departementsvoorzitter dit vermoeden van fraude direct medegedeeld aan de Rector Magnificus van Tilburg University, die op zijn beurt onmiddellijk met de klokkenluiders heeft gesproken. De Commissie Levelt heeft geconstateerd dat er in Tilburg eerder andere klokkenluiders geweest zijn. Drie jonge onderzoekers hebben in een vroeger

stadium aan de bel getrokken bij bestuurders van de Faculteit vanwege onregelmatigheden in de geleverde datasets van de heer Stapel en voor hen waren de risico's even groot. Ook twee hoogleraren hebben eerder geconstateerd dat data 'te mooi waren om waar te zijn'. Geen van deze eerdere meldingen hebben echter tot verdere stappen geleid. Daarvoor bestonden verschillende redenen, waarop in de hoofdstukken 4 en 5 ingegaan wordt.

1.6 Karakter rapportage

Dit rapport bevat de gezamenlijke eindrapportage van de Commissies Levelt, Noort en Drenth. Er wordt een compleet overzicht gegeven van de publicaties van de heer Stapel en van alle dissertaties waarbij de heer Stapel actief als (co)promotor betrokken is geweest. Indien fraude is vastgesteld of indien er aanwijzingen voor fraude zijn gevonden wordt dit aangegeven (zie ook paragraaf 2.3 en 2.4.2). Ook wordt de aard van de fraude en de onderzoekscultuur waarbinnen deze heeft plaatsgevonden beschreven.

De Commissies hebben er voor gekozen het in oktober 2011 gepresenteerde interim-rapport over door de heer Stapel gemaakte inbreuk op de wetenschappelijke integriteit uit te breiden en aan te passen om zo tot één definitief eindrapport van de drie gezamenlijke Commissies te komen.

2 Werkwijze Commissies

2.1 Samenwerking Commissies en verdeling werkzaamheden

Gegeven de opdracht is elk van de drie Commissies onafhankelijk en heeft elke commissie een eigen verantwoordelijkheid. Bij de uitvoering van deze opdracht hebben de Commissies Levelt, Noort en Drenth in goed overleg samengewerkt waarbij de coördinatie lag bij de Commissie Levelt van Tilburg University. De werkverdeling tussen de drie Commissies met betrekking tot het onderzoeken van de publicaties van de heer Stapel was als volgt. Als eerste uitgangspunt zijn de proefschriften genomen: deze zijn steeds onderzocht door de Commissie van de universiteit waar de promotie heeft plaats gevonden, evenals de artikelen die onderdeel uitmaken van een proefschrift (zij het soms in wat gewijzigde vorm ten opzichte van het hoofdstuk in het proefschrift). Het merendeel van de promoties heeft in Groningen plaatsgevonden (13) omdat ook na zijn vertrek naar Tilburg de heer Stapel vanwege aangevane verplichtingen in Groningen als (co)promotor optrad. Deze dertien dissertaties en de daarbij behorende artikelen (20) kwamen daardoor voor rekening van de Commissie Noort. In Tilburg zijn door de Commissie Levelt 5 dissertaties onderzocht met 9 bijbehorende artikelen. In Amsterdam is de heer Stapel niet als (co)promotor bij promoties betrokken geweest.

Voor de overige publicaties geldt dat de Commissie Drenth van de Universiteit van Amsterdam de (32) publicaties onderzocht heeft die verschenen zijn in de periode van 1993 t/m 1999, waarbij speciale aandacht werd geschonken aan het onderzoek waarop de dissertatie van de heer Stapel is gebaseerd. De Commissie Noort van de Rijksuniversiteit Groningen onderzocht de 31 overige publicaties van 2000 t/m 2006 en de Commissie Levelt van Tilburg University, 45 overige publicaties van 2007 t/m 2011. (ASPO) Proceedings zijn niet als publicaties beschouwd door de Commissies.

2.2 Formele status commissies

De *Commissie Levelt* is een vertrouwenscommissie ingesteld door de Rector Magnificus volgens de Regeling Wetenschappelijke Integriteit van Tilburg University. De werkwijze van de Commissie wordt beheerst door deze regeling. Volgens artikel 4 van deze regeling geschiedt behandeling van de zaak achter gesloten deuren. Dit betekent dat informatie niet op naam zal worden gedeeld met derden tenzij dit voortvloeit uit de opdracht van de Commissie. De Commissie kan wel anoniem gebruik maken van de informatie beschikbaar gesteld door de informanten om haar bevindingen te staven. De beraadslagingen van de Commissie geschieden op basis van het beginsel van hoor en wederhoor. De Commissie overweegt en beoordeelt de standpunten van partijen in elk geval binnen de kaders en normen zoals die zijn geformuleerd in de Notitie Wetenschappelijke Integriteit van de Koninklijke Nederlandse Akademie van Wetenschappen. Ditzelfde geldt op de wezenlijke onderdelen ook voor de Commissies Noort en Drenth. De *Commissie Noort* is te beschouwen als een speciale ad hoc commissie van de Commissie voor Wetenschappelijke Integriteit (CWI) van de RUG. Twee van haar leden zijn lid van het CWI, het derde, externe lid is als deskundige afkomstig van de Universiteit Twente. De *Commissie Drenth* is een speciale, ad hoc commissie ingesteld door de Rector Magnificus van de UvA.

2.3 Procedure

De Commissies zijn gestart met het verzamelen van alle publicaties van de heer Stapel, verschenen in de periode 1993 t/m 2011, inclusief beschikbare datasets, vragenlijsten, stimulusmateriaal, hypothesen en mailwisselingen. Het bestuderen van al dit materiaal mede met behulp van de statistici heeft het grootste deel van de onderzoekstijd van de Commissies in beslag genomen.

De Commissie Noort heeft bij de verzameling van al het relevante materiaal gebruik gemaakt van een vragenlijst (bijlage 2) die is ontworpen voor de betreffende coauteurs. Hierin lag de nadruk op de wijze van verzamelen en bewerken van de data voor het betreffende artikel of hoofdstuk en op de rol van de coauteur en die van de heer Stapel bij de publicatie. De Commissie Levelt heeft hetzelfde type materiaal verzameld via specifieke schriftelijke en mondelinge vragen aan de coauteurs. Voor de Commissie Drenth was dit type materiaal niet meer beschikbaar gelet op de ouderdom van de onderzoeken en publicaties.

Aan de heer Stapel is gevraagd zelf aan te geven bij welke publicaties gebruik is gemaakt van gefingeerde data. Voor zijn publicaties van 1993 tot en met medio 2011 heeft hij in september 2011 een eerste lijst verstrekt met daarin aangemerkt bij welke publicaties, naar zijn beste herinnering, wel en bij welke geen gebruik is gemaakt van gefingeerde data.

Met betrekking tot een tweede lijst, samengesteld door de Commissie Levelt en voornamelijk bestaande uit hoofdstukken in boeken en te verschijnen artikelen, heeft de heer Stapel in eerste instantie aangegeven lichamelijk en geestelijk niet in staat te zijn het mogelijk frauduleuze karakter van deze publicaties aan te geven. Maar in december 2011 heeft de Commissie Levelt alsnog een reactie van de heer Stapel op deze tweede lijst ontvangen, met daarbij de kanttekening dat hij gezien zijn gezondheidstoestand, niet de garantie kan geven dat de lijst met frauduleuze publicaties volledig is. Ook gaf hij aan bij sommige publicaties niet meer helemaal zeker te zijn van de gang van zaken.

Daarnaast is de heer Stapel ook nog om reacties gevraagd over het frauduleuze karakter van de dissertaties die hij begeleid heeft en (opnieuw) over een aantal publicaties die hij zelf in een eerdere fase niet had aangemerkt als frauduleus, maar waarvan de betreffende Commissie op grond van haar onderzoek concludeerde dat ze wellicht of waarschijnlijk toch frauduleus waren. Ook deze publicaties, alsmede de dissertaties heeft de heer Stapel in drie categorieën ingedeeld: goed, twijfel of besmet.

De Commissies hebben veel betrokkenen geïnterviewd: klokkenluiders, (voormalige) promovendi, coauteurs, collega's en (oud)bestuurders, dit alles om een zo volledig mogelijk beeld te krijgen van wat zich heeft afgespeeld. Waar mogelijk zijn deze gesprekken in persoon gevoerd, in andere gevallen per skype of telefoon. Een compleet overzicht van de gesprekspartners van de drie Commissies is te vinden in bijlage 3. De Commissies hebben geprobeerd de gesprekken met betrokkenen zo open mogelijk te voeren om zoveel mogelijk nuttige informatie te verkrijgen. Dit geldt zowel voor de gesprekken waarbij in meer algemene zin informatie werd verstrekt door de gesprekspartners als voor gesprekken waarin op specifieke vragen van de Commissies gereageerd werd. Van alle gesprekken zijn door de betreffende secretaris verslagen of notities gemaakt die benut zijn voor intern gebruik van de Commissies. In Groningen zijn na verkregen toestemming bandopnamen gemaakt van de gevoerde gesprekken.

Op basis van alle beschikbare informatie hebben de Commissies een onderscheid gemaakt tussen

- i. Publicaties waarin fraude is vastgesteld
- ii Publicaties waarin fraude niet kon worden vastgesteld.

Voor de (oudere) publicaties uit de Amsterdamse periode en voor een deel van de Groningse publicaties (17 artikelen) was de beschikbare informatie dusdanig dat publicaties beoordeeld zijn naar de mate dat er *aanwijzingen* voor fraude zijn gevonden. In paragraaf 2.4.2 wordt hier nog verder op ingegaan.

Ten aanzien van de bekendmaking van de resultaten van het onderzoek, waren de Rectores Magnifici als opdrachtgevers en de Commissies van het begin af aan van oordeel dat de conclusies van de Commissies openbaar gemaakt moesten worden: wetenschap moet ten principale een openbare en controleerbare aangelegenheid zijn. Voor een juist begrip zij overigens vermeld dat de heer Stapel zich nooit heeft verzet tegen openbaarmaking van de conclusies.

Na zorgvuldige afwegingen hebben de Commissies er voor gekozen hun conclusies per publicatie incrementeel bekend te maken via een openbare website (<https://www.commissielevelt.nl/>). Op deze wijze werd zo snel en volledig mogelijk tegemoet gekomen aan de behoefte aan informatie in de wetenschappelijke wereld en daarbuiten, maar vooral ook aan de behoeften van diverse coauteurs, expliciet kenbaar gemaakt aan de Commissies, om zo spoedig mogelijk over het commissieoordeel te kunnen beschikken. Dit woog op tegen het nadeel dat betrokkenen steeds weer met de Stapel-fraude zouden worden geconfronteerd. Bij publicatie op het web werd doorgaans voorrang gegeven aan dissertaties boven andere publicaties, aan recente publicaties boven oudere, en aan artikelen waarvoor dringende ad-hoc redenen bestonden ze snel te behandelen.

Voorafgaand aan de presentatie op de website van de status van de publicaties werd de volgende procedure doorlopen:

1. De betreffende Commissie komt tot een oordeel met betrekking tot de publicatie.
2. In het contact met de coauteurs worden doorgaans additionele vragen gesteld en wordt het oordeel besproken. Als de publicatie onder categorie (i) (frauduleus) valt of, in het geval van de Commissie Drenth en deels de Commissie Noort, aanwijzingen voor fraude bevat, wordt over het voornemen tot plaatsing van dit oordeel op de website gecommuniceerd, waarbij de coauteurs feitelijke onjuistheden kunnen melden.
3. De coördinerende Commissie informeert de heer Stapel over de voorgenomen plaatsing van de oordelen over de publicaties, waarbij hij de gelegenheid krijgt feitelijke onjuistheden aan te geven.
4. Ongeveer één keer per maand rapporteren de Commissies aan hun Rectores en Colleges van Bestuur over de publicaties waar op dat moment zekerheid over bestaat. De drie Commissies informeren ook elkaar.
5. De Colleges van Bestuur van de universiteiten informeren na aanvaarding van de rapportage de betrokken tijdschriften en subsidieverstrekkingen, direct of na aanbidding van het eindrapport, over de conclusies met betrekking tot de publicaties.
6. De Commissies Levelt, Noort en Drenth publiceren gezamenlijk de bevindingen op de website, gecoördineerd door de Commissie Levelt. Voor iedere publicatie waarin fraude is vastgesteld wordt de essentiële informatie waarop de conclusie is gebaseerd, gepresenteerd.

De Commissie Levelt heeft een volledige lijst van de onderzochte publicaties opgesteld en een lijst met publicaties waarin de Commissie fraude heeft kunnen vaststellen. Deze laatste lijst wordt hieronder weergegeven in hoofdstuk 3 en in bijlage 4. In deze bijlage is ook de volledige lijst van bestudeerde publicaties vermeld.

De Commissie Drenth heeft haar oordelen op een andere wijze uitgedrukt omdat zij in geen enkel geval over de oorspronkelijke data kon beschikken waarop de analyses en conclusies in de publicaties gebaseerd waren. Bovendien was het ten aanzien van deze oude publicaties moeilijk om nog betrouwbare informatie van coauteurs te krijgen. In dit rapport, zoals ook op de website, wordt door de Commissie Drenth een onderscheid gemaakt tussen bewijs van fraude ('beyond reasonable doubt') en sterke aanwijzingen voor fraude ('evidence of fraud'). De Commissie Drenth rapporteert dan ook een volledige lijst van de onderzochte publicaties en een lijst met publicaties waarin sterke aanwijzingen voor fraude zijn gevonden. Deze lijsten zijn te vinden in hoofdstuk 3 en in bijlage 6.

De Commissie Noort volgt voor een aantal publicaties de werkwijze van de Commissie Levelt. Voor een ander deel van de publicaties volgt de Commissie Noort de werkwijze van de Commissie Drenth, om soortgelijke redenen. Deze lijsten zijn te vinden in hoofdstuk 3 en in bijlage 5. In deze laatste bijlage is ook de volledige lijst van onderzochte publicaties opgenomen.

2.3.1 Informatiebronnen

Om tot de conclusie te kunnen komen of een publicatie al dan niet frauduleus is, beschikten de Commissies op basis van de gevolgde procedure over de volgende informatiebronnen:

- a. *De teksten van alle wetenschappelijke publicaties van de heer Stapel, als auteur of coauteur, inclusief alle door hem als (co)promotor begeleide dissertaties.*

Voor zover de Commissies dit hebben kunnen nagaan, is deze lijst volledig. Vrijwel alle publicaties zijn verslagen van de uitkomsten van nieuw experimenteel empirisch onderzoek. Enkele publicaties zijn uitsluitend theoretisch van aard in die zin dat er geen verslag wordt gedaan van nieuw experimenteel onderzoek. Deze theoretische publicaties zijn in afgeleide zin beoordeeld op fraude en wel naar de mate waarin de belangrijke theoretische conclusies in het betreffende artikel tot stand zijn gekomen op basis van publicaties waarin sprake was van fraude.

Door de heer Stapel geredigeerde boeken zijn niet beoordeeld door de Commissies aangezien hierbij alleen hoofdstukken van diverse auteurs gebundeld worden. Eventueel door de heer Stapel zelf geschreven hoofdstukken voor zover die experimenten bevatten, zijn uiteraard wel onderzocht door de Commissies.

- b. *De experimentele data.*

Veel moeite is gedaan de, al of niet frauduleuze, datasets te verkrijgen waarop de analyses en conclusies in de publicaties zijn gebaseerd. Lang niet in alle gevallen waren de data nog beschikbaar, zeker niet wanneer het onderzoek langer geleden was uitgevoerd. De coauteurs hebben allen met de Commissies meegewerkt om de data boven tafel te krijgen. In sommige gevallen bleek het niet mogelijk de data te verkrijgen, omdat deze niet beschikbaar of onvolledig waren of soms ook in een vorm die zorgvuldige analyse onmogelijk maakte.

De Commissie Levelt heeft 46 artikelen van experimenteel onderzoek beoordeeld. Hiervan waren voor 38 artikelen de data beschikbaar (en in twee gevallen ook nog de ruwe data in de vorm van ingevulde vragenlijsten).

De Commissie Noort heeft 45 artikelen van experimenteel onderzoek beoordeeld. Hiervan waren voor

29 artikelen de data beschikbaar.

De Commissie Drenth heeft 32 publicaties waarvan de heer Stapel (co)auteur was onderzocht, inclusief 6 artikelen die deel uitmaken van zijn dissertatie. Van geen van deze publicaties waren nog oorspronkelijke data beschikbaar of terug te vinden.

c. Het onderzoeksmateriaal.

Dit zijn de schriftelijke stukken die de opzet en precieze uitvoering van het experiment weergeven.

Hieronder vallen zowel de afgenomen vragenlijsten en het stimulusmateriaal als voorspellingen van de uitkomsten van de experimenten en mailwisselingen tussen de heer Stapel en de onderzoeker. Ook deze zijn via de coauteurs verkregen. In lang niet alle gevallen waren deze echter nog beschikbaar.

De Commissie Levelt had in 33 van de 46 artikelen van experimenteel onderzoek, onderzoeksmateriaal van tenminste één experiment in het artikel tot haar beschikking.

De Commissie Noort had in 12 van de 45 artikelen van experimenteel onderzoek, onderzoeksmateriaal van tenminste één experiment in het artikel tot haar beschikking.

De Commissie Drenth had bij geen van de 32 onderzochte publicaties onderliggend onderzoeksmateriaal ter beschikking.

d. De verklaringen van de heer Stapel over het frauduleuze karakter van de data.

e. De verklaringen van de coauteurs over de analyses en de wijzen waarop het experiment is ingericht, al dan niet in antwoord op specifieke vragen van de Commissies.

f. De rapportages van de statistici.

2.3.2 Allesomvattend onderzoek

Hoewel de bekentenis van de heer Stapel en zijn aanwijzing van frauduleuze publicaties een vrijwel voldoende grond vormen voor het frauduleus verklaren van een publicatie, hebben de Commissies toch besloten om niet alleen met vele betrokkenen te praten, maar ook om al het beschikbare materiaal, alle publicaties en alle experimenten in deze publicaties te bestuderen. Een belangrijke reden hiervoor is het feit dat de heer Stapel bij diverse publicaties heeft aangegeven niet precies meer te weten of de publicatie al dan niet met gefingeerde data tot stand is gekomen en hij in algemene zin een voorbehoud heeft gemaakt waarin hij aangeeft zich uiteraard te kunnen vergissen, zeker ook gezien zijn lichamelijke en psychische gezondheidstoestand van dat moment. Publicaties door hem aangeduid als frauduleus zouden dus ook eventueel niet-frauduleus kunnen zijn, maar ook door hem als ‘goed’ gekwalificeerde publicaties zouden toch besmet kunnen zijn. In principe bevat elk afzonderlijk artikel een unieke bijdrage aan de wetenschap. Het is dus van groot belang voor elk afzonderlijk artikel een zo groot mogelijke zekerheid te verkrijgen over het eventuele frauduleuze karakter.

Een verdere reden voor de gekozen werkwijze van de Commissies is gelegen in de opdracht aan de Commissies waarbij niet alleen vastgesteld moest worden welke publicaties van de heer Stapel frauduleus zijn maar ook of er anderen direct of indirect bij betrokken waren (de omvang van de fraude).

Eveneens lag het in de opdracht aan de Commissies te onderzoeken welke omstandigheden en procedures er toe geleid hebben dat de fraude deze omvang in aantal en tijd heeft kunnen krijgen (de onderzoekscultuur). Daarvoor is een grondige kennis van de feitelijke publicatiepraktijk nodig, welke onder andere verkregen moest worden uit de bestudering en heranalyse van al het beschikbare materiaal.

2.3.3 Conclusie per publicatie

Bij het beoordelen van het frauduleuze karakter van een artikel zijn alle bovenstaande informatiebronnen in samenhang met elkaar benut en ook gekoppeld aan aanwijzingen over de betrokkenheid van de heer

Stapel bij de verzameling, analyse en rapportage van de data. Hierbij speelde de rapportage van de statistici een belangrijke rol. De uiteindelijke beoordeling geschiedde door de Commissies na bestudering van de rapportage in overleg met de statistici en na een check door de coauteurs die in de gelegenheid werden gesteld op de bevindingen te reageren. In sommige gevallen is daarbij na verkregen informatie van de coauteurs het aanvankelijke oordeel van de Commissies aangepast. De uiteindelijke conclusies van de Commissies zijn ook voorgelegd aan de heer Stapel voor de vaststelling van feitelijke onjuistheden. Dit laatste heeft in geen enkel geval geleid tot aanpassing van het finale oordeel. Daarna werd overgegaan tot publicatie van de bevindingen en hoofdoverwegingen op de website. Op de webpagina en in de bijlagen 4, 5 en 6 zijn per Commissie de hoofdoverwegingen voor het bepalen van het frauduleuze karakter kort vermeld.

2.3.4 Hoor en wederhoor: algemeen

De Commissies hebben het principe van hoor en wederhoor toegepast. Dit gold met name bij de vaststelling van fraude in de publicaties. In eerste instantie zijn de coauteurs door de Commissies in de gelegenheid gesteld te reageren op vragen van de Commissies, veelal naar aanleiding van de rapportages van de statistici. Vervolgens hebben zij, voor zover van toepassing, ook de gelegenheid gehad feitelijke onjuistheden in de conclusies van de drie Commissies te signaleren.

2.3.5 Hoor en wederhoor: de heer Stapel

De heer Stapel is in de gelegenheid gesteld commentaar te geven op de conclusies en rapportages van de Commissies. Ook is hij uitgenodigd voor een persoonlijk onderhoud met de Commissies. Hij heeft op 14 oktober 2011, voor het verschijnen van het interim-rapport, evenwel aan de Commissie Levelt aangegeven daartoe niet in staat te zijn. Vervolgens is een concept-versie van het interim-rapport aan hem en zijn advocaat voor commentaar ter hand gesteld.

Op 15 mei 2012 is de heer Stapel wel op de uitnodiging van de Commissie Levelt in kunnen gaan voor een persoonlijk gesprek in het kader van hoor en wederhoor.

Op 9 oktober 2012 heeft hij ook gevolg gegeven aan een uitnodiging van de Commissie Noort voor een uitgebreid persoonlijk gesprek in het kader van hoor en wederhoor.

Met de Commissie Drenth heeft geen persoonlijk onderhoud plaats gevonden, maar de conclusies van de Commissie Drenth zijn eveneens voor commentaar voorgelegd aan de heer Stapel. De heer Stapel heeft via zijn advocaat als volgt gereageerd:

‘Cliënt heeft reeds eerder erkend in veel van zijn onderzoek te hebben gefraudeerd. Cliënt heeft eveneens erkend dat hij slordig is geweest, zaken slecht heeft vastgelegd en data van hun mooiste kant heeft laten zien. Voor zover cliënt zich kan herinneren, is in zijn Amsterdamse periode van fraude echter geen sprake geweest.’

Ook deze voorliggende eindrapportage ‘Falende wetenschap’ is wederom aan de heer Stapel voorgelegd met het verzoek feitelijke onjuistheden aan te geven. De heer Stapel en zijn advocaat hebben schriftelijk gereageerd op een concept-versie van het rapport. Door hen aangegeven feitelijke onjuistheden zijn vervolgens in de eindversie gecorrigeerd.

2.4 Vaststelling fraude

2.4.1 Afbakening begrip fraude

Wat verstaan de Commissies onder ‘fraude’? Bij de bestudering van alle beschikbare informatie (publicaties, data, onderzoeksmateriaal, inhoud van gevoerde gesprekken enz.) kwamen vele schendingen van de regels van behoorlijk wetenschappelijk onderzoek aan het licht. Niet alleen werden vaak kleinere onregelmatigheden gevonden, ook werden veelvuldig ernstige schendingen van de wetenschappelijke integriteit geconstateerd. In de hoofdstukken 3, 4 en 5 wordt hier uitvoerig op ingegaan.

Besloten is echter, in het licht van de opdrachten aan de Commissies, de definitie van fraude te beperken tot het aanmaken, vervalsen, of aanvullen van data, dan wel het geheel of gedeeltelijk verzinnen of vervalsen van analyse-uitkomsten, dan wel het op cruciale punten onjuist weergeven van de inrichting en aard van het experiment. Het begrip data verwijst daarbij naar de gecodeerde ruwe onderzoeksgegevens zoals die voorkomen in de datamatrix waarin voor elke onderzoekspersoon de al dan niet gefingeerde scores zijn vastgelegd voor alle variabelen.

De Commissies Levelt en Noort noemen een (empirische) publicatie frauduleus als ten minste één van de deelstudies in de publicatie gebaseerd op experimentele data, frauduleus is.

De overige overtredingen van de wetenschappelijke spelregels worden beschouwd als (ernstige) vormen van *questionable research practices*, maar daarvoor wordt hier niet de term ‘frauduleus’ gehanteerd. Bij de bespreking van de Onderzoekscultuur (hoofdstuk 5) komen veel aangetroffen varianten van zulke, vaak ernstige, overtredingen van de regels van het wetenschappelijk onderzoek aan de orde in alle onderzoeksfasen.

2.4.2 Bewijs van en aanwijzingen voor fraude

Bij de uiteindelijke vaststelling van fraude van een bepaalde publicatie gaat het vaak om een combinatie van verschillende fraude indicatoren. Sommige indicatoren zijn minder dwingend dan andere, maar deze vormen dan in combinatie met andere constatering wel een extra bewijs voor de vaststelling van fraude. Op de website en in deze eindrapportage (bijlagen 4, 5 en 6) worden per publicatie in het kort de belangrijkste gronden voor de vaststelling van fraude weergegeven.

De Commissies spreken enkel van bewijs van fraude bij een publicatie als daaromtrent geen redelijke twijfel kan bestaan.

Aangezien de Commissie Drenth geen oorspronkelijke data of ander onderzoeksmateriaal meer tot haar beschikking had, hebben de analyses van deze Commissie zich uitsluitend moeten richten op verdachte onregelmatigheden en, van uit statistisch oogpunt, zeer onwaarschijnlijke resultaten. De analyses van de Commissie Drenth hebben geleid tot een oordeel over de waarschijnlijkheid van fraude per artikel, op een schaal met de ankers: niet van toepassing - geen - verwaarloosbaar - licht - tamelijk sterk - sterk. De Commissie Drenth heeft de laatste twee categorieën genomen als grond voor het oordeel ‘evidence of fraud’. De Commissie Drenth wijst erop dat het hier gaat om een statistisch gezien grote mate van waarschijnlijkheid, maar dat dit, vanwege het niet beschikbaar zijn van de oorspronkelijke data, en het feit dat de heer Stapel ter zake geen frauduleus gedrag erkent, niet gelijkstaat met een juridisch bewijs van fraude.

Ook voor de Commissie Noort gold dat ze voor een aantal artikelen (17) geen data of ander onderzoeksmateriaal meer tot haar beschikking had. Voor deze artikelen heeft de Commissie Noort dezelfde werkwijze gevolgd als de Commissie Drenth, zij het dat alleen ‘sterke’ aanwijzingen voor fraude hebben geleid tot classificatie ‘evidence of fraud’.

2.5 Concrete werkwijze: toelichting en voorbeelden

In deze paragraaf wordt beschreven op welke wijze de Commissies tot de concrete vaststelling van fraude zijn gekomen.

De aangeleverde datasets, indien beschikbaar, zijn onderzocht op eventuele merkwaardigheden en verdachte uitkomsten. Ook zijn de datasets geanalyseerd en zijn de uitkomsten vergeleken met de gerapporteerde resultaten in de publicatie. De aangeleverde vragenlijsten en het stimulusmateriaal zijn vergeleken met de aangeleverde datasets en met de inhoud van de publicatie. Ook de publicatie zelf is onderzocht op verdachte onregelmatigheden. Mededelingen van de coauteurs met betrekking tot de publicatie zijn beoordeeld, mede als reactie op vragen van de betreffende Commissie. Tenslotte is het oordeel van de heer Stapel over de publicatie in beschouwing genomen.

Een nauwkeuriger beeld van de werkwijze van de Commissies kan verkregen worden door de belangrijkste concretisering van de hierboven genoemde algemene stappen en werkwijzen weer te geven. Om de variëteit aan gebruikte methoden en benaderingen overzichtelijk te houden, zijn ze hieronder verdeeld over de belangrijkste bronnen van informatie waarop ze betrekking hebben. Allereerst wordt aangegeven hoe het *onderzoeksmateriaal* over opzet en uitvoering concreet gebruikt is, waarbij het vooral ging om de vraag in hoeverre de informatie vervat in dit materiaal overeenstemt met hetgeen in de data (indien beschikbaar) en in de publicatie gevonden werd. Dan zijn de *data op zich* onderzocht op merkwaardigheden. Vaak gebeurde deze visuele inspectie van de datamatrix als er op grond van andere overwegingen reden was om fraude te vermoeden. Ook hebben *(her)analyses van de data* plaats gevonden, zowel om na te gaan of de gepubliceerde resultaten reproduceerbaar waren als om merkwaardigheden en onjuistheden in de analyses en in de datasets op te sporen. Tenslotte richtte de aandacht zich op de *publicatie zelf*, soms uitsluitend op hetgeen in het artikel of hoofdstuk vermeld is bij gebrek aan verdere informatie, soms in relatie tot het onderzoeksmateriaal en de data. De Commissie Levelt heeft vaak alle drie de bronnen kunnen gebruiken, de Commissie Noort heeft hoofdzakelijk de laatste twee bronnen gebruikt, en de Commissie Drenth heeft zich met name op de derde bron moeten baseren.

Het zou te ver voeren een compleet overzicht te geven van alle specifieke concrete methoden die de Commissies gehanteerd hebben. Maar de algemene werkwijze en de belangrijkste variaties daarin worden hieronder aangegeven, mede door middel van concrete voorbeelden van resultaten die zo verkregen kunnen worden.

2.5.1 Inrichting en uitvoering experiment

Een belangrijke en vaak eerste stap in het onderzoekswerk van de Commissies was na te gaan hoe het experiment opgezet en uitgevoerd was. Het ging daarbij om de vraag of de informatie over de inrichting en uitvoering van het experiment in de diverse informatiebronnen (het onderzoeksmateriaal, het artikel, de data en de gesprekken) consistent was. Zo werd bijvoorbeeld gecontroleerd of de experimentele condities,

de volgorde van de condities en het stimulusmateriaal in het artikel correct gerapporteerd werden, dat wil zeggen in overeenstemming met het toegezonden onderzoeksmateriaal en de data. Ook werd gecheckt of de datasets alle variabelen van het stimulusmateriaal bevatten en of alle variabelen in de datasets ook voorkwamen in het stimulusmateriaal. In deze confrontaties werden regelmatig discrepanties aangetroffen. Steeds is nagegaan, veelal door gerichte schriftelijke of mondelinge vragen aan de coauteurs, of de discrepanties en merkwaardigheden te verklaren waren, anders dan door datamanipulatie. In de meeste gevallen waren er ofwel goede inhoudelijke verklaringen voor de discrepanties, ofwel moest een en ander op het conto van ‘slordigheid’ of ‘slechte rapportage’ geschreven worden. In de overblijvende gevallen leidde dit tot de overtuiging van datafabricage.

Wat betreft de inrichting van het experiment is bijzondere aandacht geschonken aan onderzoek op scholen. Volgens de verklaringen van de heer Stapel waren veel op scholen afgenomen onderzoeken frauduleus en hebben deze eenvoudigweg nooit plaats gevonden. Voor sommige experimenten is het vanwege allerlei praktische problemen ook moeilijk voor te stellen dat ze op die scholen gerealiseerd hadden kunnen worden. Dit laatste geldt ook voor de experimenten op stations en in treinen of onder rechters. Eveneens is extra kritisch gekeken naar onderzoeken waarbij de heer Stapel, terwijl hij in Tilburg verbleef, tegen Tilburgse promovendi of coauteurs zei dat hij dat onderzoek wel mee kon laten lopen in Groningen en tegen Groningse promovendi dat hij dat wel in het laboratorium van Tilburg kon laten doen.

De aanwijzingen verkregen in deze categorie vormden bewijzen voor fraude indien vastgesteld kon worden dat het onderzoek nooit plaats had gevonden, maar ook leidden de verkregen aanwijzingen vaak tot vermoedens van fraude, die dan verder onderzocht werden.

2.5.2 Merkwaardigheden in de datamatrix

Een belangrijke bron van informatie voor de bepaling van mogelijke fraude in een publicatie, ligt in de aangeleverde datasets. Zo kunnen de data vergeleken worden met het onderzoeksmateriaal en met de informatie in de publicaties, om discrepanties in opzet en uitvoering te ontdekken (paragraaf 2.5.1). Ook kunnen de aangeleverde datasets opnieuw geanalyseerd worden om de uitkomsten in de publicaties te controleren (paragraaf 2.5.3). Maar vooral kan ‘visuele inspectie’ van de datamatrices patronen zichtbaar maken die minstens de verdenking van fraude oproepen. Een belangrijk bewijs van fraude is het kopiëren van de data van het ene experiment naar het andere. Er is, waar daar aanleiding voor was, minutieus onderzocht of de scores op bepaalde variabelen in de diverse datasets van elkaar gekopieerd waren. In de oorspronkelijke datasets is dit meestal niet direct te zien, maar na het herordenen van de respondenten, niet naar (oplopend) respondentnummer maar naar bijvoorbeeld de scores op de afhankelijke variabelen, wordt dit kopiëren zichtbaar. Voor sommige experimenten werd met behulp van simulatie op basis van geobserveerde frequentieverdelingen en correlaties vastgesteld hoe onwaarschijnlijk het aantal identieke waarnemingen was.

Bijzondere aandacht is ook geschonken aan de afwezigheid van *missing data* en de spreiding van de scores van de proefpersonen binnen en tussen de onderzoekcondities (experimentele en controle conditie). Indien in een experiment de metingen van de variabelen plaatsvinden met behulp van vragenlijsten die de proefpersonen ‘met de hand’ invullen, zijn ‘missing data’ in een of andere vorm vrijwel onontkoombaar. Het totaal ontbreken van zulke scores roept dus een verdenking van frauduleuze datamanipulatie op. Eveneens vallen in de verdachte categorie uitkomsten zoals: alle personen in de experimentele groep behalen zonder uitzondering de hoogste scores (bijv. 6 en 7) en alle personen in de controlegroep de laagste

(bijv. 1 en 2). In ‘normaal’ uitgevoerd onderzoek zijn dit zeer onwaarschijnlijke uitkomsten en is er dus reden voor verder onderzoek.

2.5.3 (Her)Analyse van de data

Indien de data beschikbaar waren, zijn die gebruikt om de belangrijkste analyse-uitkomsten in de publicaties te checken, zoals de gerapporteerde waarden van gemiddelden en standaarddeviaties, van de betrouwbaarheden, toetsgrootheden, effect sizes en p-waarden. Speciale aandacht kregen die discrepanties tussen de uitkomsten van de heranalyses en de gepubliceerde bevindingen die de conclusies in de publicaties zouden aantasten. Steeds is nagegaan, vaak via contacten met de coauteurs, of er alternatieve verklaringen voor deze discrepanties gegeven konden worden, anders dan datafraude. Zo doet zich het selectief weglaten van uitzonderlijke scores of van bepaalde onderzoekspersonen voor. Daar kunnen goede redenen voor zijn, die echter vermeld hadden moeten worden.

Behalve dat deze heranalyses nuttig waren om mogelijke fraude op te sporen, hebben ze de Commissies vooral ook een goed inzicht gegeven in de achterliggende onderzoekscultuur en de feitelijke onderzoekspraktijken bij de totstandkoming van publicaties.

Naast deze strikte heranalyses zijn er ook veelvuldig aanvullende analyses op de data uitgevoerd. Onder andere zijn standaarddeviaties van diverse variabelen berekend, evenals correlaties tussen verschillende items van dezelfde schaal en correlaties per conditie. Ook zijn ongerapporteerde betrouwbaarheden berekend en zijn effect sizes vergeleken over studies en condities heen. De uitkomsten van deze analyses, zoals zeer grote of juist zeer kleine standaarddeviaties, konden aanleiding geven de antwoordpatronen op de betreffende variabele in de datamatrix nauwkeuriger te bekijken (als in paragraaf 2.5.2).

Veel van die aanvullende analyses hadden als doel na te gaan hoe reëel de gepubliceerde uitkomsten waren. Zij zullen in de volgende paragraaf 2.5.4 verder besproken worden.

2.5.4 Publicaties

De informatie in de publicaties zelf kan eveneens gebruikt worden om frauduleuze praktijken aan te tonen. Zo bevat vrijwel elke publicatie gegevens over de inrichting van de experimenten. Het kan dan bijvoorbeeld duidelijk worden dat een bepaald experiment in de beschreven omstandigheden om praktische, logistieke redenen zo niet uitgevoerd had kunnen worden. Dit kan een sterke aanwijzing van fraude opleveren, een vermoeden dat vervolgens afgezwakt of versterkt kan worden door informatie van de betrokken onderzoekers of door de informatie in het onderzoeksmateriaal en de data.

Ook de gerapporteerde analyse-uitkomsten in de publicaties kunnen aanwijzingen bevatten voor het overtreden van de basisregels van goed wetenschappelijk onderzoek in het algemeen, en van datafraude in de hier gebezigde zin in het bijzonder. Er zijn globaal gesproken twee soorten datamanipulaties die tot verdachte onderzoeksuitkomsten kunnen leiden. De eerste is het weliswaar werken met echte data, maar met selectieve en onvermelde weglating van uitzonderlijke scores op variabelen of van uitzonderlijke proefpersonen of van uitzonderlijke onderzoeksresultaten, waarbij ‘uitzonderlijk’ inhoudt dat de weggelaten informatie niet strookt met wat de onderzoeker graag als resultaat ziet. De analyseresultaten van dergelijke datamanipulaties wijken doorgaans af van de resultaten verkregen met niet selectief gemanipuleerde data. In de tweede plaats leidt het geheel of gedeeltelijk ‘naïef’ verzinnen van data tot uitkomsten en resultaten die in echte data niet of zelden voorkomen. Indien de datafabricatie, zoals bij de heer Stapel, plaatsvindt door zelf op simpele wijze de datamatrix in- of aan te vullen om het gewenste resultaat te bereiken,

worden vrijwel zeker uitkomsten verkregen die afwijken van reële uitkomsten, in tegenstelling tot wanneer statistisch meer geavanceerde methodes van datafabricatie worden gebruikt.

Die afwijkingen kunnen zich met name uiten in te geringe toevalsfluctuaties, in te grote effecten, in vreemde multivariate verbanden, in onderlinge afhankelijkheid van waarnemingen, en in andere onwaarschijnlijke analyse-uitkomsten. Deze punten worden hieronder nader uitgewerkt, want dat alles is aangetroffen in de frauduleuze publicaties van de heer Stapel.

Te geringe toevalsfluctuaties

Gegevens van twee onafhankelijke onderzoeken van dezelfde situatie zullen toch van elkaar verschillen vanwege toevalsfluctuaties. Mensen onderschatten doorgaans de rol van toeval. Bij het zelf invullen van een datamatrix zullen frauderende onderzoekers de neiging hebben de resultaten van twee ‘identieke’ onderzoeken te zeer op elkaar te laten lijken. Een dergelijk gebrek aan (toevals)variatie kan ook veroorzaakt worden door het systematisch weglaten van ‘uitzonderlijke’ uitkomsten.

Zo komen in frauduleuze publicaties van de heer Stapel vaak onafhankelijk herhaalde onderzoeken voor, van wat in essentie dezelfde experimentele (c.q. controle) condities zijn. Wanneer de gemiddelde gerapporteerde waarden van de afhankelijke variabele (en/of van de standaarddeviaties) voor al die replicatie-condities exact gelijk zijn, tot in vier cijfers achter de komma, zoals in een artikel aangetroffen werd, dan kan tot frauduleus kopieergedrag van data geconcludeerd worden. Maar ook als de gemiddelden niet exact gelijk zijn, maar wel ongeveer, kunnen de resultaten wijzen op frauduleuze manipulatie. Via een statistisch model kan dan bepaald worden hoe waarschijnlijk het is dat zulke kleine verschillen optreden bij onafhankelijke replicaties van dezelfde metingen onder de aanname dat ze ‘in werkelijkheid’ gelijk zijn. Als die kans zeer klein is (bijvoorbeeld 10^{-10}), vormt dit een duidelijke aanwijzing voor datamanipulatie.

Te grote effecten en samenhangen

Deels verband houdend met het bovenstaande zal het weglaten van onwelgevallige resultaten gemakkelijk leiden tot te grote effecten, te groot in de zin van ‘groter dan men doorgaans aantreft in reële niet gemanipuleerde data’. Ook bij naïeve invulling van de datamatrix door de onderzoeker is de tendens te bespeuren irreëel sterke verbanden tussen variabelen te veronderstellen en irreëel grote effecten van de experimentele manipulaties aan te maken. Bij de heer Stapel kwam dit veelvuldig voor. Zo ontstaan zeer grote gemiddelde verschillen op de afhankelijke variabele tussen experimentele en controlegroep met nauwelijks fluctuaties in scores binnen de groepen. ‘Grote effecten’, gerapporteerd of door de statistici berekend, kunnen dus een belangrijke aanwijzing voor datafabricage vormen.

Nu vraagt uiteraard de constatering ‘te groot’ om een standaard: wanneer is een samenhang of effect size te hoog, zodat het leidt tot een vermoeden van datamanipulatie? Zoals gevonden in meta-analyses is de gemiddelde effectgrootte in psychologisch onderzoek waarin twee groepen of condities met elkaar worden vergeleken, zoals gemeten met Cohen’s *d*, ongeveer 0,5 (o.a. Meyer et al., 2001; Richard, Bond, & Stokes-Zoota, 2003). Dit komt overeen met een middelgroot effect, volgens de terminologie van Cohen.

Ter vergelijking met publicaties waar de heer Stapel (co)auteur was en ter verantwoording van de gevolgde werkwijze is eind 2011 door Research Master studenten in Tilburg onder begeleiding van de statisticus van de Commissie Levelt, nagegaan wat in enkele belangrijke internationale tijdschriften op het gebied van de experimentele sociale psychologie, waarin de heer Stapel veelvuldig publiceerde, het gemiddelde en de spreiding van de effect sizes was. Daarbij zijn van het European Journal of Social Psychology (EJSP) issues 1 tot en met 3 uit 2011 bekeken en van het Journal of Personality and Social Psychology (JPSP).

issues 100 (1) tot en met 100 (3) uit 2011, en issues 96 (1) uit 2009, 94 (4) uit 2008, en 93 (5) uit 2007. Van alle artikelen in deze issues zijn, mits gerapporteerd, de Cohen's d en de totale verklaarde variantie van de afhankelijke variabele (η^2) in elke studie genoteerd, althans wat betreft de 'between-subjects designs', het design waar de heer Stapel meestal gebruik van maakte in zijn onderzoeken.

De gemiddelde gerapporteerde Cohen's d in deze issues was 0,69 ($n=158$), iets hoger dan die gevonden in de eerder genoemde meta-analyses. In 71,5% van de gevallen was Cohen's d lager dan 0,8, wat overeenkomt met een groot effect. De gemiddelde totale verklaarde variantie was gelijk aan 22%, en in 80% van de gevallen kleiner dan 35%.

De verbanden zoals gevonden in diverse publicaties van de heer Stapel, met 85% of zelfs 95% procent verklaarde variantie blijken daarbij uitermate zeldzaam; zelfs verbanden met bijvoorbeeld 55% verklaarde variantie komen niet vaak voor. Dergelijk hoge waarden vragen dus om nadere analyse en zijn mogelijk frauduleus.

De grootte van het effect van de experimentele condities op de afhankelijke variabele kan ook geëvalueerd worden in relatie tot de betrouwbaarheid van de meting van de afhankelijke variabele. Zo kan de grootte van een effect nooit sterker zijn dan de betrouwbaarheid van die afhankelijke variabele (onder de gebruikelijke aannamen van onafhankelijke waarnemingen en meetfouten). Als dit toch het geval is, is dit een mogelijke aanwijzing voor fraude. Correcties voor onbetrouwbaarheid kunnen ook leiden tot onwaarschijnlijk grote effect sizes (bijvoorbeeld 90% verklaarde variantie of meer). Dat kwam regelmatig voor in publicaties van de heer Stapel. Onwaarschijnlijk hoge effect sizes en zeer sterke verbanden hebben bijvoorbeeld aanleiding gegeven om te onderzoeken of die corresponderen met bepaalde merkwaardigheden in die datamatrix.

Ongebruikelijke multivariate verbanden

Datafabricage, zoals die van de heer Stapel, kan ook leiden tot vreemde multivariate uitkomsten. Op zich kunnen alle bivariate correlaties er 'normaal' (en in de gewenste richting) uitzien, maar er kunnen dan wel eigenaardige uitkomsten in een multivariate analyse ontstaan, die weliswaar niet direct zichtbaar zijn, maar die in reële data zelden aangetroffen worden. Bijvoorbeeld, een fraudeur geeft scores aan de onafhankelijke variabele X en eveneens aan de afhankelijke variabele Y en wel dusdanig dat er een groot positief effect van X op Y ontstaat. Vervolgens geeft die onderzoeker scores aan een covariaat Z en wel zo dat er een positief verband ontstaat tussen Y en Z . Op zich zien dan de bivariate verbanden X - Y en Y - Z er vrij 'normaal' uit maar dat hoeft niet het geval te zijn voor de bivariate relatie X - Z of voor het partiële verband tussen Y en Z , gecontroleerd voor X , dat wil zeggen binnen de categorieën van X .

Ook kan het bivariaat fabriceren van data tot vreemde uitkomsten leiden bij de zgn. manipulatiechecks. Bij deze checks wordt nagegaan of de experimentele manipulatie X , bijvoorbeeld zich al dan niet in een angstige situatie bevinden, inderdaad leidt tot een hogere of lagere score op de angstvariabele Z . Het effect van de experimentele manipulatie X op de afhankelijke variabele Y wordt dan geacht te lopen via de interveniërende variabele Z . Het effect is theoretisch in een standaard (regressie)analyse gelijk aan het product van het effect van de manipulatie op de interveniërende en het directe effect van de interveniërende op de afhankelijke variabele. Via de zogenaamde manipulatiechecks wordt nagegaan of de experimentele conditie inderdaad op de verwachte wijze samenhangt met de scores op de interveniërende variabele.

Theoretisch zou dan het effect van de onafhankelijke variabele X op de afhankelijke variabele Y klein, zelfs nul, moeten zijn indien de interveniërende variabele Z constant wordt gehouden. Bij het paarsgewijs fabriceren van data wordt hiermee geen rekening gehouden en blijft het effect van X op Y groot, ook als

het effect van interveniërende variabele Z wordt geëlimineerd. Wanneer iets dergelijks werd aangetroffen in publicaties van de heer Stapel was dat aanleiding voor nader onderzoek, zeker wanneer het effect van X op Y groter was dan het effect van X op Z.

Afhankelijke waarnemingen

In experimenteel onderzoek zijn de data van verschillende onderzoekspersonen onafhankelijk van elkaar. Dat wil zeggen dat de kennis van de scores op de diverse variabelen van een willekeurige proefpersoon niets zegt over de scores die een andere willekeurige proefpersoon heeft behaald. In gefabriceerde data is dit moeilijk te realiseren omdat mensen slecht zijn in het toekennen van toevalsfluctuaties. Het is mogelijk met behulp van een statistisch model uit te rekenen wat de kans is dat de gerapporteerde (variantie-analyse) uitkomsten afkomstig zijn van onafhankelijke waarnemingen dan wel van afhankelijke waarnemingen en daarmee van gefabriceerde of gemanipuleerde onderzoeksresultaten. Ofschoon aan zulke statistische modellen vooronderstellingen ten grondslag liggen, zijn ze vrij robuust tegen schendingen van deze assumpties.

Overige onwaarschijnlijke analyse-uitkomsten

Ook zonder de aanwezigheid van datasets kunnen uit de gepubliceerde uitkomsten verdachte datapatronen afgeleid worden. Zo kon uit de gerapporteerde waarde van de standaarddeviatie van een bepaalde variabele bij een gegeven gemiddelde afgeleid worden dat alle scores op die variabele zonder uitzondering uiterste scores waren op de uiteinden van de schaal; een ongebruikelijke uitkomst die kan wijzen op fraude. Ook kon soms bepaald worden dat de kans op bepaalde (toetsings)uitkomsten vrijwel onmogelijk is zonder datamanipulatie. Zo werd bij 24 F-toetsen ter toetsing van de hypothese van geen verschil tussen de experimentele en de controlegroep gerapporteerd dat deze hypothese 24 keer 'bevestigd' werd met $F < 1$. Het is mogelijk met de gegevens uit het artikel zelf de kans op een dergelijke serie van kleine F-waarden uit te rekenen. Als een dergelijke kans, zoals in dit geval, 'onmogelijk' klein is, kan deze in feite alleen verklaard worden door selectieve datamanipulatie.

Zo kan in een publicatie, ook zonder data, bijvoorbeeld nagegaan worden of er onwaarschijnlijke of onmogelijke gemiddelden en standaarddeviaties voorkomen.

Ook geven foutief gerapporteerde vrijheidsgraden (in vergelijking tot het aantal respondenten) en p-waarden die niet corresponderen met de toetsgrootte indicaties voor onregelmatigheden.

Bij de uiteindelijke bepaling van fraude in een artikel en in een concreet experiment zijn bovenstaande criteria en concrete werkwijzen zo veel mogelijk in samenhang met elkaar gebruikt om tot een definitief oordeel over fraude en datamanipulatie te komen.

3 Bevindingen over de aard en omvang van de fraude

3.1 Omvang van de fraude

De Commissies zijn tot de conclusie gekomen dat de omvang van de fraude door de heer Stapel zeer aanzienlijk is. De Commissies zijn gestuit op in totaal 55 publicaties waarin fraude is vastgesteld.

Commissie Levelt

De Commissie Levelt heeft fraude vastgesteld in 34 publicaties. Hiervan zijn, volgens de acknowledgements, dertien publicaties geschreven op basis van de door NWO aan de heer Stapel verstrekte Pionier subsidie en één op basis van de onderzoekssubsidie van Transforum. Ook in drie door de heer Stapel begeleide dissertaties blijken één of meer hoofdstukken gebaseerd op gefingeerde data. Een volledig overzicht van de onderzochte publicaties, de eventuele aan de heer Stapel toegekende subsidies, alsmede de verkorte weergave van de belangrijkste gronden voor fraude zijn terug te vinden in bijlage 4 (zie ook de website www.commissielevelt.nl).

De publicaties waarin de Commissie Levelt fraude heeft vastgesteld zijn:

- 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 of 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 of 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

De dissertaties waarvan in één of meer hoofdstukken datafraude is vastgesteld door de Commissie Levelt zijn:

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

Commissie Noort

De Commissie Noort heeft fraude vastgesteld in 21 publicaties en zeven dissertaties. Daarnaast gaat het in drie publicaties om 'evidence of fraud'. Hiervan zijn, volgens de acknowledgements, 18 publicaties geschreven op basis van de door NWO aan de heer Stapel verstrekte Pionier subsidie, 3 publicaties op basis van de door NWO aan de heer Stapel verstrekte subsidie T32 MH19728 en 2 publicaties op basis van zijn fellowship van de KNAW. Ook in zeven door de heer Stapel begeleide dissertaties blijken één of meer hoofdstukken gebaseerd op gefingeerde data. Een volledig overzicht van de onderzochte publicaties, de eventuele aan de heer Stapel toegekende subsidies, alsmede de verkorte weergave van de belangrijkste gronden voor fraude zijn terug te vinden in bijlage 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 of 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 of 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 of 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 of 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 of 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 of 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 of 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 of 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 of 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)
- De dissertaties waarvan in één of meer hoofdstukken datafraude is vastgesteld door de Commissie Noort zijn:

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

Op basis van de uitgevoerde statistische analyses van de publicaties van de heer Stapel en volgens de in paragraaf 2.4.2 genoemde beslissingsregels heeft de Commissie Noort geconcludeerd dat er sprake is van ‘evidence of fraud’ in de volgende artikelen:

- 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)

De volgende boekhoofdstukken en tijdschriftartikelen zijn (deels) gebaseerd op bevindingen van artikelen waarvan de Commissies ‘evidence of fraud’ hebben vastgesteld:

- 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).

Commissie Drenth

Op basis van de uitgevoerde statistische analyses van de publicaties van de heer Stapel en volgens de in paragraaf 2.4.2 genoemde beslissingsregels heeft de Commissie Drenth geconcludeerd dat er sprake is van 'evidence of fraud' in de volgende artikelen:

- 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. Dissertatie, Hoofdstuk 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. Dissertatie, Hoofdstuk 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.

In hoofdstuk 3 en 5 van de dissertatie van de heer Stapel zijn door de Commissie Drenth sterke aanwijzingen voor fraude gevonden. De Commissie Drenth had specifiek als opdracht na te gaan of sprake was van fraude bij het onderzoek waarop de dissertatie van de heer Stapel was gebaseerd en of dat gevolgen zou moeten hebben voor de aan hem door de UvA verleende doctorsgraad. Deze laatste vraag behoeft uiteindelijk niet te worden beantwoord omdat de heer Stapel zijn doctorsbul op eigen initiatief bij de Rector Magnificus van de UvA heeft ingeleverd. Voor de principiële vraag naar de mogelijkheid van de intrekking van een doctorsgraad zie de aanbevelingen en de juridische opmerkingen van de Commissie Noort in bijlage 7.

De bevindingen van de Commissies wijzen uit dat het fabriceren van data in enigerlei vorm al voor de Tilburgse tijd is begonnen. De eerste publicatie in Groningen waarin fraude is vastgesteld dateert uit 2004 en de eerste publicatie waarin aanwijzingen voor fraude zijn gevonden dateert uit 2001. Voor de Amsterdamse periode geldt dat de eerste publicatie waarin aanwijzingen voor fraude zijn gevonden dateert uit 1996.

3.2 Aard van de fraude

De Commissies hebben op grond van hun bevindingen vastgesteld dat de heer Stapel datasets gemanipuleerd heeft en/of zelf gefabriceerd. Uit de informatie die in gesprekken met coauteurs en andere betrokkenen en collega's verstrekt is, komt het volgende beeld naar voren van de wijze waarop de manipulatie van de gegevensverzameling, -analyse en -rapportage doorgaans heeft plaatsgevonden.

In een eerste variant verliep de voorbereiding voor deze onderzoeken normaal. Na uitvoerige discussies met de coauteur(s) werden stimuli en vragenlijsten ontwikkeld en werden het aantal proefpersonen en de te volgen procedure vastgesteld. Het uiteindelijke afnemen van de vragenlijsten vond echter niet plaats. De heer Stapel creëerde in die gevallen zelf de datasets waarin de gedane voorspellingen grotendeels uitkwamen.

In een tweede variant wijzigde de heer Stapel gegevens in bestaande datasets. De data werden hierbij wel degelijk verzameld en waren doorgaans ingevoerd door een student-assistent. Deze stuurde de dataset vervolgens naar de heer Stapel. Na het manipuleren van de data stuurde de heer Stapel de dataset vervolgens aan de promovendus, die van te voren zijn of haar verwachtingen over de uitkomst al had moeten formuleren. De Commissies zijn in het bezit van datasets die ingevoerd zijn door student-assistenten en de datasets die uiteindelijk bij de promovendus terecht kwamen. De verschillen tussen de datasets laten duidelijk de wijzigingen zien die zijn aangebracht.

In een derde variant stuurde de heer Stapel, na contacten met collega-onderzoekers, hen datasets met de mededeling dat hij deze al lange tijd had liggen, en met de vraag of zij het onderzoek wilden analyseren en opschrijven. Ook hierbij leverde de heer Stapel gefabriceerde datasets, waarvan de herkomst obscuur was. Waar de Commissie Drenth op grond van onregelmatigheden, onjuistheden (bijvoorbeeld incorrecte vrijheidsgraden en p-waarden) en uiterst onwaarschijnlijke (verdeling van) gegevens tot de conclusie kwam dat er in een aantal onderzoeken tijdens de Amsterdamse periode sprake was van ‘evidence of fraud’ heeft de heer Stapel zelf aangegeven zijn wijze van onderzoek in Amsterdam ‘grijs’ te noemen. Daar heeft hij zich naar eigen zeggen schuldig gemaakt aan wat hij thans ziet als wetenschappelijk ongeoorloofde praktijken, zoals het weglaten van condities en variabelen of het strategisch ‘draaien’ van groepen. Volgens de heer Stapel is zijn onderzoekspraktijk in Groningen van ‘grijs’ ‘zwart’ geworden, waarbij volgens hem in Tilburg de wijze van fingeren ‘steeds gekker en sneller en vreemder’ werd. Een uitvoerige beschrijving van de werkwijze van de heer Stapel is te vinden in de paragrafen 4.2 tot en met 4.5.

3.3 Meldingen van ernstige inbreuk op wetenschappelijke integriteit

Naast frauduleuze schendingen van wetenschappelijke integriteit, zijn er ook andere schendingen van behoorlijk wetenschappelijk onderzoek aangetroffen door de Commissies, die geanonimiseerd besproken worden in hoofdstuk 5.

Omdat promovendi nog in een leerproces zijn, kan hen een dergelijke verkeerde werkwijze, voor zover die al aanwezig was, niet in volle omvang persoonlijk aangerekend worden. Dit geldt echter niet op dezelfde wijze voor postdocs en stafleden. Van hen moet een zelfstandige, kritische en behoorlijke beoefening van de wetenschap verwacht worden; zij zijn ook aangesteld om de studenten en promovendi op te leiden en hebben de opdracht hen een kritische opstelling bij te brengen.

De Commissie Levelt heeft ten aanzien van een klein aantal personen geconstateerd dat er sprake was van dusdanige ernstige schending van wetenschappelijke normen dat de Commissie zich genooddaakt zag de namen van betrokkenen vertrouwelijk te melden aan de Rector Magnificus van Tilburg University.

3.4 Coauteurs

De Commissies hebben geen aanwijzingen gevonden dat coauteurs doelbewust met de heer Stapel hebben meegewerkt aan datavervalsing. Wel hebben de Commissies Levelt en Noort moeten concluderen dat een meer kritische houding verwacht had mogen worden van sommige coauteurs, bijvoorbeeld waar kopieergedrag in datasets terug te zien was in de resultaten in de tabellen van publicaties. Ook hebben de Commissies Levelt en Noort vastgesteld dat in een aantal publicaties sprake is van wetenschappelijke onzorgvuldigheid door coauteurs. Het gaat hierbij om gedrag dat niet aan de fundamentele regels van deugdelijk wetenschappelijk onderzoek beantwoordt. Bij wijze van voorbeeld: er werden diverse gelijksoortige experimenten uitgevoerd, waarbij er maar één of enkele de verwachte resultaten opleveren; deze werden dan ter publicatie aangeboden zonder nieuwe onafhankelijke replicatie en zonder vermelding van de niet succesvolle exemplaren. Ook werden er onvermeld en op twijfelachtige gronden grote aantallen proefpersonen uit de analyses gelaten of werden bepaalde (toetsings)uitkomsten verkeerd gerapporteerd of bewust achterwege gelaten. Dit wijst op een cultuur waarin wetenschappelijke integriteit niet hoog in het vaandel staat; in de cultuurparagraaf wordt hier nog uitvoeriger op teruggekomen.

3.5 Dissertaties Tilburg

De Commissie Levelt heeft met zekerheid vastgesteld dat ook in door de heer Stapel begeleide Tilburgse promotieonderzoeken gewerkt is met gefingeerde data. De Commissie heeft allereerst geen fraude vastgesteld in het proefschrift van dr. F. van Horen en van dr. C. Wiekens. Anderzijds staat vast dat alle hoofdstukken in de dissertaties van dr. Y.R. Avramova en dr. M.K. Noordewier berusten op gefingeerde data. In het proefschrift van dr. L. van den Broek is in één hoofdstuk fraude geconstateerd. Het proefschrift van mevrouw M.H.C. Meijers, waarvan de verdediging gepland was voor 7 oktober 2011, is op haar eigen initiatief bij het vernemen van de datafraude teruggetrokken. De Commissie heeft zich in dat geval beperkt tot het beoordelen van het artikel dat reeds gepubliceerd was.

De Commissie Levelt heeft zich vervolgens de vraag gesteld of, als uitvloeisel van deze bevindingen, dit consequenties zou moeten hebben voor een reeds verleende graad. Ze heeft hierbij de volgende maatstaf gebruikt: Wisten de betreffende promovendi dat de data gefingeerd waren of was er sprake van ‘culpa’, dat wil zeggen verwijtbare onwetendheid? In antwoord hierop heeft de Commissie geconstateerd dat in geen van de gevallen sprake is geweest van opzet of ‘culpa’. Deze promovendi-coauteurs zijn allen voornamelijk op zeer geraffineerde wijze om de tuin geleid en in belangrijke mate ook opgeleid in de hierboven kort aangegeven foutieve onderzoekscultuur. De promovendi hebben bovendien, naast de hoofdstukken waarin gebruik is gemaakt van gefingeerde data, ook onderzoeken uitgevoerd waarin zij wel degelijk zelf data hebben verzameld. Deze hebben echter niet tot publicaties geleid aangezien de bevindingen beduidend minder interessant waren dan de resultaten van de aangereikte data.

De Commissie is dan ook van mening dat dit in geen van de gevallen gevolgen zou moeten hebben voor een verleende graad. Wel dient uiteraard de onzuiverheid van de data gebruikt in deze dissertaties te worden bekend gemaakt, zodat mogelijk ten onrechte getrokken conclusies en gegeven interpretaties geen verdere effecten hebben.

3.6 Dissertaties Groningen

Na gesprekken met alle in Groningen gepromoveerde onderzoekers over hun vooraf of op basis van de dissertatie gepubliceerde artikelen/hoofdstukken is de Commissie Noort tot dezelfde conclusie gekomen als de Commissie Levelt. Van het kennis hebben van of bewust meewerken aan de fraude is bij deze promoti geen sprake geweest. Ook van ‘culpa’, zoals door de Commissie Levelt omschreven, kan bij deze Groningse junioronderzoekers niet gesproken worden. De Commissie Noort deelt dan ook het oordeel van de Commissie Levelt, dat de datafraude door de heer Stapel geen gevolgen zou moeten hebben voor een verleende graad van deze promoti.

Anders dan over de promoti zelf is het oordeel over de dissertaties gemengd. Voor de Commissie Noort is daarbij het eerste criterium geweest in hoeverre data door de onderzoekers zelf zijn verzameld, ingevoerd, geanalyseerd en geïnterpreteerd. Maar er werd wel degelijk in een aantal gevallen gebruik gemaakt van door de heer Stapel aangereikte of bewerkte datasets. Ook kwam het voor dat data in een bepaald stadium van interpretatie buiten bereik van de promovendus in handen van de heer Stapel waren, waarbij het voorkwam dat de data door de heer Stapel gewijzigd werden.

Géén fraude in bovenbedoelde zin heeft de Commissie Noort bij de proefschriften van dr. M.B. Braun-Ekker, dr. E. Kamans, dr. S.J. Ko, dr. J. Lammers, dr. H. Oldenhuis, dr. S.W. van der Velde kunnen vaststellen. In deze gevallen was de heer Stapel alleen om formele redenen promotor of staat vast dat de promovendus alle data van begin tot eind in eigen beheer had. Anders ligt het met de dissertaties van dr. Janneke F. Joly, dr. J. Grob, dr. S.A. Schwinghammer, dr. D. Trampe, dr. A. van den Bos, dr. M. Maringer en dr. L. Renkema. Uit de analyses van de Commissie Noort blijkt dat er bij deze dissertaties in een of meer hoofdstukken sprake is van door de heer Stapel uitgevoerde frauduleuze handelingen, zie paragraaf 3.1 en bijlage 5.

3.7 Gevolgen van de fraude

De Commissies hebben geconstateerd dat coauteurs en met name promovendi in ernstige mate gedupeerd zijn als gevolg van de door de heer Stapel gepleegde fraude. Het is de Commissies in de gevoerde gesprekken gebleken dat de consequenties voor de betrokkenen formeel en informeel van aard kunnen zijn. In formele zin worden de getroffen mensen belemmerd in hun carrière, onder andere bij verlengingen van tijdelijke contracten en bij aanvragen van subsidies. Gepromoveerd zijn met de heer Stapel als begeleider was altijd een aanbeveling voor zijn promoti; thans is het tegenovergestelde het geval. Op individueel niveau zijn sommige slachtoffers gedemoraliseerd. In informele zin vindt stigmatisering plaats die langdurig kan nawerken in de verdere loopbaan. Voor sommige gedupeerden zouden de gevolgen wel eens ingrijpender kunnen zijn dan zij op dit moment kunnen voorzien.

Niet alleen personen zijn getroffen, het imago van de wetenschap en in het bijzonder van de sociale psychologie, is door toedoen van de heer Stapel fors beschadigd. Andere slachtoffers zijn de universiteiten van Tilburg, Groningen en Amsterdam, die reputatieschade hebben geleden. Vooral kleinere instellingen zijn extra kwetsbaar als het gaat om internationale reputatie. Met name het psychologisch onderzoek in

Tilburg zal dan ook nog lang geconfronteerd worden met deze fraudezaak. In dit kader is het van belang dat het College van Bestuur van Tilburg University onmiddellijk een Commissie heeft ingesteld om deze fraudezaak tot op de bodem uit te zoeken.

Ook wetenschappelijke uitgevers die gepubliceerde artikelen moeten terugtrekken en onderzoeksfondsen wier subsidies frauduleus zijn gebruikt zijn slachtoffers, alsmede collega onderzoekers die door de aan de heer Stapel toegekende subsidies of ondersteuning uit de eerste geldstroom zelf deze subsidies of ondersteuning mogelijk zijn misgelopen. Daarnaast is ook buiten Tilburg en Groningen veel tijd en geld verspild aan het tevergeefs proberen te repliceren van dan wel verder te bouwen op de resultaten die de heer Stapel gepubliceerd heeft.

De Commissies hebben geconstateerd dat deze fraudezaak geleid heeft tot een grotere internationale bewustwording van problematische aspecten van de gangbare onderzoekspraktijk. De discussies over replicatie, data-archivering en de onderzoekscultuur in het algemeen worden met meer nadruk dan ooit gevoerd en zullen zeker de komende tijd geïntensiveerd worden. Dit rapport komt hierop terug in het hoofdstuk 'Aanbevelingen'.

4 Onderzoekscultuur: werkwijze en onderzoeksomgeving

4.1 Inleiding

Naast het vaststellen van de aard en omvang van de fraude, hadden de Commissies ook als opdracht na te gaan welke factoren hebben bijgedragen aan het feit dat deze omvangrijke fraude zich over zoveel jaren heeft kunnen uitstrekken. Een antwoord op die vraag is relevant voor het doen van aanbevelingen om herhaling te voorkomen. Die vraag is des te klemmender omdat een aantal aspecten van de werkwijzen en resultaten van de heer Stapel, ook zonder direct vermoeden van fraude, gezien hadden kunnen of zelfs moeten worden als ongebruikelijk en ongewenst. Bovendien zijn die merkwaardigheden soms, bij uitzondering, wel eerder opgemerkt, maar heeft dit geen verdere gevolgen gehad.

De Commissies hebben geconstateerd dat promovendi en coauteurs niet bewust meegewerkt hebben aan de fraude van de heer Stapel (paragraaf 3.4). Ook na intensieve bestudering van de publicaties en gesprekken met de coauteurs blijft de conclusie uit het interim-rapport dat promovendi en coauteurs niet medeplichtig zijn overeind. Wel hebben de Commissies moeten constateren dat de voorstelling van zaken door de heer Stapel over inrichting en loop van het onderzoek, de opmerkelijke, frauduleuze data en in het algemeen de bevindingen van de heer Stapel veelal kritiekloos geaccepteerd zijn, ook wanneer twijfel voor de hand had gelegen. Daarbij waren de coauteurs de eerst aangewezenen om de fraude te doorzien. Waarom gebeurde dat dan niet? In het volgende bespreken wij een aantal factoren binnen maar ook buiten de onderzoeksomgeving van de heer Stapel die dit falen van kritiek hebben bevorderd. Aan deze analyse ligt een veelheid van gesprekken ten grondslag die de Commissies met collega-onderzoekers van de heer Stapel, met de heer Stapel zelf, met bestuurders, promoti en coauteurs hebben gevoerd.

De eerste en belangrijkste groep factoren moet gezocht worden in de werkwijze en onderzoeksomgeving van de heer Stapel. Hij maakte misbruik van het vertrouwen dat de basis vormt van alle wetenschappelijke samenwerking. Zijn positie van groot prestige en macht gebruikte hij om zijn fraude uit te voeren en om elke mogelijke twijfel aan zijn werkwijze in de kiem te smoren. Deze aspecten komen in de volgende paragrafen aan bod. Hoewel hierin overal dezelfde tendensen te bespeuren zijn, zeker in Tilburg en Groningen, worden deze aspecten toch afzonderlijk per universiteit behandeld. De positie van de heer Stapel was verschillend binnen de drie instellingen waar hij werkte en zijn werkwijze hing daarmee samen.

De tweede groep factoren is gelegen in de algemene wijze waarop het onderzoek binnen de werkomgeving van de heer Stapel, maar ook daarbuiten uitgevoerd werd. Dit wordt besproken in hoofdstuk 5.

4.2 Algemeen

Vóór alles verdient aandacht dat wederzijds vertrouwen een essentieel onderdeel van elke collegiale samenwerking is, ook in de wetenschap. Dit is bij uitstek het geval voor een meester-gezelrelatie zoals die zich voordoet in de wetenschap (promotor-promovendus of senior-junior onderzoeker). De junior moet absoluut kunnen vertrouwen op de integriteit van de ‘meester’. Het is dit vertrouwensaspect, dat in de hele fraude-affaire een cruciale rol heeft gespeeld: een wetenschapper fraudeert niet; het is bijna onvoorstelbaar dat dit wel gebeurt.

De fraude en daarmee het misbruik van vertrouwen werden daarom ook niet opgemerkt omdat de fraude vaak geraffineerd voorbereid en uitgevoerd werd, in die zin dat er door de heer Stapel voor gezorgd werd dat het hele verloop van het onderzoek er realistisch uitzag. Bovendien opereerde hij ten aanzien van het uitvoeren van het onderzoek en het begeleiden van zijn promovendi erg solistisch. In de beschrijvingen van de Groningse en Tilburgse situatie worden hierover vele details gegeven.

In die meer gedetailleerde lokale beschrijvingen komt ook naar voren dat de heer Stapel een aanzienlijke machtspositie innam, in ieder geval binnen de Rijksuniversiteit Groningen en meer nog binnen Tilburg University. Aan de Universiteit van Amsterdam had hij al de reputatie van ‘Golden Boy’. Die ‘machtspositie’ maakt het zeker moeilijker fraude op te merken en aan te kaarten. De heer Stapel werd ook ervaren als een charismatisch leider: personen die men bewondert zal men niet snel van fraude verdenken en men zal eerder ‘alternatieve verklaringen’ voor eventuele merkwaardigheden bedenken dan fraude vermoeden. Als promotor had de heer Stapel bovendien meestal een zeer vriendschappelijke band met zijn promovendi. Ook dit bemoeilijkt een kritische opstelling door de promovendi ten aanzien van een mogelijk frauduleus handelen van de promotor. Tenslotte is het de Commissies uit diverse gesprekken gebleken dat de heer Stapel niet aarzelde zijn macht, prestige en charisma in te zetten om fraudeontdekking te voorkomen.

Deze algemene conclusies kunnen op veel punten meer concreet onderbouwd worden. Illustraties daarvan zijn te vinden in de weergaven van de lokale onderzoeken en gesprekken.

4.3 Werkwijze en Onderzoeksomgeving: Universiteit van Amsterdam

De Amsterdamse periode (1993/1999) is belangrijk omdat de heer Stapel daar als promovendus en junior onderzoeker in de praktijk het vak geleerd heeft. Waar de Commissie Drenth op basis van de resultaten, vermeld in de publicaties in een aantal artikelen, sterke aanwijzingen voor onregelmatigheden heeft aangetroffen, zou een verder onderzoek naar de ‘onderzoekscultuur’ en de gangbare onderzoekspraktijken in het toenmalige departement waar de heer Stapel werkte, van groot belang zijn. Een en ander is echter lang geleden. Het feit dat er mede daardoor voor geen enkele publicatie nog data of verder onderzoeksmateriaal beschikbaar waren en gesprekken met betrokkenen geen betrouwbare gedetailleerde inzichten meer konden bieden, kon de Commissie Drenth slechts tot algemene conclusies komen ten aanzien van de ‘cultuur’ inzake wetenschappelijke integriteit binnen de Faculteit Psychologie aan de Universiteit van Amsterdam.

Binnen de wetenschappelijke onderzoeksomgeving in Nederland bestond tot het einde van de vorige eeuw geen expliciete aandacht, laat staan een formele Code, voor de bestrijding of het voorkomen van wetenschappelijk wangedrag. Een en ander bleef meer impliciet en werd overgelaten aan de individuele onderzoekers. Dat was in de Faculteit Psychologie van de UvA niet anders. Dit betekent niet dat de Faculteit Psychologie het met de normen van wetenschappelijke integriteit niet zo nauw nam, maar eerder dat er geen gerichte aandacht en geen formele structuur bestond, gericht op de handhaving van deze normen. Daar komt bij dat er in die faculteit ook geen cultuur bestond van transparantie, delen van informatie, collegiale controle en gezamenlijke verantwoordelijkheid. Onderzoekers konden gemakkelijk hun eigen gang gaan. Zij werden niet op de vingers gekeken en deelden hun gegevens niet met andere leden van de verschillende onderzoeksgroepen en werden niet collegiaal getoetst. Die wat naïeve delegatie van vertrouwen en het nalaten van gedetailleerde controle op ruwe en gecodeerde gegevens, met name van de zijde van (co)promotoren, leden van leescommissies, werkgroepeliders en coauteurs, hebben een werkomgeving en werkcultuur gecreëerd waarin een individuele onderzoeker, als deze eens een poging deed de grenzen van integriteit te overschrijden, een te grote kans liep dat dit onopgemerkt bleef. Men vertrouwde op elkaars wetenschappelijke integriteit: geen feedback, geen waarschuwingen, geen correctie. Er ontbrak dus een cultuur waarin de wetenschappelijke integriteit een hoge prioriteit krijgt via expliciete aandacht, gedetailleerde controle, impliciete voorbeeldfunctie van senior medewerkers en een collectieve verantwoordelijkheid binnen onderzoeksgroepen, waarbij de leden elkaar corrigeren en scherp houden.

Anders dan in de (frauduleuze) schoolonderzoekingen in de latere, vooral Tilburgse periode, zijn de data uit de Amsterdamse periode minder gebaseerd op onderzoek bij middelbare scholieren, hoewel dit soort (school)data wel zijn gebruikt, onder andere voor enige artikelen die aan zijn proefschrift ten grondslag liggen, bijvoorbeeld hoofdstuk 3. In het merendeel van de gevallen zijn de gegevens afkomstig uit onderzoek op eerste of jongerejaars studenten (Amsterdam, Eindhoven) of undergraduate students (University of Michigan, University of Georgia), bijvoorbeeld hoofdstuk 5. Daarmee voldeden de eerstejaars psychologie aan de Universiteit van Amsterdam aan de verplichting zich een aantal uren als proefpersoon beschikbaar te stellen. Ook deden studenten soms tegen een vergoeding mee aan het onderzoek. In de betreffende jaren heeft inderdaad onderzoek met eerstejaars studenten plaatsgevonden, en de heer Stapel heeft daarin meegedaan.

De heer Stapel heeft vrijwel altijd geïsoleerd, ongecontroleerd en ongetoetst de dataverzameling/codering/bewerking voor zijn rekening genomen. Hij werd beschouwd als een zeer veelbelovend en talentvol wetenschapper, die opviel als een informeel leider onder zijn leeftijdgenoten.

4.4 Werkwijze en Onderzoeksomgeving: Rijksuniversiteit Groningen

De heer D.A. Stapel

Per 1 mei 2000 werd de heer Stapel, tot op dat moment KNAW-fellow aan de UvA, benoemd tot hoogleraar A op een Van der Leeuw-leerstoel Cognitieve Sociale Psychologie in de toenmalige faculteit PPSW, nu Faculteit Gedrags- en Maatschappijwetenschappen (GMW) aan de Rijksuniversiteit Groningen (RUG). Deze Van der Leeuw-leerstoelen waren bedoeld om jonge wetenschappers versneld naar gewone leerstoelen te laten doorstromen. De benoemingsadviescommissie voor de Van der Leeuw-leerstoel werd voorgezeten

door de sociaal-psycholoog prof.dr. A.P. Buunk en droeg de heer Stapel in een enkelvoudige voordracht voor. Men had zeer hoge verwachtingen van de kandidaat Stapel vanwege zijn grote internationale wetenschappelijke productie. Hij werd geacht inspirerend en stimulerend te zijn en kon als briljant, zo niet geniaal onderzoeker omschreven worden. Zijn c.v. gaf daar overigens aanleiding toe. Hij had twee doctoraal examens (Psychologie en Communicatiewetenschap) cum laude gehaald (UvA), was 1997 eveneens cum laude gepromoveerd in de sociale psychologie (waarvoor hij de ASPO Proefschrift Prijs kreeg), had de nodige tijd in het buitenland doorgebracht en kon 1999 de Jos Jaspars Early Career Award van de EAESP in ontvangst nemen. Zijn publicatielijst over de periode 1993-1999 was indrukwekkend. Op 13 maart 2000 meldde de heer Stapel dat hij de benoeming ‘met enthousiasme’ zou aanvaarden en hij inaugureerde op 11 september 2001 met een rede over de hoofdrol van de sociale psychologie in de sociale wetenschappen getiteld ‘De koningin is aan zet.’ Zes jaar later, op 28 mei 2006, schreef hij de decaan van de Faculteit GMW, dat hij zijn ontslag indiende om per 1 september 2006 aan de Universiteit van Tilburg te gaan werken. Hoewel hij naar eigen zeggen met plezier aan de RUG gewerkt had, meende hij ‘dat op dit moment van mijn leven de UvT beter bij mijn persoonlijke ontwikkeling en ambitieniveau past.’

De datum van zijn inaugurele oratie, 11 september 2001, was voor de heer Stapel belangrijk, maar om heel andere redenen dan waarom die in het collectieve geheugen gegrift staat. De heer Stapel verwijst in zijn oratie naar een conferentie 45 jaar eerder, waar de ‘cognitieve revolutie geboren was’. Voor sociaal-psychologen was de cognitieve benadering volgens de heer Stapel niet nieuw. Zij zagen de mens al ‘van nature als een verwerker en transformator van informatie’, maar de cognitieve revolutie bracht methodisch ‘verfijning en precisie’. Zijn programmatische ambitie reikte ver:

‘In het hart van het land der menswetenschappen woont de sociale psychologie. En dat maakt de sociale psychologie de koningin van dat wetenschapsland. Met in haar ene hand een scheermes (als symbool voor de precisie waarmee zij theoretische, retorische wildgroei wegsnijdt) en in haar andere hand een fototoestel (als symbool voor haar aandacht voor het alledaagse) vormt zij de ideale mediator tussen de maatschappij- en gedragswetenschappen. Sociale psychologie. De natuurkunde mag zichzelf dan tot koning hebben gekroond, de sociale psychologie is de koningin’ (Stapel, D.A. (2001)).

En ook de volgende quasi-verontschuldiging in zijn rede voor de kroning van de sociale psychologie als koningin der wetenschappen onderstreept die eerder dan dat ze die ondermijnt. Bij de ambitie van een koppeling tussen het alledaagse en het precieze, met ‘aandacht voor de echte, belangrijke problemen en aan de andere kant een methodiek die hoort bij de precieze analyses van de natuur- of cognitiewetenschappen’ is sprake van een bipolariteit. Het primaat van de theorie – en daarmee de ondergeschikte rol van de data? – werd al eerder door hem geformuleerd:

‘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)).

Dat ‘beschuldigen’ van het experiment deed hij in een veld waarin verschillende opvattingen bestonden met betrekking tot het primaat van de theorie dan wel de data, en vanuit een opvatting waarin de data volstrekt ondergeschikt zijn aan de theorie, kon worden geconstateerd dat lang niet altijd de vereiste nauwkeurigheid ten aanzien van de data werd nagekomen. ‘Dat is bij ons zo gebruikelijk’, viel een aantal keren in het kader van de gevoerde vraaggesprekken, wanneer bij de bespreking van de grenzen die in de omgang met de data

in acht moeten worden genomen, gebrek aan nauwkeurigheid met betrekking tot de data aan de orde kwam.

Alle geïnterviewden, dat wil zeggen coauteurs of in bredere zin leerlingen van de heer Stapel en de meeste collega's, uitten zich – ondanks woede en/of verdriet – bijzonder lovend over zijn capaciteiten. Dat geldt vooral de copromotor, die een stap terug deed achter zijn opkomende ster, tot en met de student- of onderzoeksassistent die er trots op was deel uit te maken van het team van de heer Stapel. De erkenning, die hij ook na zijn hoogleraarsbenoeming kreeg, was niet gering blijkens onder meer een Pionier-subsidie van NWO (2002), gelden vanuit de Breedtestrategie en andere subsidies. Aan geld voor assistenten, promovendi, projecten was geen gebrek. De media wisten hem te vinden en soms ook omgekeerd. Naast de lof voor dynamiek, ideeënrijkdom, toegankelijkheid en werkkraft, viel het woord 'ambitieuw' het meest. En dat in een overtreffende trap.

De groep Stapel

Het onderzoek van de faculteit PPSW (GMW) vond plaats in relatief kleine verbanden met een grote mate van autonomie. Binnen deze 'archipel' nam de groep Stapel een eigen positie in.

In dit kader maakten meerdere geïnterviewden, waaronder prof. Buunk zelf, gewag van een moeizame relatie c.q. verwijdering tussen prof. Buunk en de heer Stapel en het, als gevolg daarvan, ontbreken van een 'meekijkende' hoogleraar, zowel bij eigen onderzoek als onderzoek van promovendi. In het gesprek met de Commissie heeft de heer Stapel aangegeven zelf de relatie met prof. Buunk niet als moeizaam te hebben ervaren. Voor zover sprake was van een verwijdering, lag dat volgens hem aan een verschil in inhoudelijke belangstelling.

De promovendi-groep van de heer Stapel had weliswaar een zekere eigen identiteit en trots, maar het voert duidelijk te ver om te concluderen dat sprake was van een gesloten groep promovendi: promovendi gingen niet alleen met promovendi van de eigen groep (Stapel) om, maar ook met die van andere groepen.

Een uitzondering daargelaten, hadden de promovendi van de heer Stapel altijd een tweede begeleider. Het accent bij de rol van de tweede begeleider lag op het schrijfproces en inhoudelijke betoog; voor de data was weinig aandacht. Er bestond bij de tweede begeleiders geen argwaan jegens de kwaliteit van gebruikte data. Binnen het raamwerk van zijn eigen onderzoekslijn en eigen promovendi werd de heer Stapel als onbetwiste leider erkend. Binnen dat raamwerk zagen zij slechts een volgende en zijdelingse rol voor zich weggelegd.

De groep van de heer Stapel hield een talentenklasje voor gemotiveerde studenten sociale psychologie, waarvoor regelmatig bijeenkomsten werden gehouden waar onderzoek werd besproken. Dit clubje studenten was een belangrijke wervingsbron voor promovendi. Veel van zijn promovendi nam de heer Stapel werk uit handen door een student-assistent in te schakelen bij de data-invoer. Vaak zorgde de promovendus wel zelf voor de dataverzameling, maar de ingevoerde data werden door de student-assistent rechtstreeks aan de heer Stapel gegeven waarbij het voorkwam dat deze gewijzigd werden voordat ze naar de promovendus gingen. Ook werd ten minste een keer, naar aanleiding van tijdschriftcommentaar, de oorspronkelijke dataset door de heer Stapel teruggenomen ter nadere analyse. Deze was aangevuld, zo bleek later, met een tiental nieuwe observaties, zodat alle hypothesen nu wel fraai bevestigd werden. De hulp bij de data-invoer en -verwerking werd door promovendi als een luxe ervaren (niet per se als vanzelfsprekend) omdat het veel tijd scheelde. Het leidde soms ook tot enige jaloezie onder promovendi, die in een aantal gevallen nog werd aangewakkerd door de vermeende kwaliteit van de data. Overigens werd het bovenop

de data zitten door de heer Stapel ook wel gewaardeerd als een grote betrokkenheid bij het onderzoek van zijn promovendi. In zijn gesprek met de Commissie Noort merkte de heer Stapel in dit verband op dat hij er zelf van overtuigd was zijn promovendi te helpen. Bij het verzamelen van de data ging het naar eigen zeggen om het vervelendste deel van het onderzoek, waar promovendi zo snel mogelijk doorheen geholpen moesten worden. De paradox ‘helpen’ door vervalsen is hem pas later duidelijk geworden.

De heer Stapel werd door zijn onderzoeksgroep in het algemeen als een charismatische, aimabele onderzoeksleider gezien (andere woorden die respondenten gebruikten, zijn onder andere: ‘briljant’, ‘creatief’, ‘ijverig’, ‘gedreven’, ‘ambitieuus’). Hij was echt de baas, zo komt uit de vraagesprekken duidelijk naar voren. Bekend was dat statistiek noch zijn sterkste punt was noch zijn grootste belangstelling had. Zo nodig legden promovendi hun vragen van statistische aard daarom aan binnen de faculteit beschikbare statistisch deskundige onderzoekers voor. De heer Stapel had daar geen problemen mee. Dat de data die promovendi in handen hadden er in een aantal gevallen zo fraai uitzagen, werd veelal toegeschreven aan sterke theorievorming. Bovendien waren er niet louter succesverhalen: er waren ook promovendi met ‘slechte’ data, die daarom geen mooie onderzoeksresultaten wisten te boeken.

Ten tijde van de aankomst van de heer Stapel te Groningen had Groningen een inhaalslag te maken waar het ging om dataverzameling (het is in het algemeen geen sinecure om aan voldoende gegevens te komen). Die inhaalslag resulteerde in de totstandkoming van het datalab. Regelmatig gehouden labdagen (testdagen) waarin studenten werden bevraagd (verplichte deelname) garandeerden een regelmatig verkrijgen van nieuwe onderzoeksdata. Aan de totstandkoming van het lab te Groningen hebben meerdere personen bijgedragen. Het fenomeen van de testdagen had de heer Stapel uit zijn Amsterdamse tijd meegenomen. Duidelijk is dat lab en testdagen voor het onderzoek van de heer Stapel en zijn groep van vitaal belang waren. Niet alleen had de heer Stapel nu de onderzoeks ideeën, maar ook de noodzakelijke onderzoeksdata. Overigens was het, naast zijn uitstekende reputatie en ideeënrijkdom, juist ook zijn mogelijkheid om aan voldoende data te komen, waardoor de heer Stapel een aantrekkelijke onderzoekspartner was voor buitenlandse onderzoekers, aldus een aantal geïnterviewden.

Op in beginsel wekelijks gehouden labmeetings werd onderzoek besproken, waarbij tevens aandacht aan methodologie werd besteed. Op dit punt was er echter geen duidelijk verhaal. Protocolen over bijvoorbeeld het verzamelen van data (waaronder standaarden voor vragenlijsten) of over onderzoeksrapportage ontbraken in de groep Stapel. De promovendi in de groep van de heer Stapel werden niet vertrouwd gemaakt met vaste en eenduidige standaarden. Bij dit alles heeft de heer Stapel aangetekend dat er te Groningen veel over methodologie en het omgaan met de data werd gesproken en over dit laatste wel degelijk afspraken werden gemaakt. Volgens hem week de Groningse situatie op dit punt niet af van wat elders gebruikelijk was. Hij onderstreepte het gebrek aan (vaste en eenduidige) standaarden; maar dat was evenmin (verre van) een lokale kwestie volgens de heer Stapel.

Een enkele keer hebben promovendi in informeel verband uitgesproken dat sommige data te mooi leken om waar te zijn, aldus een paar geïnterviewden. De leiderschapspositie van de heer Stapel, de erkende kwaliteit van diens hypothesevorming en het gegeven dat lang niet alle data fraai waren, kunnen volgens hen mede verklaren waarom er geen serieuze twijfel omtrent datafraude is gerezen. Een enkele keer ook, is in een daarop gericht gesprek door een promovendus twijfel over de kwaliteit van de data geuit tegenover een naaste collega van de heer Stapel (dus niet bij de facultaire vertrouwenspersoon voor promovendi). Omdat

het slechts om een vermoeden ging, bleef het daarbij. Gelet op de ernst van het betrokken verwijt werden sterke aanwijzingen noodzakelijk geacht om verder actie te ondernemen.

De Commissie Noort meent dat het gesignaleerde eilandenrijk van vrij kleine, zelfstandige onderzoeksgroepen niet bijzonder voor alleen de Groningse situatie is. Indertijd was sprake van de opkomst van tenure tracks (loopbaan op basis van individuele prestaties met regelmatige beoordeling daarvan) en ruim baan voor talent. Daarbij past dat individuen eigen onderzoeklijnen ontwikkelen. De Commissie Noort heeft geconstateerd dat de heer Stapel weinig tegenspel van senior-onderzoekers van zijn onderzoeksgroep kreeg. Ofschoon de groep Stapel een eigen identiteit had, kan niet duidelijk worden geconcludeerd dat er sprake was van afscherming van onderzoekers van andere groepen. Merkwaaardig is de bijzondere aanpak van de heer Stapel bij de data-invoer en -verwerking. Dit heeft in de groep Stapel amper aanleiding gegeven tot twijfel omdat promovendi daar zelf voordeel van hadden. De Commissie Noort acht de concrete verdenking die een promovendus tegenover een naaste collega van de heer Stapel heeft geuit opmerkelijk, maar heeft onvoldoende inzicht kunnen krijgen in het besprokene en de achtergrond daarvan om de stellige conclusie te kunnen trekken dat het signaal aanleiding had moeten vormen voor nader onderzoek.

4.5 Werkwijze en Onderzoeksomgeving: Tilburg University

De heer Stapel trad als hoogleraar in dienst bij Tilburg University per 1 september 2006 ten behoeve van het instituut TIBER in oprichting, een samenwerkingsverband tussen de faculteit Sociale Wetenschappen en de Economische faculteit. Op 1 september 2010 werd de heer Stapel benoemd tot decaan van de faculteit, welke dan de naam Tilburg School of Social and Behavioral Sciences draagt. De faculteit bestond in 2010 uit 9 departementen, waaronder het Departement Sociale Psychologie.

Voor iedereen die wetenschappelijk samenwerkte met de heer Stapel, in het bijzonder ook de collega's en junioren binnen het Departement Sociale Psychologie van Tilburg University, was het vertrouwen in de wetenschappelijke integriteit van de heer Stapel absoluut. Het laatste wat collega's, medewerkers en studenten konden vermoeden was dat uitgerekend de wetenschappelijke coryfee van het Departement, en decaan van de Faculteit, en de man die zelf de cursus wetenschapsethiek gaf voor zijn Departement, dat vertrouwen systematisch zou schenden. Velen zagen hem als charismatisch leider met grote toewijding aan medewerkers en studenten. Slechts weinigen hebben aan datafraude kunnen, durven of willen denken.

Het was ook niet direct eenvoudig, gezien zijn ingenieuze aanpak, de fraude te achterhalen. De meest voorkomende werkwijze van de heer Stapel was globaal de volgende. Hij ontwierp samen met een andere wetenschapper (doorgaans een research master student, een promovendus of postdoc, maar soms ook een senior collega), in intensieve één-op-één interactie, een theorie die intelligent en creatief aansloot op een onderzoeksvraag, welke samen met de partner werd geformuleerd. Ter toetsing van die theorie werden ingenieuze experimentele manipulaties ontworpen en tot in detail voorbereid en op papier uitgewerkt. De heer Stapel gaf daarbij blijk van grote literatuurkennis en van een aanzienlijke vaardigheid in het ontwerpen van de gepaste experimentele manipulaties. De materialen voor het experiment, zoals precieze omschrijvingen van de experimentele set up, gedetailleerde instructies voor de proefleiders ter plekke, de stapels vragenlijsten, experimentele benodigdheden zoals plaatjes, belonend snoepgoed, etcetera, werden,

na zorgvuldige bespreking, door de wetenschappelijke partner voorbereid in de benodigde hoeveelheden. Tevens produceerde de laatste vaak, op verzoek van de heer Stapel, een tabel met de verwachte empirische uitkomsten van alle te toetsen hypothesen.

Tot zover was alles in orde. Maar dan volgde er een volstrekt fictieve fase. De experimenten werden (zogenaamd) uitgevoerd onder de volledige supervisie van de heer Stapel alleen. De heer Stapel had, naar eigen zeggen, uitstekende contacten met een groot aantal onderwijsinstellingen in het land. Die waren steeds weer bereid om in goed overleg met hem persoonlijk zulke onderzoeken uit te voeren, soms geholpen door (zogenaamd) betaalde research assistenten. Ter compensatie voor de inspanningen van de scholen gaf de heer Stapel er (zogenaamd) voordrachten en schonk hij (zogenaamd) de betreffende scholen van tijd tot tijd computers en beamers. Uit de verklaring van de heer Stapel aan de Commissie Levelt blijkt, dat de vragenlijsten en extra materialen die achter in zijn auto waren geplaatst (zoals bijvoorbeeld chocoladerepen die bij een experiment gebruikt zouden worden) niet naar scholen gebracht werden maar in een container gedumpt werden. De heer Stapel ging dan zelf met een lege vragenlijst achter zijn computer zitten om een bijbehorend databestand aan te maken. Naar zijn partners (promovendi, coauteurs, collega's) toe gaf hij de indruk dat op die scholen data werden verzameld door (onbekende) assistenten, die de data ook verwerkten en codeerden. De aldus 'verkregen' gegevens werden dan rechtstreeks aan de heer Stapel gegeven, nooit aan de partners. Al die 'inspanningen' resulteerden doorgaans enkele weken later in een (gefingeerd) databestand dat de partner in zijn geheel ter beschikking werd gesteld voor nadere analyse, of ook direct in de vorm van tabellen met de benodigde gemiddelden, standaardfouten, betrouwbaarheden, toetsingsuitkomsten, etcetera. De partner kon direct aan het schrijven van het artikel beginnen, al dan niet in verdere intensieve samenwerking met de heer Stapel.

Deze werkwijze beslaat het grootste deel van de door de Commissie Levelt aangetroffen fraude. Er waren ook wel varianten. Soms raakte de heer Stapel in gesprek met een collega over een interessant onderzoek waar die collega mee bezig was en zei hij te beschikken over een oude dataset, ooit eerder verzameld, waar hij nog niets mee gedaan had, maar die perfect paste bij wat de collega nodig had. Deze fictieve dataset werd dan vrij snel ter beschikking van de collega gesteld, die daar een mooi artikel van kon maken met uiteraard de heer Stapel als coauteur. Ook vertelde hij wel in Groningen of elders dat hij data verkregen had in Tilburg van het sociaal-psychologisch lab aldaar of er voor kon zorgen dat die data er vanuit Tilburg kwamen. In Tilburg kon het omgekeerde richting Groningen gebeuren. Ook heeft de Commissie Levelt een onderzoek aangetroffen dat werkelijk uitgevoerd was, waarvan de ruwe data aan de heer Stapel werden gestuurd en waarbij de promovendus een paar weken later een tabel met de (mooier gemaakte) uitkomsten toegestuurd kreeg. En tenslotte werden onderzoekingen ook in hun geheel verzonden en solo 'uitgevoerd', bijvoorbeeld een onderzoek op een treinstation en met onbekende onderzoeksassistenten. Hierbij schreef de coauteur enkel mee aan het tijdschriftartikel; de dataset werd nooit gedeeld.

In discussies bij labmeetings werden kritische vragen in de richting van de dataverzameling niet op prijs gesteld. De heer Stapel was daarbij altijd duidelijk in zijn communicatie wanneer men moest stoppen met het stellen van vragen. Ook hebben diverse promovendi in gesprekken met de Commissie Levelt aangegeven dat de heer Stapel zijn machtspositie misbruikte om hen de mond te snoeren. Bij een research master studente die verdachte patronen vond in de data gaf de heer Stapel aan: 'Als je hier aangenomen wilt worden moet je laten zien dat je iets af kan maken en de resultaten gewoon opschrijven.' Bij een promovenda uit een andere vakgroep gebeurde het dat de kamer opgeruimd diende te worden. De heer

Stapel zag dozen met ingevulde vragenlijsten staan en wilde dat deze weggegooid werden. De promovenda gaf aan dat ze deze vijf jaar moest bewaren. De heer Stapel vertelde haar dat dit niet hoefde en voegde nog toe: ‘Als je wilt frauderen lukt het toch wel.’

In de wijze van begeleiden van zijn promovendi trof de Commissie Levelt nog een opmerkelijk aspect aan. Aan de ene kant had de heer Stapel een intensieve één-op-één werkwijze met de jonge onderzoekers en vele promovendi beschouwden hem als een persoonlijke vriend. Men kwam bij hem thuis, had gezamenlijke diners, ging gezamenlijk naar het theater etcetera. Aan de andere kant echter was daar de dreiging als er kritische vragen gesteld werden. De promovendus werd dan duidelijk gemaakt dat dit gezien werd als een gebrek aan vertrouwen en dat er geen vragen gesteld moesten worden. Juist de feitelijke nauwe relatie met de heer Stapel maakt het voor een junior onderzoeker dan moeilijk hier iets anders in te zien dan goed bedoelde opbouwende kritiek van de senior partner.

In een gesprek van de Commissie Levelt met de heer Stapel is met hem zijn hierboven beschreven, door de Commissies gereconstrueerde werkwijze besproken. De heer Stapel heeft in dat gesprek aangegeven zijn precieze werkwijze bij de datafabricage niet nader te willen toelichten omdat dat moeilijk is voor hem. Hij gaf alleen aan dat er geen systeem zat in de datafraude en dat hij vlug dingen veranderde omdat hij het zelf eigenlijk ook niet wilde weten. Wel gaf hij aan zich niet te herkennen in het beeld dat van hem geschetst werd als een manipulator om zijn fraude mogelijk te maken.

Veel van genoemde merkwaardigheden in zijn werkwijze (en uitkomsten) zijn pas achteraf uitdrukkelijk naar voren gekomen met de kennis van nu. Maar er zijn ook herhaaldelijk mensen geweest die het een of ander direct gezien hebben. Een aantal heeft het bij een geïsoleerde gedachte gelaten, omdat ze niet konden geloven dat er gefraudeerd werd of zich te machteloos voelden ten opzichte van de heer Stapel. Maar er zijn ook anderen geweest die er uitvoeriger aandacht aan hebben besteed en er ook met anderen overleg over hebben gepleegd. Waarom heeft dit laatste niet tot een eerdere ontdekking geleid?

Eén factor heeft in Tilburg wellicht bijgedragen aan vertraagde onderkenning van vermoedens van fraude. Gedurende het laatste jaar hebben een zestal jonge onderzoekers twijfels geuit over de wetenschappelijke integriteit van de heer Stapel. Aan wie uit je die twijfels? Volgens de regels van het Landelijk Orgaan Wetenschappelijke Integriteit (LOWI) dient dit te worden gemeld aan een door de universiteit aangestelde vertrouwenspersoon. In afwijking van wat het LOWI voorschrijft (‘De functie van vertrouwenspersoon is onverenigbaar met die van lid van het (College van) Bestuur, faculteitsdecaan, directeur van een onderzoekschool of onderwijs- of onderzoekinstituut’), was in Tilburg de Rector Magnificus de aangewezen vertrouwenspersoon. Dat is, zeker voor een student of jonge promovendus, hoog op en ver weg. De Rector Magnificus is een van de machtigste mensen binnen de universiteit. Daar ga je niet heen, zo is de voor de hand liggende gedachte, als je je vermoedens niet keihard kunt maken. Daar kwam bij dat de heer Stapel vanaf zijn aanstelling en als decaan zich verzekerd wist van aanzienlijke steun door het toenmalige College van Bestuur. Zo blijkt uit een financieel onderzoek van de Tilburg School of Social and Behavioral Sciences (TSB) dat de heer Stapel als hoogleraar al een bijzondere positie had in de Faculteit. Er werden hem rond zijn aanstelling een aantal toeslagen en vergoedingen toegekend door het College van Bestuur die zo uitzonderlijk waren voor TSB dat deze alleen konden verlopen via de Economische Faculteit. Verder werden zijn declaraties niet op de gebruikelijke wijze gecontroleerd, terwijl daar wel degelijk aanleiding toe was. ‘Hij werd niet of nauwelijks ter verantwoording geroepen.’ (financieel rapport TSB, 9 juni 2012).

Dat werd onder zijn eigen decanaat niet beter. De heer Stapel werd uitermate ruim gefaciliteerd en als onaantastbaar behandeld. Dergelijke signalen ten aanzien van de bijzondere positie van de heer Stapel werden voortdurend afgegeven en drongen uiteraard door tot de ‘facultaire werkvloer’.

De opdracht van de Commissie houdt mede in dat aandacht besteed moeten worden aan de vraag welke cultuur er bestond die de fraude heeft gefaciliteerd en in stand heeft gehouden. Dit is van belang om ‘aanbevelingen te formuleren om herhaling te voorkomen’. In de loop van het onderzoek en met name in de laatste fase bleek de Commissie dat zelfs na zijn ontslag de Faculteit de facilitering van de heer Stapel niet geheel gestopt heeft en een informatiestroom met zeer vertrouwelijke elementen op gang heeft gehouden. De Commissie heeft hiervan melding gedaan aan het huidige College van Bestuur. Het College heeft naar aanleiding hiervan direct een onderzoek ingesteld. Het is niet aan de Commissie om vooruit te lopen op de uitkomsten van dit nog lopende onderzoek.

In die context gaat iemand, bij vermoeden van fraude, in eerste instantie naar een naaste collega (hoogleraar of ander stafflid), in de hoop dat die de juiste vervolgstappen weet te zetten. En ook dat is niet zonder problemen, want wat voor relatie hebben die zelf met de vermeende fraudeur? Het wordt klokkenluiders veel gemakkelijker gemaakt wanneer er een onafhankelijke universitaire vertrouwenspersoon is voor wetenschappelijke integriteit, die geen onderdeel vormt van de bestuurslijn.

Er zijn in 2010 en 2011 drie meldingen verricht aan leden van de wetenschappelijke staf binnen de psychologie. De eerste twee hebben in eerste of tweede instantie niet geleid tot follow-up. Daarin heeft de haast onaantastbare positie van de heer Stapel wellicht een rol gespeeld. De derde melding, aan de departementsvoorzitter en buitengewoon zorgvuldig voorbereid door een drietal jonge en zeker in hun positie kwetsbare klokkenluiders, is wel direct professioneel opgepakt met het bekende resultaat.

Ook bij collega-hoogleraren zijn gedurende het laatste jaar tweemaal verdenkingen gerezen ten aanzien van door de heer Stapel aangereikte data. Aan die verdenkingen is geen follow-up gegeven. De Commissie concludeert dat de drie jonge klokkenluiders meer moed, alertheid en speurzin hebben getoond dan zittende hoogleraren.

5 De Onderzoekscultuur: gebrekkige wetenschapsbeoefening

5.1 Inleiding

Naast de genoemde factoren in hoofdstuk 4 is een andere belangrijke groep factoren ter verklaring van omvang en duur van de fraude gelegen in de algemene wijze waarop het onderzoek binnen de werkomgeving van de heer Stapel, maar ook daarbuiten uitgevoerd werd. Het gaat om een meer algemeen falen van de wetenschappelijke kritiek en om een onderzoekscultuur die te zeer gericht is op het onkritisch vaststellen van het eigen gelijk en, zoals uit de gesprekken met de Commissies bleek, op het vinden van interessante maar theoretisch oppervlakkige ad hoc resultaten. De methodologisch *sloppy science* die hieruit ontstaat is op grote schaal in het werk van de heer Stapel en zijn medewerkers aangetroffen en wordt in dit hoofdstuk verder besproken. In de slotparagraaf wordt ingegaan op het falen op alle niveaus van de wetenschappelijke reviewprocedures.

De Commissies hebben indertijd besloten tot de uitgebreide bestudering van al het onderzoeksmateriaal, beschikbare data en publicaties, en hebben uitvoerige gesprekken met betrokkenen gevoerd om een beter inzicht te krijgen in de omvang en aard van de fraude, en hoe deze mogelijk is geweest in de onderzoeksomgeving waarin de heer Stapel, zijn promovendi en naaste collega's werkten. In eerste instantie ging het onderzoek door de statistici van de dissertaties en publicaties om een verdere, betere en definitievere vaststelling van de fraude. De statistici en de Commissies hebben de dataverzameling, -codering, -analyse en -rapportage aan een kritisch onderzoek onderworpen.

De Commissies konden echter niet hun ogen sluiten waar bleek dat basisregels van integer wetenschappelijk onderzoek werden geschonden. Steeds meer moesten de Commissies tot de conclusie komen dat er, ook zonder strikte fraude, in algemene zin sprake was van een cultuur waarin op slordige, selectieve en niet-kritische wijze met onderzoek en data werd omgegaan. Het ging niet om kleine 'normale' onvolkomenheden in de statistische bewerkingen of opzet en uitvoering van het experiment, maar om schendingen van fundamentele regels van ordentelijk wetenschappelijk onderzoek die de conclusies van het onderzoek zeer ernstig kunnen beïnvloeden. Hierin ligt volgens de Commissies een deel van de verklaring waarom de fraude niet eerder is opgemerkt.

5.2 Generaliseerbaarheid van de bevindingen van lokale naar nationale en internationale cultuur

De ontdekking van de methodologische tekortkomingen, de 'bijvangst' van dit onderzoek, riep wel voor de Commissies de cruciale vraag op of de genoemde onderzoekscultuur, die hieronder in meer technisch detail wordt beschreven, ook algemene opgang maakt binnen het vakgebied van de sociale psychologie, nationaal en internationaal. Is het wellicht zo dat in het algemeen aan de werkwijze die binnen deze discipline gebruikelijk is aspecten kleven die uit het oogpunt van wetenschappelijkheid en wetenschappelijke integriteit als onjuist moeten worden aangemerkt?

In dit verband is het van belang dat de Commissies alle publicaties van de heer Stapel onderzocht hebben. Dit is weliswaar een groot aantal maar ze vormen uiteraard geen aselechte steekproef van publicaties uit de sociale psychologie. De Commissies kunnen derhalve op grond daarvan geen uitspraken doen over de sociale psychologie als zodanig. Het zou echter ook weer veel te simpel zijn om de hieronder beschreven bevindingen rond de onderzoekscultuur als slechts een lokale aberratie te karakteriseren. Daarvoor heeft de heer Stapel in te veel verschillende plaatsen en in te veel verschillende hoedanigheden gewerkt: als beginnend onderzoeker aan de Universiteit van Amsterdam, als beginnend hoogleraar in Groningen en als coryfee in Tilburg. Daarvoor heeft hij ook met te veel mensen van allerlei universiteiten in binnen- en buitenland samengewerkt. Bovendien heeft hij gepubliceerd in vrijwel alle gerenommeerde internationale tijdschriften op zijn vakgebied. Zeer zelden werden zijn al te fraaie uitkomsten onderworpen aan ernstige twijfel, zelfs niet in de promotiecommissies van ‘zijn’ promoties, zelfs niet in de internationale review procedures en zelfs niet waar de fraude uitermate zichtbaar was. Dit alles tezamen versterkt het beeld van een internationale onderzoeksgemeenschap waar de heer Stapel, zijn promovendi en naaste collega’s deel van uit maakten en waarvan de onderzoeksmethodologische gebruiken en bijbehorende normen en waarden gedeeld werden door de heer Stapel c.s.

Een duidelijk teken is ook dat een aantal coauteurs die de analyses wel zelf uitgevoerd hadden, en niet allen afkomstig uit de school van Stapel, de ernstige en minder ernstige schendingen van een behoorlijke wetenschappelijke werkwijze in gesprekken verdedigden met: zo heb ik dat geleerd in de praktijk, zo doet iedereen dat in mijn onderzoeksomgeving, en iedereen met wie we praten op internationale congressen. Uiteraard suggereren de Commissies hiermee niet dat ondeugdelijke onderzoekspraktijken gemeengoed zijn binnen de sociale psychologie. De Commissies kunnen of willen geen uitspraken doen over de sociale psychologie in het algemeen, al menen ze wel dat er voldoende redenen zijn voor de sociale psychologie in binnen- en buitenland om op grond van de bevindingen in dit rapport een eigen grondig onderzoek in te stellen naar de gang van zaken in het vakgebied. Hierbij kunnen de paragrafen 5.3 tot en met 5.6 wellicht diensten bewijzen. In de ‘Aanbevelingen’ zal dit verder opgenomen worden. De Commissies oordelen zeer positief over diverse initiatieven die op dit gebied reeds zijn genomen.

5.3 Verificatiebias en ontbrekende replicaties

Eén van de meest fundamentele regels van wetenschappelijk onderzoek is dat het onderzoek zo moet worden ingericht dat feiten die de onderzoekshypothesen zouden kunnen weerleggen minstens even veel kans krijgen in het onderzoek naar voren te komen als feiten die de onderzoekshypothesen bevestigen. Schendingen van deze grondregel, zoals het een aantal malen herhalen van een experiment tot het uitkomt of het weglaten van onwelgevallige onderzoekspersonen of resultaten, leiden als vanzelf tot een bevestiging van de onderzoekshypothesen van de onderzoeker en maken de hypothesen in essentie immuun voor de feiten.

In de grote meerderheid van de onderzochte publicaties bleek er sprake te zijn van het hanteren van procedures die leiden tot wat hier aangeduid wordt als verificatiebias. Er zijn niet veel artikelen waarin alle hieronder aangegeven schendingen van behoorlijke wetenschappelijke werkwijze tegelijkertijd aangetroffen werden. Aan de andere kant: er zijn ook nauwelijks publicaties gevonden waarin niet een of meer van deze strategieën gehanteerd werden.

Bij verificatiebias gaat het niet om de ‘gebruikelijke’ publicatiebias. Met dit laatste wordt bedoeld dat

negatieve of zwakke bevindingen die de theoretische verwachtingen niet duidelijk bevestigen, maar wel in goed uitgevoerd onderzoek verkregen zijn, doorgaans toch hun weg niet vinden naar de tijdschriften. Dit in tegenstelling tot 'positieve' resultaten. Met verificatiebias wordt iets ernstigers bedoeld: de onderzoeksprocedures worden zo gehanteerd dat negatieve resultaten op een of andere wijze 'onderdrukt' worden. De verificatiebias komt in de bestudeerde publicaties op uiteenlopende wijzen tot stand, waarvan de belangrijkste hier puntsgewijs genoemd worden.

- Een experiment levert niet de verwachte statistisch significante resultaten. Het experiment wordt herhaald, vaak met kleine wijzigingen in de manipulatie of andere omstandigheden en men rapporteert dan de uitkomst van alleen het experiment dat wel de verwachte resultaten oplevert. Het is niet duidelijk waarom de aangebrachte veranderingen theoretisch wel de verwachte resultaten zouden moeten opleveren. In het artikel wordt geen melding gemaakt van deze explorerende werkwijze; men doet het voorkomen alsof er via een eenmalig experiment een toetsing heeft plaats gevonden van de a priori verwachtingen. Het moge duidelijk zijn dat, zeker bij de doorgaans geringe aantallen proefpersonen, een dergelijk gebruik van het experiment gemakkelijk kan leiden tot een opeenstapeling van toevalsbevindingen. In dit verband is het ook saillant dat uit het materiaal van sommige studies blijkt dat er meerdere versies van vragenlijsten gebruikt zijn, waarbij de onderzoekers niet meer wisten welke versie in het artikel was gehanteerd.
- Een variant op voorgaande werkwijze: een bepaald experiment leidt niet tot statistisch significante verschillen tussen experimentele en controlegroep. Men vergelijkt dan de experimentele groep met een controlegroep uit een ander experiment - 'het zijn toch allemaal gelijke toevalsgroepen' - en vindt zo wel de gezochte significante verschillen. Ook dit wordt niet als dusdanig vermeld in het artikel.
- Het verwijderen van experimentele condities. In een experiment heeft de experimentele manipulatie bijvoorbeeld drie waarden. Elk van deze condities (zoals drie verschillende kleuren van het overigens identieke stimulus materiaal) moet een bepaald specifiek verschil opleveren voor wat betreft de afhankelijke variabele ten opzichte van de andere twee. Twee van de drie condities gedragen zich volgens de onderzoekshypothesen, een derde niet. Zonder enige vermelding in het artikel wordt die derde conditie uit het artikel gelaten, zowel in theoretisch opzicht als in de resultaten. Verwant hieraan is de aangetroffen verificatieprocedure waarbij de experimentele condities bepaalde effecten zouden moeten hebben op verschillende afhankelijke variabelen. Men rapporteert alleen de effecten op die afhankelijke variabele(n) waar de hypothesen uitkomen, meestal zonder vermelding van de niet-significante effecten op de andere afhankelijke variabelen en zonder nadere verklaring.
- Het samenvoegen van data uit verschillende experimenten. Het bleek zowel uit diverse datasets als uit de gesprekken met de coauteurs dat men vrij selectief en vooral ook achteraf de data van verschillende experimenten bij elkaar voegde om zo tot grotere aantallen en wel significante resultaten te komen.
- Onderzoeksbevindingen worden zonder vermelding in het artikel op slechts een deel van de onderzoekspersonen gebaseerd. Enerzijds werden 'outliers', extreme scores op een meestal afhankelijke variabele, uit de analyse verwijderd bij niet-significante resultaten. Door deze eliminatie neemt de variantie van de afhankelijke variabele af en neemt de kans op een 'statistisch significante' bevinding toe. Er kunnen goede redenen zijn om, zeker in een exploratieve fase, outliers te verwijderen. Maar dat

moet dan ook als zodanig vermeld worden.

- Anderzijds vonden de Commissies ook dat extreme scores van een of twee personen in de analyse gehandhaafd werden terwijl ze bij verwijdering significante verschillen niet-significant zouden maken; nergens werd melding gemaakt van het feit dat de significantie op een of enkele personen gebaseerd was.
- Tenslotte werden ook hele groepen respondenten weggelaten, met name indien de uitkomsten de hypothesen in eerste instantie niet bevestigden, opnieuw zonder dat te vermelden. De in de gesprekken genoemde redenen hadden een ad hoc karakter: ‘die student-proefpersonen hadden al eens aan zoiets meegedaan’; ‘die studenten vulden zomaar wat in’, maar diezelfde studenten waren wel in eerste instantie gewoon meegenomen in het experiment en de analyse. Als de weggelaten respondenten andere resultaten te zien hadden gegeven, waren ze wel in de analyses opgenomen.
- Met (vermelding van) betrouwbaarheden van de gebruikte meetschalen werd vaak selectief omgegaan en zeker ook ten eigen voordele, dat wil zeggen ter bevestiging van de onderzoekshypothesen. Met behulp van onbetrouwbare meetinstrumenten kan men geen hypothesen ‘toetsen’ en dus werd bijvoorbeeld de betrouwbaarheid berekend op een deel van de onderzoeksgroep met onvermelde weglating van ‘onbetrouwbare’ personen, of er werd per studie een selectie gemaakt van die items die wel leidden tot een betrouwbaar instrument zonder dat men zich realiseerde dat dit ten koste ging van de onderlinge vergelijkbaarheid tussen studies.
- Indien er discrepanties gevonden werden tussen de door de onderzoeker gerapporteerde betrouwbaarheden (veelal de coëfficiënt alfa) en de door de statistici berekende waarden waren meestal de gerapporteerde waarden duidelijk hoger.
- Indien de betrouwbaarheid van een afhankelijke variabele of een covariaat niet vermeld werd, bleek die veelal te laag ten opzichte van de gangbare normen (bijvoorbeeld lager dan 0,60).
- Soms werd de betrouwbaarheid bewust niet gerapporteerd, met name niet indien die erg laag was. Zo meldde een coauteur dat de begeleiders erop aandrongen de data zo goed mogelijk te verkopen en niet zelf de data onderuit te halen. Dat zou editors en reviewers op een spoor kunnen zetten. Mochten ze er toch naar vragen dan konden de ontbrekende gegevens alsnog geleverd worden.
- Er werd ook selectief omgegaan met de meetschalen, afhankelijk van wat men wilde bewijzen. Aan eenzelfde verzameling items werden in verschillende experimenten, afhankelijk van wat het beste uitkwam in het licht van de onderzoekshypothesen, één dan wel twee dimensies (onderliggende variabelen) ten grondslag gelegd.
- Ook kwam de volgende situatie voor. Een bekend meetinstrument bestaat uit zes items. In het artikel wordt verwezen naar dat instrument maar uit de dataset blijkt dat er slechts vier items meegenomen zijn; twee items zijn er uit weggelaten zonder enige vermelding. In weer een ander experiment met weer hetzelfde meetinstrument gebeurt dit opnieuw maar nu zijn er twee andere items verdwenen, opnieuw zonder vermelding. De enige verklaring voor dit gedrag lijkt te zijn bevestiging te

verkrijgen van de onderzoekshypothesen. In de gesprekken werd gemeld dat de weggelaten items zich niet conform de verwachtingen gedroegen. Daarvoor werden uiteraard ‘goede’ ad hoc redenen gevonden, maar dat alles was niet gemeld in de publicatie en kon alleen geconstateerd worden door de systematische vergelijking van het beschikbare onderzoeksmateriaal en de publicatie.

- Indien bij de heranalyse verschillen in significantieniveau (p-waarden) gevonden werden bij toepassing van dezelfde statistische toetsen, was het doorgaans zo dat de in het artikel gerapporteerde waarden ‘gunstiger’ zijn voor de verwachtingen van de onderzoeker. Zo werden er ook verkeerde afrondingen gevonden, bijvoorbeeld: $p = 0.056$ wordt $p = 0.05$.

Veel van de hierboven vermelde verificatieprocedures kunnen zinvol zijn in een meer exploratieve setting, maar dat moet dan expliciet worden gemeld. De resultaten van een dergelijke exploratie zouden dan in een nieuw onafhankelijk replicatieonderzoek bevestigd moeten worden alvorens tot publicatie over te gaan. Het is verwonderlijk dat dit niet of nauwelijks gebeurde, waar dit type experimenten vaak toch relatief simpel te herhalen is.

In gesprekken met betrokkenen en met de heer Stapel kwam duidelijk naar voren dat veel van de hiervoor genoemde procedures, die overigens door hen niet als onzorgvuldig onderzoek werden gekwalificeerd, vooral gericht waren op een snelle positieve afhandeling van dataverzameling, -analyse en -rapportage, met het oog op het halen van publicatiecriteria. Ook het nauwelijks uitvoeren van replicatieonderzoek kan hieronder worden gerangschikt. De publicatiedruk werd door velen genoemd als belangrijke oorzakelijke verklaring voor de door hen gevolgde werkwijzen. De Commissies doen geen uitspraak over of deze verklaring steek houdt, dan wel alleen als excuus wordt gebezigd. Het verdient aanbeveling hiernaar in groter verband nader onderzoek te verrichten.

5.4 Onvolledige of foutieve informatie over de gevolgde onderzoeksprocedures

Uit de grondregel van openbaarheid en controleerbaarheid volgt dat onderzoeksprocedures dusdanig beschreven moeten worden dat een bepaald experiment nauwkeurig gerepliceerd kan worden. In de vorige paragraaf is al aangegeven dat onderdelen van de onderzoeksprocedures die selectief gebruikt werden om hoe dan ook te bereiken dat de hypothesen bevestigd werden, vaak niet gemeld worden. Daarnaast zijn tal van discrepanties gevonden tussen de wijze waarop het experiment feitelijk verliep, zoals afgeleid kon worden uit de beschikbare data en het onderzoeksmateriaal, en hetgeen in het artikel vermeld werd. Die discrepanties hoefden niet op verificatie gericht te zijn, maar maakten wel de feitelijke gang van zaken ondoorzichtig en daarmee oncontroleerbaar. Ook was de gegeven informatie vaak te vaag om het onderzoek te kunnen repliceren. Uit gesprekken met coauteurs bleek zelfs dat zij soms niet de specifieke aard van het stimulus materiaal kenden voordat ze het artikel gingen schrijven. De inrichting van het experiment werd dan ‘creatief’ beschreven, dat wil zeggen zoals men dacht dat het wel gegaan zou zijn, of in algemene en vage termen. Dit is bevreemdend omdat in de gesprekken met de Commissies door onderzoekers herhaaldelijk naar voren is gebracht dat het in de sociale psychologie vaak om heel precieze en subtiële effecten en manipulaties gaat. De heer Stapel stond juist bekend als een bij uitstek inventief en kundig experimentator, wat zich slecht verhoudt tot een slordige weergave van experimentele opzet en uitvoering.

Enkele concrete voorbeelden, een kleine greep uit de talloze slordigheden en vaagheden, zijn de volgende.

- De onderzoeksgroep werd slechts zeer globaal aangeduid, bijvoorbeeld scholieren in Nederland.
- Er werd verwezen naar een bestaand meetinstrument, maar onvermeld bleef dat hiervan een eigen variant gebruikt was.
- Waar een zeven-puntsschaal gebruikt was, werd een vijf-puntsschaal gemeld.
- Gerapporteerd werd dat ‘aantrekkelijkheid’ gemeten was door middel van de opinie van een andere aanwezige bij het experiment, maar het ging in feite om zelfbeoordeling.
- Niet gemeld werd dat het experiment in een sessie had plaats gevonden waarin nog meer experimentele condities gevarieerd werden en nog meer afhankelijke variabelen werden gemeten. Zulke metingen kunnen elkaar beïnvloeden.
- De vermelde aantallen proefpersonen weken soms naar beneden af van het feitelijk onderzochte aantal; had men alleen een deelgroep onderzocht?
- In de datasets kwamen meer variabelen voor dan vermeld in het artikel, maar ook het omgekeerde was het geval: was die informatie, onvermeld, uit andere experimenten met dezelfde personen gehaald?
- Een veelgehanteerde techniek in de onderzochte publicaties is ook het hanteren van zogenaamde *fillers*. Dit zijn afleidingsmanoeuvres die de proefpersoon de indruk geven van een relevante variabele maar die niet in de analyse worden opgenomen. Dit dient te worden gemeld, maar dat werd doorgaans achterwege gelaten.
- Missing data werden naar aard en omvang niet gemeld, maar simpelweg onvermeld uit de analyse weggelaten.
- ‘Namen’ (John, Mary) die vaak een belangrijke rol spelen in de experimentele condities werden in het artikel verzonnen of verdraaid ten opzichte van de werkelijke gebruikte namen, of onvermeld gelaten.

Deze en veel andere slordigheden in de onderzochte publicaties geven het beeld van een onderzoeksomgeving rond de heer Stapel, waarin het zorgvuldig en kritisch omgaan met onderzoek en data niet hoog in het vaandel stond en geen onderdeel uitmaakte van de praktische onderzoeksopleiding van promovendi.

5.5 Statistische onvolkomenheden

Door de statistici werden tal van onvolkomenheden aangetroffen bij de bestudering en heranalyse van al het onderzoeksmateriaal. Dat is onvermijdelijk; men mag niet verwachten dat sociaal-psychologische onderzoekers op de hoogte zijn van de laatste, specialistische technieken op statistisch gebied. Maar verontrustend is wel dat het hier toch vaak ging om onbekendheid met elementaire onderdelen van de statistiek. Zo werd in een artikel een hele serie uitkomsten aangetroffen van t-toetsen op het verschil tussen twee gemiddelden, die van het soort waren: $t = 0,90$, $p < 0.05$ en dus ‘significant’. De coauteur was oprecht verbaasd toen de Commissie Levelt opmerkte dat een coauteur dit als onzinnig had moeten opmerken. Een andere coauteur bleek geen notie te hebben van wat men nu eigenlijk ‘deed’ in een factoranalyse of wat de betrouwbaarheidscoëfficiënt alfa voorstelt, laat staan wat de berekende onbetrouwbaarheid van bijvoorbeeld de covariaten voor consequenties kan hebben voor de schatting van het experimentele effect. Ook werden in een bepaald geval de ruwe data vervangen door ‘geschatte data’, verkregen uit een of andere vorm van ‘smoothing’ of ‘curve fitting’. Deze werden voorgesteld als de echte ruwe data. De coauteur zag hier niet echt een probleem in en had overigens ook geen idee hoe dit gebeurd was: ‘via een optie in Excel’ (de statistici konden de geschatte data op geen enkele manier reproduceren of koppelen aan de oorspronkelijke data).

Het zijn slechts enkele voorbeelden van statistische incompetentie en desinteresse. Door coauteurs werd dit in de gesprekken veelal niet als een gemis of probleem ervaren.

5.6 Falende wetenschappelijke kritiek

Zoals hierboven beschreven hebben veel factoren aan de late ontdekking van de fraude van de heer Stapel bijgedragen. Toch blijft de klemmende vraag waarom deze fraude en de zozeer verbreide schendingen van een behoorlijke wetenschappelijke werkwijze in de normale controleprocedures binnen de wetenschap nooit zijn ontdekt.

De data en uitkomsten waren in vele opzichten te mooi om waar te zijn. De onderzoekshypothesen werden vrijwel altijd bevestigd. De effecten waren onwaarschijnlijk groot. *Missing* of onmogelijke, *out-of-range* data komen niet of nauwelijks voor. Duidelijk zichtbaar onmogelijke uitkomsten werden niet opgemerkt. Het is bijna onvoorstelbaar dat coauteurs die de data intensief geanalyseerd hebben of reviewers van de internationaal 'leading journals' die geacht worden ter zake deskundig te zijn, niet zagen dat een gerapporteerd experiment praktisch vrijwel onuitvoerbaar was, niet in de gaten hadden dat de statistische rapportage niet deugde, zoals wanneer bijvoorbeeld aan een serie t-waarden echt onmogelijke p-waarden gekoppeld waren, of niet merkten dat in de gepubliceerde tabellen hele reeksen gemiddelden exact gelijk waren, in vele decimalen. Vrijwel niets van al de onmogelijkheden, merkwaardigheden en *sloppyness* die in dit rapport aan de orde zijn gekomen werd opgemerkt door al deze lokale, nationale en internationale vakgenoten en tot een verdenking van fraude kwam het al helemaal niet.

Waar de directe werkomgeving van de heer Stapel een factor was, betrof dit alleen promovendi en postdocs. Er was geen reden voor andere coauteurs, collega-hoogleraren in binnen- en buitenland, leden van promotiecommissies of reviewers van internationale tijdschriften om hun kritische wetenschappelijke functie geen gestand te doen. De Commissies kunnen niet anders concluderen dan dat er van laag tot hoog sprake was van een algemene veronachtzaming van fundamentele wetenschappelijke standaarden en methodologische eisen.

Dit laatste gold zeker ook voor editors en reviewers van internationale tijdschriften. Daar komt nog bij dat veel tijdschriften hun auteurs vóór publicatie formulieren laten invullen waarin de auteurs in diverse varianten correcte onderzoeksprocedures en beschikbaarheid van data en onderzoeksmateriaal garanderen. Auteurs blijken zich daar vaak niet aan te houden (zie onder andere Wicherts et al. 2006). De tijdschriften oefenen hier verder geen enkele controle op uit.

Ook meldden coauteurs in gesprekken met de Commissies meer dan eens dat reviewers aanzetten tot vreemde praktijken. Zo liet een coauteur weten dat editors en reviewers nogal eens verzochten om het weglaten van bepaalde variabelen omdat dat beter in de redenering en in de lijn van het verhaal past, met weglating dus van onwelgevallige resultaten. Ook verzochten reviewers wel om het niet rapporteren van alle uitgevoerde analyses, bijvoorbeeld om het simpel onvermeld laten van condities waarin geen effecten werden gevonden, hoewel die oorspronkelijk wel verwacht waren. Soms drongen reviewers aan op het alsnog achteraf doen van pilot studies, die dan natuurlijk gerapporteerd werden alsof ze vooraf hadden plaatsgevonden. Hiermee werden achteraf de gebruikte experimenten en keuzes van items gelegitimeerd. Reviews waren niet zelden sterk gericht op het vertellen van een interessant, mooi, kort en bondig verhaal, eventueel ten koste van de nodige wetenschappelijke zorgvuldigheid. Het is duidelijk dat de prioriteiten zo verkeerd worden gelegd. Het is toch simpel om alle voor het artikel relevante informatie op een website te

plaatsen en daarnaar expliciet te verwijzen in het artikel.

De voorkeur van tijdschriften voor mooie, bondige, aansprekende bevindingen spoort met de inschatting van diverse gesprekspartners van de Commissies dat de heer Stapel geen belangrijke bijdragen aan de wetenschappelijke theorievorming heeft geleverd. Zijn bijdragen lagen vooral in inventieve experimentele toepassingen en varianten, en in ‘leuke’ uitkomsten. ‘Hij had een goed gevoel voor wat ‘in’ was’, zo werd ons gezegd. Zijn citatiescore is niet hoger, eerder lager dan het gemiddelde van zijn Tilburgse onderzoeksgroep, zoals blijkt uit diverse citatie indices met en zonder de publicaties van de heer Stapel. Dat veel van zijn publicaties in toptijdschriften zijn verschenen doet daar niets aan af. Dat werpt de vraag op in hoeverre toptijdschriften zich voldoende kritisch opstellen tegenover publicaties die niet essentieel bijdragen aan de theorievorming in het vakgebied.

Steeds weer werd door tijdschriften en ervaren onderzoekers op het gebied van de sociale psychologie geaccepteerd dat de hypothesen van de heer Stapel alsmear in een enkel experiment bevestigd werden, met zeer grote effect sizes. Alleen hij was blijkbaar in staat om de precieze manipulaties te realiseren die nodig waren om de subtiele effecten zichtbaar te maken. Men accepteerde, indien men zelf al probeerde te repliceren, dat men faalde omdat men niet de kunde van de heer Stapel had. Maar doorgaans replicateerde men niet en zeker niet onafhankelijk. De enkele keer dat het wel systematisch gebeurde, en faalde, werd nooit bekend, want dat was niet publiceerbaar. Inmiddels is de Commissies een aantal van zulke gevallen in binnen- en buitenland voorgelegd, waarbij veel onderzoeksgeld en kostbare onderzoekstijd verknoeid is.

Als ‘bijvangst’ van hun onderzoekingen hebben de Commissies moeten constateren dat in aanzienlijker mate dan zij aanvankelijk veronderstelden ook aan de discipline zelf een aantal aspecten kleeft die uit het oogpunt van wetenschappelijkheid en wetenschappelijke integriteit als ongelukkig of zelfs onjuist moeten worden aangemerkt.

De kritische functie van de wetenschap heeft, in het geval van de fraude van de heer Stapel, gefaald op alle niveaus. Basisprincipes van wetenschappelijke methodologie werden genegeerd of als irrelevant ter zijde geschoven. Dat heeft naar het oordeel van de Commissies in belangrijke mate bijgedragen aan de zeer uitgestelde ontdekking van die fraude. Het siert de Tilburgse klokkenluiders die schendingen van wetenschappelijke integriteit wél ontdekt te hebben en daar ook naar gehandeld te hebben.

Echter, de bespreking van factoren die de fraude van de heer Stapel hebben gefaciliteerd, zijn machtige positie van coryfee en de falende kritische functie van zijn wetenschappelijke wereld laten onverlet de conclusie van de Commissies, dat de fraude begaan door de heer Stapel dusdanig is dat hij in die factoren en omstandigheden geen enkel excuus voor zijn handelen kan vinden.

6 Aanbevelingen

Ruim een jaar na de presentatie van het interim-rapport, kan gezegd worden dat de casus Stapel heeft gewerkt als een katalysator voor nadere bezinning en verbetering. Het lijkt erop dat nieuwe gevallen van inbreuk op de wetenschappelijke integriteit nu meer dan vroeger in de openbaarheid komen en onderzocht worden. Onderzoekers die met statistische technieken publicaties op fraude onderzoeken, krijgen steeds meer aandacht. Ook zijn er diverse algemene initiatieven genomen ter voorkoming van fraude en hebben onder andere de KNAW, de VSNU, de Federatie van Europese Academies van Wetenschappen (ALLEA) en de Inter Academy Council (IAC) met het instellen van commissies en het uitbrengen van rapporten bijgedragen aan het debat over de schending van wetenschappelijke integriteit en het voorkomen daarvan. Zo biedt het onlangs gepresenteerde rapport van de Commissie Schuyt (KNAW) goede suggesties ten aanzien van de vraag hoe zorgvuldig en integer om te gaan met onderzoeksgegevens. Daarbij is geconcludeerd dat het niet zozeer nodig is meer regels op te stellen, maar dat het belangrijker is om onderzoekers de inhoud en de noodzaak van de bestaande regels in te scherpen. In deze zelfde lijn benadrukken alle vier bovengenoemde onderzoekscommissies de institutionele noodzaak een onderzoeksomgeving te scheppen waarin onderzoeksintegriteit bevorderd wordt. Het blijkt dat, hoezeer ook onderling vertrouwen een grondslag blijft van de wetenschappelijke samenwerking, dit uitgangspunt niet kritiekloos gehanteerd mag worden. De gesignaleerde tekortkomingen moeten niet leiden tot georganiseerd wantrouwen of een overdreven bureaucratie die het wetenschappelijke werk onnodig belemmert, maar er dient een onderzoeksomgeving gecreëerd te worden waarin onderzoekers door onderwijs, training, maar ook effectieve controle gestimuleerd worden de regels van het zorgvuldig en integer uitvoeren van wetenschappelijk onderzoek in acht te nemen.

Ook binnen het veld van de psychologie zijn diverse initiatieven genomen. De ASPO (Associatie van Sociaal-Psychologische Onderzoekers) heeft recentelijk een rapport uitgebracht, *Sharpening scientific policy after Stapel*. (<http://www.sociale-psychologie.nl/>), met aanbevelingen over scholing, data-opslag en replicatieonderzoek binnen de sociale psychologie. Ook zijn in 'de Psycholoog', het blad van het Nederlands Instituut van Psychologen diverse artikelen verschenen met analyses van de Stapel-casus en met suggesties voor de preventie van wetenschappelijk wangedrag. Het tijdschrift *Perspectives on Psychological Science* bracht in november 2012 een special issue uit gewijd aan replicatie binnen het psychologisch onderzoek.

Daarnaast is door een aantal onderzoekers het grootschalige 'Reproducibility Project' gestart (<http://openscienceframework.org/project/EZcUj/wiki/home>), waarbij voor de studies gepubliceerd in 2008 in drie prominente psychologische tijdschriften (*Journal of Personality and Social Psychology*, *Psychological Science*, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*) de mogelijkheid wordt nagegaan om de originele resultaten te repliceren. De commissie wetenschapsintegriteit binnen de Faculteit Psychologie van de Universiteit van Amsterdam heeft een aantal meer precieze en strakke richtlijnen voor verantwoord onderzoek (o.a. een verplicht 'data protocol') voorgesteld. In Tilburg heeft de decaan gehandeld in lijn met de aanbevelingen uit het interim-rapport met betrekking tot het beheer en de opslag van data en de inrichting van onderzoek. Ook in Groningen zijn dergelijke acties ondernomen.

In het interim-rapport zijn diverse aanbevelingen gedaan door de Commissies. De Commissies zijn verheugd dat de aanbeveling om ruimhartig te zijn ten aanzien van de benadeelde promovendi

overgenomen is door de betrokken Rectores Magnifici. Inmiddels hebben zij maatregelen genomen en zijn er gesprekken met betrokkenen gaande.

De aanbeveling om strafrechtelijke aangifte te doen bij het Openbaar Ministerie is door Tilburg University en de Rijksuniversiteit Groningen opgevolgd en er loopt nu een strafrechtelijk vooronderzoek. Er hebben financiële onderzoeken plaatsgevonden in Tilburg en Groningen. Ten aanzien van de correcte besteding van externe en interne subsidiestromen bestaat nog onduidelijkheid, die door nader onderzoek zal dienen te worden weggenomen.

De aanbeveling te onderzoeken of later vertoond uiterst onwaardig wetenschappelijk gedrag grond is voor het ontnemen van de doctorsgraad is naar het oordeel van de Commissie Drenth met betrekking tot de heer Stapel niet meer actueel, omdat deze uit eigen beweging zijn doctorsbul heeft ingeleverd. Zij houdt echter wel voor de toekomst haar algemene relevantie. Over de juridische mogelijkheid van de intrekking van een doctorsgraad bestaan uiteenlopende opvattingen. Alleen al om bestaande onduidelijkheden weg te nemen lijkt aanpassing van de wetgeving aangewezen. Waar het gaat om intrekking van een doctorsgraad in verband met *na* de promotie verrichte gedragingen lijkt de WHW met een specifieke intrekkingbevoegdheid te moeten worden aangevuld. Zie verder bijlage 7 van de Commissie Noort.

Conform de aanbevelingen van de Commissie Levelt heeft Tilburg University een onafhankelijke vertrouwenspersoon integriteit in het leven geroepen. Het College van Bestuur van Tilburg University heeft prof.dr. Cyrille Fijnaut, emeritus hoogleraar strafrecht en criminologie aan Tilburg Law School, bereid gevonden de taak op zich te nemen. Ook is de regeling wetenschappelijke integriteit aangepast. Daarnaast heeft Tilburg University haar beleid met betrekking tot de begeleiding van promovendi veranderd. Zo dient iedere promovendus voortaan door twee of meer (co)promotores begeleid te worden.

De aanbevelingen uit het interim-rapport die nog niet (volledig) geïmplementeerd zijn, volgen hieronder, aangevuld met een aantal extra aanbevelingen.

Verdere afwikkeling van de affaire Stapel

De Rectores Magnifici van de betreffende universiteiten dienen er op toe te zien dat de publicaties ten aanzien waarvan fraude is vastgesteld of een ernstig vermoeden van fraude is geconstateerd, gemeld worden aan de betreffende tijdschriften. Hoewel het de verantwoordelijkheid is van die tijdschriften de artikelen terug te trekken en ontoegankelijk te maken in de databases waar deze artikelen opgevraagd kunnen worden danwel als frauduleus aan te merken bij aanvraag, dienen de universiteiten na te gaan of dit ook daadwerkelijk gebeurd is.

Wat de (deels) frauduleuze dissertaties betreft, moeten de betrokken universiteiten erop toezien dat bij opvragen van of toegang geven tot zo'n dissertatie verwezen wordt naar het (deels) frauduleuze karakter van die dissertatie. Dit kan bijvoorbeeld plaatsvinden door verwijzing naar dit eindrapport.

De betreffende universiteiten dienen de onderzoeksfondsen die aan de financiering van het onderzoek van de heer Stapel hebben bijgedragen te informeren over eventuele frauduleuze publicaties die uit dat onderzoek zijn voortgekomen. Zij dienen ook na te gaan hoe die subsidies op het niveau van Faculteit en Departement zijn gebruikt.

Code of conduct en toezicht

Op facultair niveau dienen condities geschapen te worden voor de bevordering van wetenschappelijk integer handelen en een adequate afhandeling van klachten of verdenkingen van fraude. Volgens de richtlijnen zoals landelijk overeengekomen in het LOWI dient er een vertrouwenspersoon wetenschappelijke integriteit aangesteld te zijn en dient een regeling te bestaan die de bescherming van klokkenluiders waarborgt in geval van een melding van een vermoedelijke inbreuk op wetenschappelijke integriteit. Iedere student en medewerker dient geïnformeerd te worden over het bestaan en de inhoud van de regeling wetenschappelijke integriteit.

De Commissies bevelen aan dat nieuw benoemde staffleden en promovendi bij hun aanstelling geïnformeerd worden over de binnen de instelling geldende 'code of conduct' en dat hun benoeming de acceptatie van deze code impliceert. Aanbevolen 'codes of conduct' zijn 'De Nederlandse Gedragscode Wetenschapsbeoefening' (VSNU, 2004, herziening 2012) en de 'European Code of Conduct for Research Integrity' (ESF-ALLEA 2011). Voor diegenen die statistiek toepassen, is de 'ISI Declaration on Professional Ethics' van het International Statistical Institute ook van belang; zie:

<http://www.isi-web.org/about-isi/professional-ethics>

In het opleidingsprogramma van promovendi dienen de relevante grondprincipes van wetenschapsfilosofie, methodologie, ethiek en statistiek aan de orde te komen die een verantwoorde beoefening van de wetenschap mogelijk maken. Op basis van die inzichten moeten research master studenten en promovendi door hun begeleiders ook praktisch geschoold worden in de toepassing van de regels voor adequaat en integer uitvoeren van wetenschappelijk onderzoek, waarbij onder andere niet wenselijke gedragingen zoals datamassage aan de orde dienen te komen. De Graduate Schools zouden hierop expliciet moeten toezien.

Promotieonderzoek

De data waarop een experimenteel-psychologisch proefschrift gebaseerd is, dienen als regel door de promovendus zelf verzameld, beheerd en geanalyseerd te zijn.

Promotor en copromotor dienen zich te realiseren dat verantwoordelijkheid voor het gehele proefschrift ook de controle en toetsing van de kwaliteit van de oorspronkelijke gegevens omvat. Zij dienen zich ervan te vergewissen dat de data op een integere en verantwoorde wijze zijn verzameld, gecodeerd en bewerkt.

De promotiecommissie dient zich een helder beeld te vormen over de wijze waarop de onderzoeksdata zijn vergaard. Het feit dat een hoofdstuk reeds aanvaard is voor publicatie in een peer-reviewed tijdschrift kan geen voldoende argument zijn om het goed te keuren.

De promotieformules dienen melding te maken van zowel de rechten als de *plichten* die door recht of gewoonte jegens wetenschap en samenleving aan het doctoraat verbonden zijn.

Psychologische onderzoekspraktijk en publicaties

Bij de aanbevelingen willen de Commissies zich niet alleen richten op preventie of terugdringen van fraude, maar ook op het verbeteren van de onderzoekscultuur. De Europese Code spreekt van 'minor misdemeanors': enige datamassage, een paar onwelkome gegevens weglaten, 'gunstig' afronden of

samenvatten, e.d. Deze misdragingen vallen dan wel niet direct onder de ‘grote drie’ (verzinnen, vervalsen, plagiëren), maar zijn in beginsel even onaanvaardbaar, en kunnen, als het niet ontdekt of gecorrigeerd wordt, gemakkelijk leiden tot zwaardere overtredingen van de integriteitsnormen. Zo zijn de Commissies gestuit op vele aspecten van de onderzoekspraktijk die *sloppy science* en verificatiebias verraden. Dit vormt de aanleiding tot de volgende aanbevelingen.

Veel meer dan usance is in de psychologische onderzoekspraktijk moet onderzoeksreplicatie gaan behoren tot het basisinstrumentarium van de discipline en tenminste enkele tijdschriften zouden ruimte moeten bieden voor de publicatie van gerepliceerd onderzoek. Het verrichten van onderzoeksreplicaties dient een onderdeel te vormen van de research master opleiding.

Onderzoeksdata die aan psychologische publicaties ten grondslag liggen dienen gearchiveerd te blijven en op aanvraag ter beschikking gesteld te worden aan andere wetenschapsbeoefenaren. Dit geldt niet alleen voor de dataset die uiteindelijk voor de analyse gebruikt is, maar ook voor de ruwe laboratoriumdata en al het relevante onderzoeksmateriaal, zoals ingevulde vragenlijsten, band- en video-opnamen, etc. Aan te bevelen is een systeem te gebruiken waarbij na afloop van het experiment gebruikte protocollen en data zodanig opgeslagen worden dat deze niet meer veranderd kunnen worden. Het moet duidelijk zijn wie verantwoordelijk is voor opslag en toegankelijkheid van de data. De publicatie dient te vermelden waar de ruwe data zich bevinden en hoe ze blijvend toegankelijk zijn gemaakt. De conclusies dienen altijd controleerbaar herleid te kunnen worden tot de oorspronkelijke data. Tijdschriften zouden pas dan artikelen mogen accepteren indien de betreffende data op deze wijze toegankelijk zijn gemaakt.

Een publicatie dient helder te vermelden welke coauteurs volledig verantwoordelijk zijn voor het gepresenteerde onderzoek en welke slechts voor een bepaald deel ervan. Indien niets vermeld wordt bij de auteursaanduiding dan geldt dat elke auteur volledig verantwoordelijk is voor het gepresenteerde onderzoek.

De Commissies bevelen de Nederlandse en internationale sociale psychologie aan een grondig onderzoek in te stellen naar de bijdrage aan de theorievorming in het vakgebied, de methodologische deugdelijkheid van gepubliceerd sociaal psychologisch onderzoek, alsmede naar de (review)procedures die deze theoretische relevantie en methodologische deugdelijkheid moeten bewaken. Zulk onderzoek kan ook voor aanpalende disciplines van betekenis zijn.

Redacties van tijdschriften wordt aanbevolen indien daartoe aanleiding is gebruik te maken van statistische technieken om aan de hand van de gerapporteerde uitkomsten waarschijnlijke onregelmatigheden te ontdekken. Omdat vermoedelijk in de toekomst de ontdekking van fraude op deze wijze bemoeilijkt zal gaan worden doordat fraudeurs gebruik zullen gaan maken van geavanceerde simulatietechnieken die beschikbaar zijn in standaard statistische pakketten, dient daarnaast het belang van volledige en toegankelijke dataopslag, replicatieonderzoek en verbeterde reviewprocedures onverminderd te worden onderschreven.

Beoordeling van onderzoeksoutput

Visitatiecommissies die de sociale psychologie hebben beoordeeld hebben er onvoldoende blijk van gegeven een aantal van de in dit rapport gesignaleerde feiten te hebben doorzien. Te denken valt aan het

onkritisch volgen van de door tijdschriften verrichte beoordelingen zowel op het punt van de methodologie als de bijdrage aan de theorievorming. Een ander aandachtspunt in dit verband is voorts in hoeverre de visitatiecommissies bijdragen aan het in stand houden van de veronderstelde grote publicatiedruk en daarmee verbonden conventies. Met name betreft dit conventies ten aanzien van het aantal, de volgorde en de verantwoordelijkheid van de coauteurs en het meervoudig publiceren van soortgelijke onderzoeksresultaten.

De Commissies adviseren de VSNU extra aandacht te besteden aan die aspecten van het Standard Evaluation Protocol (SEP) die de inhoudelijke kwaliteit voorop stellen.

Openbaarheid

De Commissies hebben ervaren dat een ‘setting’ van vertrouwelijkheid waarbinnen mensen hun verhaal kunnen doen voor de waarheidsvinding van essentieel belang is. Anders gezegd, openbaarheid kan schadelijk zijn in het kader van waarheidsvinding. Derhalve bevelen de Commissies aan ook in de toekomst bij behandeling van gevallen van wetenschapsfraude de werkzaamheden in beslotenheid uit te voeren.

De resultaten van dergelijk onderzoek moeten per geval openbaar worden gemaakt behoudens bijzondere omstandigheden.

Onder meer om redenen van vertrouwen in de wetenschap is het verschaffen van inzicht in aantallen en aard van de fraudegevallen van groot belang.

Tegen deze achtergrond zijn de Commissies van oordeel dat elke commissie wetenschappelijke integriteit jaarlijks een verslag met informatie over deze aantallen en aard dient te publiceren. Dit verslag moet openbaar worden gemaakt behoudens bijzondere omstandigheden, vergezeld van informatie over de actie die naar aanleiding van het advies is ondernomen. Dit zou per universiteit kunnen plaatsvinden, maar bij voorkeur ook in VSNU-verband, respectievelijk het verband van het Landelijk Orgaan Wetenschappelijke Integriteit (LOWI), waarin tevens de KNAW en NWO participeren.

Bijlage 1 Overzicht Commissieleden en Statistici

Commissie Levelt

Prof.dr. W.J.M. Levelt (voorzitter)

www.mpi.nl/people/levelt-pim

Prof.mr. M.S. Groenhuijsen

www.tilburguniversity.edu/webwijs/show/?uid=m.s.groenhuijsen

Prof.dr. J.A.P. Hagenaars

www.tilburguniversity.edu/webwijs/show/?uid=jacques.a.hagenaars

Ambtelijk secretaris: mw. dr.ir. S.A.M. Baert

Statisticus: dr. M.A.L.M. van Assen

Commissie Noort

Prof.dr. E. Noort (voorzitter)

www.rug.nl/staff/e.noort

Prof.dr. H.E. Bröring

www.rug.nl/staff/h.e.broring

Prof.dr. J.M. Pieters

www.utwente.nl/gw/co/Medewerkers/Vakgroepvoorzitter/pietersnl.doc

Ambtelijk secretaris: mw. dr. M. Jaspers

Statistici: Prof.dr. R.H. Koning

dr. W.J. Post

dr. M.A.J. van Duijn

Commissie Drenth

Prof.dr. P.J.D. Drenth (voorzitter)

www.knaw.nl/Pages/DEF/26/294.bGFuZz1OTA.html

Prof.dr. J.W. Zwemmer

www.uva.nl/over-de-uva/organisatie/medewerkers/content/z/w/j.w.zwemmer/j.w.zwemmer.html

Prof.dr. L.A. de Klerk

www.uva.nl/over-de-uva/organisatie/medewerkers/content/k/l/l.a.deklerk/l.a.de-klerk.html

Prof.dr. C.A.J. Klaassen (secretaris)

www.science.uva.nl/math/People/show_person.php?Person_id=Klaassen+C.A.J.

Statistici: Prof.dr. H.L.J. van der Maas

dr. D. van Ravenzwaaij

Bijlage 2 Vragenlijst Commissie Noort

Commissie Noort - Groningen

Vragenlijst coauteurs artikelen/chapters/ onderzoek Diederik Stapel

1. Artikel/Publicatie :
2. Coauteur: (1e/2e/3e/4e/5e auteur)
3. In welke periode werkte u samen met Diederik Stapel?
4. Wie superviseerde en wie deed de uitvoering van het onderzoek waarover in de publicatie wordt gerapporteerd?
5. Wie was de opdrachtgever en wie of welke instantie was de financier van het onderzoek?
6. Wie bepaalde de onderzoeksvraag?
7. Wie stelde de onderzoeksinstrumenten vast?
8. Heeft u de totale onderzoeksinstrumenten gezien en heeft u gelegenheid gehad daar commentaar op te geven?
9. Was/waren de instrumenten speciaal voor dit onderzoek ontworpen?
Zo ja, door wie?
Zo nee, van wie zijn de instrumenten?
Zo nee, zijn de instrumenten door u of een collega eerder gebruikt?
10. Waar bestonden de ruwe gegevens uit (interviews, vragenlijstantwoorden, reactietijden, ...)?
11. Wie verzamelde de ruwe gegevens?
12. Heeft U de ruwe gegevens gezien ?
13. Zijn de ruwe gegevens nog beschikbaar?
14. Heeft U zelf de ruwe gegevens geanalyseerd?
13. Is er een logboek of history list waarin de statistische analyse staat beschreven?
14. Wie schreef de eerste versie van het artikel/ de publicatie?
15. Welke bijdrage leverde Diederik Stapel aan de totstandkoming van de publicatie?
16. Zijn er in verband met de acceptatie voor publicatie peer review rapporten beschikbaar?
17. Was bij het onderzoek een begeleidingscommissie ingeschakeld?
Zo ja, welke ?
Zo ja, zijn er verslagen van besprekingen van het onderzoek?
18. Heeft U uwerzijds nog opmerkingen m.b.t. het gedane onderzoek?

Commissie Noort - Groningen

Vragenlijst coauteurs proefschrift/ publicaties onderzoek Diederik Stapel

1. Titel proefschrift:
2. In welke periode werkte u samen met Diederik Stapel?
3. Wie superviseerde en wie deed de uitvoering van het onderzoek waarover in het proefschrift wordt gerapporteerd?
4. Wie was de opdrachtgever en wie of welke instantie was de financier van het onderzoek?
5. Wie bepaalde de onderzoeksvragen?
6. Wie stelde de onderzoeksinstrumenten vast?
7. Heeft u de totale onderzoeksinstrumenten gezien en heeft u gelegenheid gehad daar commentaar op te geven?
8. Was/waren de instrumenten speciaal voor dit onderzoek ontworpen?
9. Waar bestonden de ruwe gegevens uit (interviews, vragenlijstantwoorden, reactietijden, ...)?
10. Wie verzamelde de ruwe gegevens?
11. Heeft U de ruwe gegevens gezien?
12. Zijn de ruwe gegevens nog beschikbaar?
13. Heeft U zelf de ruwe gegevens geanalyseerd?
14. Is er een logboek of history list waarin de statistische analyses staat beschreven?
15. Wie schreef de eerste versie van het artikel/ de publicatie/ proefschrift?
16. Welke bijdrage leverde Diederik Stapel aan de totstandkoming van de publicatie/ het proefschrift?
17. Zijn er in verband met de acceptatie van publicaties uit het proefschrift nog peer review rapporten beschikbaar?
18. Was bij het onderzoek een begeleidingscommissie ingeschakeld?
19. Heeft U uwerzijds nog opmerkingen m.b.t. het gedane onderzoek?

Committee Noort - Groningen

Questionnaire for coauthors 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 set up (questionnaires etc.)?
8. Have you actually seen the experimental set up and tools and were you given the opportunity to comment on them?
9. Was the set up especially 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 advise 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?

Committee Noort - Groningen

Questionnaire for authors of dissertations together with Diederik Stapel

1. Title 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 set up (questionnaires etc.)?
7. Have you actually seen the experimental set up and tools and were you given the opportunity to comment on them?
8. Was the set up especially 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 analyze 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 advise 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?

Bijlage 3 Gesprekspartners Commissies

Door de chronologische verdeling van de werkzaamheden was het onvermijdelijk dat sommige gesprekspartners door meerdere commissies geïnterviewd zijn.

Gesprekspartners Commissie Levelt

Dr. Y.R. Avramova
Prof. dr. I. van Beest
Dr. L.M. van den Broek
Prof. dr. A.P. Buunk
Drs. A.J. van Doorn
Prof. dr. P. Eijlander
Drs. M. Elshout
Drs. E.R.K. Evers
Prof. dr. G.L.M. van Heck
Dr. F. van Horen
Dr. Y. Inbar
Dr. C.S. Johnson
Dr. Janneke F. Joly
Prof. dr. G. Keren
Prof. dr. 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. dr. F.G.M. Pieters
Dr. L. Renkema
Dr. K.I. Ruys
Prof. dr. C.G. Rutte
Dr. S.A. Schwinghammer
Prof. dr. G.R. Semin
De heer D.A. Stapel
Drs. F.M. Terpstra
Dr. N. van de Ven
Prof. dr. T.M.M. Verhallen
Prof. dr. A.J.R. van de Vijver

Prof. dr. A.J.J.M. Vingerhoets
Dr. C.J. Wiekens
Drs. J. van Wolferen
Prof. dr. M. Zeelenberg

Gesprekspartners Commissie Noort

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

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

Gesprekspartners Commissie Drenth

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

Bijlage 4 Onderzochte publicaties Commissie Levelt inclusief bevindingen

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 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.
- * 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 of 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 of 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 neighbor. 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 analyzed 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 on 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 unbelievably 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).

Bijlage 5 Onderzochte publicaties

Commissie Noort inclusief bevindingen

List of publications examined by the Noort Committee

* *The Noort Committee established 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 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 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 of 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 of 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 of 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 of 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 of 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 of 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 of 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 favor 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 of 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 favor 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 data-set 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 of 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 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 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 of 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 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 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.
- 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 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 favor 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 favor 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 favor 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).

Bijlage 6 Onderzochte publicaties

Commissie Drenth inclusief bevindingen

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 krantartikelen 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 favor 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 favor 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 favor 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 favor 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 favor 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 favor of the hypothesis that the underlying data are correlated, versus the hypothesis that they are independent, as they should be.

Bijlage 7 Bevoegdheid tot intrekking van de doctorsgraad

Naar aanleiding van haar taakopdracht met betrekking tot juridische aspecten heeft de Commissie Noord mede stilgestaan bij de juridische mogelijkheid tot intrekking van een verleend doctoraat. Besluiten tot verlening van een academische graad krijgen al snel formele rechtskracht. Dit wil zeggen dat er tegen deze besluiten geen rechtsgang meer openstaat. Wel is aannemelijk dat het bevoegde bestuursorgaan onder voorwaarden de bevoegdheid heeft een toegekende academische graad terug te nemen wanneer de toekenning ervan berust op fraude door degene aan wie het doctoraat is verleend zelf. Dit betreft een aantasting van de materiële rechtskracht. Die fraude moet dan wel overtuigend door het bestuursorgaan worden aangetoond.

De WHW kent geen expliciete bevoegdheid tot intrekking van verleende graden. Verdedigbaar is dat de verleningsbevoegdheid een intrekkingbevoegdheid impliceert. Ervan uitgaande dat een geïmpliceerde bevoegdheid aanwezig is, kan worden betoogd dat de gronden voor intrekking dezelfde zijn als die waarop een graad wordt toegekend. Toekenning van een dr.-graad berust hierop, dat de promovendus blijkt heeft gegeven van zijn bekwaamheid tot het zelfstandig verrichten van wetenschappelijk onderzoek (te beoordelen aan de hand van de vraagstelling, methode, beantwoording van de vraagstelling, et cetera). Het is mede tegen deze achtergrond dat in het interim-rapport, p. 7, wordt geconcludeerd dat de door de heer Stapel gepleegde fraude in geen van de gevallen gevolgen zou moeten hebben voor de aan diens promotie verleende dr.-graden.

In geval van eigen fraude heeft een promovendus geen blijk gegeven van zijn bekwaamheid tot het zelfstandig verrichten van wetenschappelijk onderzoek. In een dergelijk geval, zo kan worden geredeneerd, kan de (geïmpliceerde) bevoegdheid zich aandienen om de promotus de hem verleende dr.-graad te ontnemen. Daarbij is van betekenis dat een beoordelingscommissie was ingeschakeld: dit zal van invloed zijn op de bewijslast. Verder is het opleidingskarakter van het PhD-traject relevant. Ten slotte is er het aspect van rechtszekerheid en tijdsverloop. Deze punten vormen extra drempels voor eventuele intrekking van een dr.-graad.

Het voorgaande betreft de mogelijkheid van intrekking van een verleend doctoraat in verband met fraude met betrekking tot de dissertatie zelf. Een andere vraag is of *na* de promotie verrichte gedragingen, die derhalve los staan van de grond voor toekenning van de dr.-graad voor de dissertatie en de verdediging daarvan, de bevoegdheid doen ontstaan om een verleende graad in te trekken en, zo ja, om wat voor soort gedragingen het dan moet gaan: alleen (vanzelfsprekend ernstige) inbreuken op wetenschappelijke integriteit, of ook anderszins (ernstige vormen van) maatschappelijk onwaardig gedrag? Het meest plausibele standpunt is dat de huidige WHW, anders dan het Duitse (Hochschul)recht, geen bevoegdheid, ook geen geïmpliceerde bevoegdheid, kent tot intrekking van een dr.-graad op een andere grond dan die waarop het doctoraat is verleend.

In het licht van het bovenstaande meent de Commissie dat de mogelijkheden tot intrekking van een

verleende dr.-graad nader aandacht verdienen, primair in academisch verband en daarnaast op departementaal niveau. Over de juridische mogelijkheid van de intrekking van een dr.-graad bestaan uiteenlopende opvattingen. Alleen al om bestaande onduidelijkheden weg te nemen, dus in het belang van de rechtszekerheid, lijkt aanpassing van de wetgeving aangewezen. Waar het gaat om intrekking van een dr.-graad in verband met *na* de promotie verrichte gedragingen lijkt de WHW met een specifieke intrekkingbevoegdheid te moeten worden aangevuld (waarbij de voorkeur van de Commissie zou uitgaan naar een beperking tot (vanzelfsprekend ernstige) inbreuken op wetenschappelijke integriteit). Bij dit alles acht de Commissie het voorts wenselijk dat meer nadrukkelijk en gepreciseerd dan tot dusverre gebruikelijk is aandacht wordt besteed aan de aan het doctoraat verbonden rechten én plichten.

Bijlage 8 Literatuurverwijzingen

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

