

DE 7 DOODZONDEN VAN DE PSYCHOLOGIE

DE  **DOODZONDEN**
VAN DE CHRIS
CHAMBERS
PSYCHOLOGIE

Pleidooi voor een cultuuromslag in
de wetenschappelijke praktijk

SCRIPTUM

Voor J, I en X,

*voor elke junior wetenschapper wiens resultaten niet goed genoeg waren,
en voor JD, die zo veel mensen heeft geïnspireerd maar nooit zal zien wat
daaruit voort zou kunnen komen.*

*Men moet in staat zijn om te zien dat dingen hopeloos zijn, en toch
vastbesloten om er verandering in te brengen. – F. Scott Fitzgerald*

Oorspronkelijke titel: *The Seven Deadly Sins of Psychology. A Manifesto for
Reforming the Culture of Scientific Practice*

Copyright © 2017 by Chris Chambers
Gepubliceerd door Princeton University Press
Binnenwerk illustraties Anastasiya Tarasenko

Nederlandse vertaling © 2018 Scriptum, Schiedam
Vertaling Vanja Walsmit
Grafische vormgeving cover en binnenwerk www.igraph.be

Niets van deze uitgave mag worden veelevoudigd in enige vorm of op
enige wijze zonder schriftelijke toestemming, met uitzondering van korte
citaten als onderdeel van kritieken en boekbesprekingen.

ISBN 978 94 6319 101 2
NUR 770 Psychologie Algemeen

info@scriptum.nl
www.scriptum.nl
[Twitter.com/ScriptumNL](https://twitter.com/ScriptumNL)
[Facebook.com/UitgeverijScriptum](https://facebook.com/UitgeverijScriptum)

INHOUD

Voorwoord 7

Doodzonde 1 <i>Bias</i>	13
Een korte geschiedenis van de ‘jaknikker’	17
Neofilie: als positief en nieuw het wint van negatief maar waar	21
Replicatie van concepten, niet van experimenten	26
Geschiedvervalsing	30
De strijd tegen bias	33
Doodzonde 2 <i>Verborgene flexibiliteit</i>	37
<i>p</i> -hacking	39
Bijzondere patronen van <i>p</i>	43
Op spokenjacht	50
Onbewuste ‘afstemming’ van analyses	52
Selectieve debugging	55
Zijn onderzoekspsychologen niet meer dan slecht betaalde advocaten?	56
Oplossingen voor verborgen flexibiliteit	57
Doodzonde 3 <i>Onbetrouwbaarheid</i>	63
Bronnen van onbetrouwbaarheid in de psychologie	65
Reden 1: Minachting voor directe replicatie	65
Reden 2: Gebrek aan power	73
Reden 3: Methodes worden niet openbaar gemaakt	79
Reden 4: Statistische dwalingen	81
Reden 5: Artikelen worden niet teruggetrokken	84
Oplossingen voor onbetrouwbaarheid	86

Doodzonde 4 <i>Data hoarding</i>	95
De onuitgesproken voordelen van het delen van data	97
Als data niet worden gedeeld	98
Restricties op delen	100
Niet delen verhult wangedrag	102
Zo verhef je het delen van data tot norm	104
Hervorming bij de basis, wortels en stokken	109
Ontsluiting van de zwarte doos	112
Zo voorkom je slechte gewoontes	116
Doodzonde 5 <i>Corrumpeerbaarheid</i>	119
De anatomie van fraude	122
Het vage grijze gebied	129
Als junior wetenschappers van het rechte pad afdwalen	136
Het verhaal van Kate	141
Twaalf slinkse tips: zo kom je weg met fraude	146
Doodzonde 6 <i>Geslotenheid</i>	151
De grondbeginselen van open-accesspublicatie	153
Waarom steunen psychologen op restricties gebaseerde publicatie?	154
Hybride OA als oplossing én als probleem	157
Hulp van de guerrilla's	162
Tegenargumenten	164
Een open route	174
Doodzonde 7 <i>Bonen tellen</i>	175
Wegen die nergens heen leiden	177
Impactfactoren en hedendaagse astrologie	177
De omgekeerde wereld	187
De ondoorgrondelijke brij van academisch auteurschap	190
Een weg die ergens heen leidt	196
Hoofdstuk 8 <i>Verlossing</i>	199
Een oplossing voor de zonden van bias en verborgen flexibiliteit	202
Registered Reports: een vaccin tegen bias	203
Preregistratie zonder collegiale toetsing	226
Oplossingen voor de zonde van onbetrouwbaarheid	228
Oplossingen voor de zonde van data hoarding	232

Oplossingen voor de zonde van corrumpeerbaarheid	235
Oplossingen voor de zonde van geslotenheid	238
Oplossingen voor de zonde van het bonen tellen	241
Concrete stappen naar hervorming	243
Sluitstuk	246

Noten	249
Register	289

VOORWOORD

Dit boek is voortgekomen uit wat ik alleen maar kan omschrijven als een diepe persoonlijke frustratie over de werkcultuur binnen de psychologische wetenschap. Ik vergelijk onze beroepscultuur altijd met een kasteel – een heiligdom van inspanningen dat lang geleden door onze voorouders is gebouwd. Net als elk ander huis heeft het constante zorg en aandacht nodig, maar in plaats van het door de jaren heen bij te houden hebben we het laten verpauperen. De ramen zijn vuil en ondoorzichtig. Het dak lekt; nog even en de regen komt erdoorheen. In de kerker wonen monsters.

Ondanks de vele tekortkomingen ervan heb ik veel aan het kasteel gehad. Het heeft me onderdak geboden tijdens mijn vormende jaren als junior onderzoeker en me zo ver gebracht dat ik nu openlijk kan praten over de noodzaak van renovatie. Ik zeg met nadruk renovatie, omdat ik niet wil suggereren dat we ons bolwerk moeten afbreken en opnieuw moeten beginnen. De fundamenten van de psychologie zijn stevig, en het vakgebied kan zich beroemen op vele ontdekkingen door de jaren heen. Onze grondleggers – Helmholtz, Wundt, James – hebben het op een duurzame manier gebouwd.

Maar na vijftien jaar te hebben doorgebracht in de psychologie en zijn neefje, de cognitieve neurowetenschap, ben ik tot een verontrustende conclusie gekomen. Als we op deze voet doorgaan, zal de reputatie van de psychologie als wetenschap steeds verder verslechteren en is de kans reëel dat ze verdwijnt. Als we de waarschuwingstekens nu negeren, wordt de psychologie over honderd jaar of eerder als de zoveelste curieuze wetenschappelijke uitspatting gezien; dan kijken we ernaar zoals we tegenwoordig naar de alchemie of de frenologie kijken. Onze nakomelingen zullen toegeeflijk glimlachen om dit staaltje van academische antiquiteit, elkaar wijs toeknikken over de protowetenschap die de psychologie was en concluderen dat we gebonden waren aan de 'beperkingen van de tijd'. Er zijn natuurlijk maar weinig wetenschappen die het oordeel van de geschiedenis glansrijk

doorstaan. Maar de psychologie zal vooral hard worden beoordeeld op haar onderzoeksmethodes en niet zozeer op haar ontdekkingen. En dat oordeel zal als volgt luiden: net als zo veel andere ‘softe’ wetenschappen zaten we gevangen in een cultuur waarin we de schijn van wetenschap een geschikt substituuut vonden voor het uitoefenen van echte wetenschap.

In dit boek laat ik zien hoe deze vervorming zijn weerslag heeft op veel aspecten van ons werk als wetenschappers. Het beeld dat ik schets is zo nu en dan weinig rooskleurig. Aan de hand van de metafoor van de zeven doodzonden leg ik uit hoe ongecontroleerde vormen van bias ons om de tuin leiden en ons laten zien wat we willen zien; hoe we fundamentele principes van de wetenschappelijke methode de rug hebben toegekeerd; hoe we de data die we verzamelen als persoonlijk eigendom in plaats van als openbare bronnen behandelen; hoe we toestaan dat academische fraude onnoemelijk veel schade toebrengt aan de meest kwetsbare leden van onze gemeenschap; hoe we publieke middelen verspillen aan verouderde publicatievormen en hoe we, bij de beoordeling van wetenschap en wetenschappers, ons deskundige oordeel hebben verruild voor een oppervlakkige obsessie met cijfers. Ik hoop u ervan te kunnen overtuigen dat we ons tijdens onze zoektocht naar echt inzicht voortdurend bewust moeten blijven van deze tekortkomingen en ze voortdurend moeten blijven rechtzetten.

In elk hoofdstuk, en in een afzonderlijk laatste hoofdstuk, draag ik verschillende suggesties voor hervormingen aan. Daarbij staan twee kernaspecten van de wetenschap centraal: transparantie en reproduceerbaarheid. Om in de eenentwintigste eeuw en daarna te kunnen overleven moeten we onze ondoorzichtige en kwetsbare cultuur omvormen tot een open, betrouwbare wetenschap – een wetenschap die openheid net zo hoog in het vaandel heeft staan als innovatie, die net zo veel waarde hecht aan robuustheid als aan vernieuwing. We moeten erkennen dat onze oude manier van werken zijn doel niet langer dient en een nieuwe weg inslaan.

In de breedste zin des woords is dit boek bedoeld voor iedereen die geïnteresseerd is in de beoefening en de cultuur van wetenschap. Zelfs mensen zonder specifieke belangstelling voor de psychologie hebben reden tot zorg over de problemen waarmee ons vakgebied wordt geconfronteerd. Binnen elke wetenschap doen zich wanpraktijken voor waarmee kostbaar publiek geld wordt verspild doordat er onderzoeklijnen worden gevolgd die misleidend of onecht blijken. Door bepaalde soorten resultaten uit de publiciteit te houden bestaat bijvoorbeeld het risico dat er ineffectieve klinische behandelingen worden geïntroduceerd voor mentale gezondheidsproblemen als depressie en schizofrenie. In het Verenigd Koninkrijk, waar de so-

ciaal-economische invloed van onderzoek regelmatig wordt gemeten als onderdeel van het Research Excellence Framework (REF), blijkt dat de psychologie invloed heeft op een breed scala aan praktijktoepassingen. Het REF van 2014 maakte bijvoorbeeld melding van meer dan 450 casestudies met maatschappelijke relevantie, waar psychologisch onderzoek bepalend is geweest voor overheidsbeleid of -werkwijzen, zoals (om maar een paar dingen te noemen) het ontwerp en het gebruik van elektrische auto's, strategieën om examenvrees te verminderen, de ontwikkeling van betere politieverhoortechnieken die rekening houden met de grenzen van het menselijke geheugen, het vaststellen van snelheidslimieten op basis van ontdekkingen in de oogheelkunde, menselijke factoren die van belang zijn bij effectieve verkenning van de ruimte, overheidsstrategieën voor het omgaan met klimaatverandering waarbij de publieke perceptie van risico's wordt ingecalculleerd, en neutrale verpakking van tabaksproducten.¹ Zowel de meest basale fundamenteën als de meest praktische toepassingsmogelijkheden van de psychologie vormen een rijk onderdeel van het openbare leven en zijn van cruciaal belang bij ons begrip van veel mondiale problemen. Daarom zijn de in dit boek besproken doodzonden een probleem voor de samenleving als geheel.

Een deel van dit boek, met name de gedeeltes over statistische methodes, zal vooral relevant zijn voor de kersverse onderzoeker – de universitaire student of de beginnende wetenschapper. Maar het bevat ook belangrijke boodschappen voor ervarener academici die hun eigen laboratorium of instituut leiden, en veel zaken zijn ook relevant voor journalisten en wetenschapsauteurs. Om de toegankelijkheid van het bronmateriaal voor verschillende doelgroepen te vergemakkelijken verwijs ik zoveel mogelijk naar literatuur die vrij toegankelijk is. Voor artikelen die dat niet zijn, levert een zoekopdracht naar de titel van het artikel in Google Scholar vaak een gratis beschikbaar elektronisch exemplaar op. Daarnaast heb ik geput uit meer hedendaagse vormen van communicatie, zoals vrij toegankelijke blogposts en social media.

Ik ben veel dank verschuldigd aan vele vrienden, academische collega's, tijdschriftredacteurs, wetenschapsauteurs, journalisten, persvoorlichters en beleidsdeskundigen voor jaren van inspiratie, cruciale gesprekken, discussies en in sommige gevallen interviews waarvan gedeeltes in dit boek zijn verwerkt: Rachel Adams, Chris Allen, Micah Allen, Adam Aron, Vaughan Bell, Sven Bestmann, Ananyo Bhattacharya, Dorothy Bishop, Fred Boy, Todd Braver, Björn Brembs, Jon Brock, Jon Butterworth, Kate Button, Iain Chalmers, David Colquhoun, Molly Crockett, Stephen Curry, Helen

Czerski, Zoltan Dienes, wijlen Jon Driver, Malte Elson, Alex Etz, John Evans, Eva Feredoes, Matt Field, Agneta Fischer, Birte Forstmann, Fiona Fox, Andrew Gelman, Tom Hardwicke, Chris Hartgerink, Tom Hartley, Mark Haselgrove, Steven Hill, Alex Holcombe, Aidan Horner, Macartan Humphreys, Hans IJzerman, Helen Jamieson, Alok Jha, Gabi Jiga-Boy, Ben Johnson, Rogier Kievit, James Kilner, Daniël Lakens, Natalia Lawrence, Keith Laws, Katie Mack, Leah Maizey, Jason Mattingley, Rob McIntosh, Susan Michie, Candice Morey, Richard Morey, Simon Moss, Ross Mounce, Nils Mulhert, Kevin Murphy, Suresh Muthukumaraswamy, Bas Neggers, Neuroskeptic, Kia Nobre, Dave Nussbaum, Hans Op de Beeck, Ivan Oransky, Damian Pattinson, Andrew Przybylski, James Randerson, Geraint Rees, Ged Ridgway, Robert Rosenthal, Pia Rotshtein, Jeff Rouder, Elena Rusconi, Adam Rutherford, Chris Said, Ayse Saygin, Anne Scheel, Sam Schwarzkopf, Sophie Scott, Dan Simons, Jon Simons, Uri Simonsohn, Sanjay Srivastava, Mark Stokes, Petroc Sumner, Mike Taylor, Jon Tennant, Erick Turner, Carien van Reekum, Simine Vazire, Essi Viding, Solveiga Vivian-Griffiths, Matt Wall, Tony Weidberg, Robert West, Jelte Wicherts, Ed Wilding, Andrew Wilson, Tal Yarkoni, Ed Yong en Rolf Zwaan. Verder wil ik Sergio Della Sala en Toby Charkin bedanken voor hun medewerking en hun standvastige inzet voor Registered Reports bij *Cortex*, Brian Nosek, David Mellor en Sara Bowman voor het feit dat zij Registered Reports met open armen hebben ontvangen bij het Center for Open Science, en de Royal Society, met name uitgever Phil Hurst en uitgeefdirecteur Stuart Taylor, die Registered Reports hebben omarmd lang voordat andere multidisciplinaire tijdschriften dat deden. Ook ben ik dank verschuldigd aan Marcus Munafò, die me bij elke stap van de promotie van Registered Reports heeft vergezeld, en aan de 83 wetenschappers die onze open brief in de *Guardian* hebben ondertekend waarin we oproepen tot overname van het format door alle tijdschriften voor de niet-exacte wetenschappen. Tot slot een speciaal bedankje aan het adres van Dorothy Bishop, wijlen (en zeer gemiste) Alex Danchev, Dee Danchev, Zoltan Dienes, Pete Etchells, Hal Pashler, Frederick Verbruggen en E.J. Wagenmakers voor hun uitgebreide lezingen en besprekingen van proefversies, dank aan Anastasiya Tarasenko voor het maken van de illustraties aan het begin van elk hoofdstuk en aan mijn redacteurs Sarah Caro en Eric Schwartz voor hun geduld en hun wijze adviezen gedurende het hele traject.

DOODZONDE 1

Bias

*Als het menselijk verstand eenmaal een
mening heeft omarmd... sleept het alles erbij
wat die mening schraagt en onderschrijft.*

– Francis Bacon, 1620



Het jaar 2011 zal misschien wel de geschiedenis ingaan als het jaar waarin de psychologie voor altijd veranderde. Het begon allemaal toen het *Journal of Personality and Social Psychology* een artikel publiceerde met de titel 'Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect.'¹ In dit artikel deed Daryl Bem van Cornell University verslag van een reeks experimenten met psi of 'precognitie', een bovennatuurlijk verschijnsel dat mensen in staat zou stellen om gebeurtenissen in de toekomst te zien. Bem, zelf een gerenommeerd psycholoog, koos voor een innovatieve methode om psi te bestuderen. In plaats van gebruik te maken van in diskrediet geraakte parapsychologische methodes als opdrachten met kaarten of dobbelsteentests koos hij voor een aantal beproefde psychologische technieken, die hij op slimme manieren aanpaste.

Een van die methodes was een omgekeerde priming-taak. Normaal gesproken moeten mensen bij een priming-taak aangeven of een afbeelding die op een computerbeeldscherm wordt getoond gekoppeld is aan een positieve of een negatieve emotie. De proefpersoon geeft bijvoorbeeld aan of een afbeelding van jonge katjes prettige of onprettige associaties bij hem oproept. Als er direct voorafgaand aan de afbeelding een woord wordt gepresenteerd dat dezelfde emoties oproept – de prime – (zoals het woord 'plezier' gevolgd door de afbeelding van de jonge katjes) vinden mensen het gemakkelijker om de emotie die de afbeelding oproept te benoemen en reageren ze sneller. Maar roepen de prime en de afbeelding tegengestelde emoties op, dan wordt de taak moeilijker (zoals het woord 'moord' gevolgd door een afbeelding van jonge katjes). Om het bestaan van precognitie te onderzoeken draaide Bem de volgorde van dit experiment om. Hij ontdekte dat ook primes die werden aangeboden nadat proefpersonen op een afbeelding hadden gereageerd, hun reactietijd leken te beïnvloeden. Op het geheugen was volgens hem sprake van vergelijkbare 'retroactieve' effecten. In een van zijn experimenten waren mensen over het geheel genomen beter in het reproduceren van specifieke woorden uit een lijst die ook in een praktische taak terugkwamen, maar het bijzondere was dat deze taak werd uitgevoerd na de geheugentaak en niet ervoor. Op basis hiervan concludeerde Bem dat de proefpersonen in het verleden profijt konden hebben van een taak die zij in de toekomst hadden uitgevoerd.

De conclusies van Bem brachten natuurlijk een stortvloed van verwarring en controverses teweeg. Hoe kon een gebeurtenis in de toekomst in hemelsnaam iemands reactietijd of geheugen in het verleden beïnvloeden? Als precognitie werkelijk bestond, ook al was het maar bij een heel klein

deel van de bevolking, hoe kon het dan dat casino's en aandelenmarkten nog winst maakten? En hoe was het mogelijk dat zo'n bizarre conclusie in een gerenommeerd wetenschappelijk tijdschrift was opgenomen?

Allereerst werd er kritisch gekeken naar de procedures die Bem voor zijn experimenten had gehanteerd. Wellicht kon een fout in zijn werkwijze de resultaten verklaren; misschien had hij de volgorde van de gebeurtenissen niet gerandomiseerd of had hij een andere subtiele fout in zijn experiment gemaakt. Maar deze aspecten van het experiment leken bij nader onderzoek in orde. Dat plaatste de onderzoeksgemeenschap voor een dilemma. Als Bems bevindingen klopten, was precognitie de meest sensationele ontdekking in de moderne wetenschap. We zouden het bestaan van tijdreizen moeten aanvaarden en onze opvattingen over oorzaak en gevolg zouden compleet op de schop moeten. Maar klopten de resultaten van Bem niet, dan betekende het dat er iets fundamenteel mis was met de standaard onderzoeksmethodes. Als geaccepteerde methodes zulke onzinnige resultaten konden opleveren, waren gepubliceerde bevindingen in de psychologie dan überhaupt nog wel te vertrouwen? Psychologen stonden voor een niet-benijdenswaardige keuze: aanvaarding van een onmogelijke wetenschappelijke conclusie, of erkenning van een onverteerbare realiteit in hun vakgebied.

De wetenschappelijke gemeenschap was instinctief sceptisch over de conclusies van Bem. In reactie op een voorpublicatie van het artikel eind 2010 zei psycholoog Joachim Krueger: 'Mijn persoonlijke mening is dat dit belachelijk is en niet waar kan zijn.'² Buitengewone beweringen vragen immers om buitengewoon solide bewijs, en ondanks dat Bems bewijs gepubliceerd was in een prestigieus tijdschrift, werd de statistische significantie ervan als verre van solide beschouwd.

Bem realiseerde zich zelf ook dat zijn resultaten om een verklaring vroegen. Hij benadrukte dat het noodzakelijk was dat onafhankelijke onderzoekers zijn bevindingen zouden repliceren. Maar dat bleek lastiger dan gedacht. Een poging tot replicatie door Chris French en Stuart Ritchie leverde geen enkel bewijs op voor precognitie, maar werd afgewezen door het tijdschrift waarin Bems oorspronkelijke artikel was gepubliceerd. Het tijdschrift nam dit keer niet eens de moeite om collegiale toetsing te regelen voor het artikel van French en Ritchie alvorens het af te wijzen. De verklaring daarvoor luidde dat er 'geen replicatieonderzoeken worden gepubliceerd, ongeacht of ze succesvol zijn.'³ Deze beslissing klinkt misschien bizar, maar zoals we ook later zullen zien komt minachtig voor replicatie binnen de psychologie veel vaker voor dan bij de meer gevestigde wetenschappen.

De meest gerenommeerde psychologietijdschriften publiceren selectief bevindingen die zij origineel, nieuw, de moeite waard en bovenal positief vinden. Deze *publicatiebias*, ook wel het ‘dossierkast- of bureaulade-effect’ genoemd, heeft tot gevolg dat onderzoeken die geen statistisch significante resultaten opleveren, of die het werk van anderen reproduceren, zo weinig prioriteit hebben dat ze effectief uit de wetenschappelijke literatuur worden geweerd. Ze eindigen in een dossierkast, of ze worden überhaupt niet uitgevoerd.

Publicatiebias is een vorm van wat ongetwijfeld de grootste tekortkoming is in het redeneringsvermogen van de mens: onze voorkeur voor bevestiging, oftewel confirmatiebias. Als we ten prooi vallen aan confirmatiebias, gaan we op zoek naar en geven we de voorkeur aan bewijs dat strookt met onze bestaande overtuigingen, terwijl we bewijs waarvoor dat niet geldt negeren of ontcrachten. Confirmatiebias corrumpeert de psychologische wetenschap op verschillende manieren. In zijn meest eenvoudige vorm leidt het tot voorkeur voor publicatie van positieve resultaten – dat wil zeggen van hypothesetoetsingen die statistisch significante verschillen of overeenkomsten tussen condities opleveren (zoals A is groter dan B; A is gerelateerd aan B, versus A is gelijk aan B; A is ongerelateerd aan B). Op een subtielere manier leidt het tot de schijn van wetenschappelijke reproduceerbaarheid waarbij het mogelijk is om eerdere bevindingen te repliceren maar niet om ze te falsifiëren, en moedigt het mensen aan om hypothesen van experimenten achteraf zodanig aan te passen dat ze ‘onverwachte’ uitkomsten voorspellen. Een van de meest verontrustende aspecten van de psychologie is dat de academische gemeenschap weigert zulk gedrag unaniem te veroordelen. Integendeel: veel psychologen leggen zich bij deze praktijken neer en zien ze zelfs als overlevingsstrategieën binnen een cultuur waarin het voor onderzoekers ‘publiceren of creperen’ is.

Een paar maanden nadat Bems beweringen over precognitie in een gerenommeerd academisch tijdschrift waren verschenen kregen ze een krachtig, zij het onbedoeld effect op de psychologische gemeenschap. Gevestigde methodes en geaccepteerde publicatiepraktijken werden opnieuw onder de loep genomen omdat zij resultaten kunnen opleveren die overtuigend lijken maar bijna zeker onjuist zijn. In een artikel waarin statistisch gezien niets heel werd gelaten van Bems paper merkten psycholoog Eric-Jan Wagenmaker en collega’s op: ‘Uit onze evaluatie komt naar voren dat er iets fundamenteel mis is met de manier waarop experimenteel psychologen hun onderzoeken opzetten en verslag doen van hun statistische resultaten.’⁴ Met die woorden barstte de storm los.

Een korte geschiedenis van de ‘jaknikker’

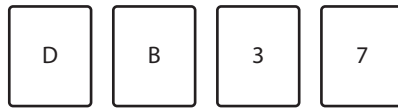
Om te begrijpen op welke manieren de hiervoor beschreven vooringenomenheid de psychologische wetenschap beïnvloedt, moeten we terug in de tijd gaan naar de historische oorsprong ervan en de eerste onderzoeken naar confirmatiebias. Filosofen en geleerden zijn al tijden bekend met de ‘jaknikker’ in relatie tot menselijk redeneren. Al in de vijfde eeuw voor Christus zinspeelde Thucydides op het bestaan ervan: ‘Wanneer een conclusie een man bevalt, accepteert hij haar zonder tegenwerpingen, maar bevalt zij hem niet, dan brengt hij er alle kracht van de logica en de rede tegen in stelling.’ Vergelijkbare bewoordingen kwamen uit de mond van Dante, Bacon en Tolstoj. Halverwege de twintigste eeuw, toen psychologen manieren bedachten om confirmatiebias te meten in gecontroleerde laboratoriumexperimenten, ontwikkelde de kwestie zich van een filosofische tot een wetenschappelijke.

Sinds halverwege de jaren vijftig zijn er verschillende onderzoeken uitgevoerd die uitwijzen dat wanneer mensen geconfronteerd worden met een reeks observaties (data) en een mogelijke verklaring (hypothese), zij de voorkeur geven aan toetsingen van die hypothese die haar proberen te bevestigen in plaats van te weerleggen. Formeel gezien betekent dit dat mensen eerder geneigd zijn om naar data te kijken waarmee zij kunnen inschatten dat een specifieke hypothese *juist* is, $p(\text{data}|\text{hypothese})$ dan dat zij de tegenovergestelde kans proberen in te schatten dat een hypothese onjuist is, $p(\text{data}|\sim\text{hypothese})$. Met andere woorden: mensen stellen liever vragen waarop het antwoord ‘ja’ is. Daarmee gaan ze voorbij aan de stelregel van filosoof Georg Henrik von Wright: ‘Geen enkel bevestigend voorbeeld van een wetmatigheid is een verificatie, maar... elk ontkennend voorbeeld is een weerlegging.’⁵

De psycholoog Peter Wason was een van de eerste onderzoekers die in het laboratorium bewijs vond van de confirmatiebias. In een van de innovatieve experimenten die hij in jaren zestig en zeventig van de twintigste eeuw uitvoerde, gaf hij deelnemers een reeks getallen, zoals 2-4-6, en vroeg hun om de regel te achterhalen die eraan ten grondslag lag (in dit geval: drie cijfers in oplopende orde van grootte).⁶ Als ze zich een hypothese hadden gevormd, mochten de proefpersonen hun eigen reeks opschrijven, waarna ze te horen kregen of deze consistent of inconsistent was met de regel. Wason ontdekte dat de proefpersonen een sterke neiging vertoonden om verschillende hypothesen te toetsen door ze te bevestigen, zelfs als ze daarmee geen plausibele alternatieven elimineerden (zoals drie even

getallen). De proefpersonen van Wason hanteerden deze strategie ondanks het feit dat hun van tevoren was verteld dat het niet het doel was om simpelweg getallen te vinden die aan de regel voldeden, maar om de regel zelf te ontdekken.

Sindsdien is er een heleboel onderzoek gedaan naar confirmatiebias in allerlei verschillende gecontroleerde laboratoriumomstandigheden. Het beroemdste daarvan is misschien wel de ingenieuze Selectietaak, die ook door Wason werd ontwikkeld (1968).⁷ De Selectietaak werkt als volgt. Stel dat ik vier kaarten voor je neerleg, waar respectievelijk D, B, 3 en 7 op staat (zie Figuur 1.1). Ik vertel je dat als een kaart een letter op de ene zijde heeft, er een cijfer op de andere zijde staat, en ik geef je een specifiekere regel (hypothese) die juist of onjuist kan zijn: ‘Als er een D op één zijde van een van de kaarten staat, staat er een 3 op de andere zijde.’ Tot slot vraag ik je om me te vertellen welke kaarten je zou moeten omdraaien om te kunnen vaststellen of deze regel juist of onjuist is. Een kaart met cruciale informatie onomgedraaid laten of een kaart die geen informatie bevat (d.w.z. waarmee de regel niet kan worden getoetst) omdraaien wordt beschouwd als een onjuist antwoord. Bedenk voordat je verder leest welke kaarten jij zou omdraaien en welke je zou laten liggen.



FIGUUR 1.1 De Selectietaak van Peter Wason voor het meten van confirmatiebias. Er worden vier kaarten omgekeerd op een tafel gelegd. Je krijgt te horen dat als er aan de ene kant een letter staat, er altijd een cijfer op de andere kant staat. Vervolgens krijg je een specifieke hypothese: als er een D op de ene kant staat, staat er een 3 op de andere kant. Welke kaarten zou je omdraaien om deze hypothese te toetsen?

Als je de D koos en de B niet, ben je in goed gezelschap. Beide antwoorden zijn correct en worden gegeven door de meerderheid van de proefpersonen. Door de D te selecteren probeer je de regel te toetsen door hem te bevestigen, en de B vermijden is correct omdat de achterkant van die kaart je geen nuttige informatie oplevert, ongeacht de uitkomst.

Koos je ook voor de 3? Wason ontdekte dat de meeste proefpersonen dat deden, terwijl ze de 3 juist zouden moeten laten liggen. Als er op de achterkant geen D staat, komen ze namelijk niets te weten; de regel stelt dat kaarten met een D op de ene zijde een 3 op de andere zijde hebben, maar niet dat

de D de enige letter is die aan een 3 gekoppeld kan zijn (het trekken van zo'n conclusie is een logische denkfout die bekend staat als 'bevestiging van het consequent'). En als de achterkant wel een D was, zou de uitkomst weliswaar consistent zijn met de regel maar zou die hem niet bevestigen, om precies dezelfde reden.

Heb je tot slot voor de 7 gekozen, of die juist vermeden? Interessant genoeg ontdekte Wason dat maar weinig proefpersonen de 7 uitkozen, hoewel dat de juiste strategie is – net zo juist als de keuze voor D. Als de achterkant van de 7 een D bleek te zijn, zou de regel namelijk categorisch zijn ontkracht – dit heet 'bewijs door contrapositie'. Hierin ligt de cruciale uitkomst besloten: het feit dat de meeste proefpersonen correct de D selecteren maar niet de 7, vormt bewijs dat mensen eerder regels of hypothesen proberen te toetsen door ze te bevestigen dan door ze te weerleggen.

Wasons bevindingen vormden het eerste in een gecontroleerde laboratoriumomgeving verkregen bewijs van confirmatiebias, maar informele observaties wezen al eeuwen sterk in die richting. In een uiterst belangrijk onderzoek merkte psycholoog Raymond Nickerson op dat er al tijdens de heksenprocessen in de middeleeuwen sprake was van confirmatiebias.⁸ In veel van die processen werden voorbarige conclusies getrokken waarmee men de schuld van de verdachte alleen maar probeerde te bevestigen. Om bijvoorbeeld te testen of een vrouw een heks was, werd zij vaak met stenen aan haar voeten in het water gedompeld. Als ze bovenkwam, vormde dat het bewijs dat ze een heks was en werd ze op de brandstapel verbrand. Verdronk ze, dan werd ze meestal onschuldig bevonden, of als een heks met minder kracht bestempeld. Hoe dan ook stond verdacht worden van hekserij gelijk aan een terdoodveroordeling binnen een juridisch raamwerk dat er alleen op uit was om beschuldigingen te bevestigen. Vergelijkbare vertekeningen zijn terug te vinden in veel aspecten van het moderne leven. Populaire tv-programma's als CSI wekken de indruk dat de forensische wetenschap vrij van vooringenomenheden en onfeilbaar is, maar in werkelijkheid wordt het vakgebied geteisterd door confirmatiebias.⁹ Zelfs forensisch onderzoekers van de meest gerespecteerde bureaus ter wereld hebben soms de neiging om bewijs zodanig te interpreteren dat het bestaande vermoedens bevestigt. Dit kan tot onterechte veroordelingen leiden, zelfs wanneer het bewijs gebaseerd is op hardere gegevens als vingerafdrukken en DNA-tests.

Confirmatiebias duikt ook op in de wereld van de wetenschapscommunicatie. Jarenlang werd aangenomen dat aanvulling van kennislacunes bij het publiek met feiten – het zogenaamde deficietmodel¹⁰ – de sleutel was tot effectievere communicatie van wetenschappelijke informatie. Recenter

is dit idee echter in diskrediet geraakt omdat het geen rekening houdt met bestaande opvattingen. Ook de mate waarin wij nieuwe informatie opnemen over kwesties als klimaatverandering, vaccinatie of genetisch gemodificeerd voedsel wordt beïnvloed door confirmatiebias; we hebben een voorkeur voor bewijs dat consistent is met onze vooronderstellingen, terwijl bewijs dat daartegen indruist genegeerd of aangevochten wordt. Als gevolg van deze bias gaan mensen niet automatisch rationeler denken als ze meer feiten krijgen aangereikt. Hetzelfde probleem doet zich voor in de politiek. In zijn gedenkwaardige boek uit 2012, *The Geek Manifesto*, betreft Mark Henderson het dat politici selectief gebruikmaken van bewijsmateriaal om een vooraf bepaalde agenda te bekrachtigen. Het daaruit voortkomende ‘op beleid gebaseerde bewijs’ is een perfect voorbeeld van confirmatiebias in de praktijk en vormt de antithese van hoe wetenschap moet worden ingezet bij de formulering van op bewijs gebaseerd beleid.

Als confirmatiebias zo irrationeel en contraproductief is, waarom bestaat het dan? Daarvoor zijn veel verschillende verklaringen, die gebaseerd zijn op cognitieve en motivationele factoren. Volgens sommige onderzoekers geeft het blijk van een fundamentele beperking van de menselijke cognitie. Het feit dat wij onvolledige informatie over de wereld hebben zou ons ertoe dwingen om af te gaan op de herinneringen die het gemakkelijkst terug te halen zijn (de zogenaamde beschikbaarheidsheuristiek). Deze afhankelijkheid zou een voorkeur in de hand werken voor wat we denken al te weten. Maar er zijn ook mensen die beweren dat confirmatiebias het gevolg is van een aangeboren ‘confirmatorische teststrategie’ – een term die in 1987 werd geïntroduceerd door de psychologen Joshua Klayman en Young-Won Ha.¹¹ We weten al dat mensen het gemakkelijker vinden om van een positieve bewering te beoordelen of zij waar of onwaar is (bijv. ‘Er zitten appels in de mand’) dan van een negatieve (‘Er zitten geen appels in de mand’). Omdat een oordeel over iets wat aanwezig is gemakkelijker is dan een oordeel over iets wat afwezig is, zou het kunnen zijn dat we positieve toetsingen van de werkelijkheid prefereren boven negatieve. Doordat we voor de gemakkelijke weg kiezen, zou deze voorkeur voor positieve gedachten ons er ten onrechte toe kunnen aanzetten bewijs te aanvaarden dat onze eerdere overtuigingen onderschrijft.

Tegen deze verklaringen voor de hardnekkigheid van irrationele vooringenomenheid brengen psychologen Hugo Mercier en Dan Sperber in dat confirmatiebias eigenlijk volkomen rationeel is in een samenleving waarin het winnen van discussies belangrijker is dan het achterhalen van de waarheid.¹² Als kind wordt ons voortdurend geleerd de opvattingen die we erop

nahouden te verdedigen en te rechtvaardigen, niet zozeer om er kritisch naar te kijken. Door nieuwe informatie zodanig te interpreteren dat zij in lijn is met onze bestaande vooronderstellingen vergroten we ons zelfvertrouwen en komen we in discussies overtuigender over. Dat zorgt er op zijn beurt voor dat anderen ons eerder macht en sociale overredingskracht toedichten. Deze observatie brengt ons tot een voor de hand liggende stelling: als de maatschappij zo in elkaar zit dat winnen lonender is dan gelijk hebben, is het niet verwonderlijk dat we zulke incentives weerspiegeld zien in onze manier van wetenschap bedrijven.

Neoflie: als positief en nieuw het wint van negatief maar waar

De carrière van elke onderzoekspsycholoog – en eigenlijk ook van veel andere wetenschappers – valt of staat met het tempo waarin hij empirische artikelen in kwalitatief goede, collegiaal getoetste tijdschriften weet te publiceren. Aangezien het proces van collegiale toetsing (soms extreem) concurrerend is, staat publiceren in de meest gerenommeerde tijdschriften in de academische wereld gelijk aan winnen.

Tijdschriftredacteuren en reviewers beoordelen aangeboden manuscripten op vele gronden. Ze kijken naar fouten in de logica achter experimenten, naar de onderzoeksmethodologie en naar de analyses. Ze bestuderen de inleiding om vast te stellen of de hypothesen stevig genoeg geworteld zijn in eerder onderzoek. Ze nemen de discussie nauwgezet onder de loep om te bepalen of de conclusies van het paper gerechtvaardigd worden door het bewijs. Maar reviewers doen meer dan kritisch kijken naar de gedachtegang, de methodologie en de interpretatie van een paper. Ze bestuderen ook de resultaten zelf. Hoe belangrijk zijn die? Hoe opzienbarend? Hoeveel hebben we van dit onderzoek geleerd? Is het een doorbraak? Een van de centrale (en zoals we zullen zien betreurenswaardige) waarheden in de psychologie is dat opwindende, positieve resultaten een cruciale factor bij – en vaak een voorwaarde voor – publicatie zijn. De boodschap aan onderzoekers is simpel: als je wilt winnen in de academische wereld, publiceer dan zoveel mogelijk papers waarin je met positieve, nieuwe resultaten komt.

Wat zijn ‘positieve’ resultaten? Positiviteit in deze context betekent niet dat de resultaten opbeurend moeten zijn of goed nieuws moeten brengen – het gaat erom dat de onderzoekers een betrouwbaar verschil in metingen of een betrouwbare relatie hebben gevonden tussen twee of meer onderzoeksvariabelen. Stel bijvoorbeeld dat je het effect van een cognitieve trai-

ning wilt meten op het succes van lijnen bij mensen die proberen af te vallen. Allereerst voer je een literatuuronderzoek uit, en op basis van eerdere onderzoeken stel je vast dat het zou kunnen helpen om de zelfbeheersing van mensen te bevorderen. Gewapend met een goed beeld van bestaand werk zet je een onderzoek op waarin twee groepen worden gevolgd. De experimentele groep voert een computertaak uit waarbij de deelnemers worden getraind om te reageren op afbeeldingen van voedingsmiddelen, maar vooral niet te reageren op afbeeldingen van bepaalde soorten junkfood. Deze taak voeren ze zes weken lang elke dag uit, en jij meet aan het einde van het experiment hoeveel ze zijn afgevallen. De controlegroep voert een vergelijkbare taak uit met dezelfde afbeeldingen, maar reageert op alle afbeeldingen. Ook bij deze groep meet je het gewichtsverlies.

De nulhypothese (H_0) luidt in dit geval dat er geen verschil in gewichtsverlies zou moeten zijn – jouw training heeft geen effect op het gewicht van de proefpersonen. De alternatieve hypothese (H_1) is dat de training bevorderlijk is voor het vermogen van mensen om junkfood te laten staan. In dat geval zou het gewichtsverlies groter moeten zijn bij de behandelgroep dan bij de controlegroep. Er zou sprake zijn van een positief resultaat als er een statistisch significant verschil in gewichtsverlies tussen de twee groepen werd gevonden (in technische termen: als H_0 werd verworpen), en van een negatief resultaat als er geen significant verschil werd gevonden (met andere woorden: onvermogen om H_0 te verwerpen). Let op het feit dat ik het woord ‘onvermogen’ gebruik. Dit taalgebruik is cruciaal; in onze huidige academische cultuur beschouwen tijdschriften zulke uitkomsten inderdaad als staaltjes van wetenschappelijk onvermogen. Ongeacht het feit dat aan beide uitkomsten dezelfde gedachtegang en methodes ten grondslag liggen, merken psychologen dat negatieve resultaten veel moeilijker te publiceren zijn dan positieve. Dit komt doordat positieve resultaten in de ogen van tijdschriften een grotere mate van wetenschappelijke vooruitgang weerspiegelen en van groter belang zijn voor lezers. Een tijdschriftredacteur zei ooit tegen mij: ‘Sommige resultaten zijn gewoon interessanter en belangrijker dan andere. Als ik een gerandomiseerd onderzoek doe naar een nieuwe interventie die gebaseerd is op een vergezochte veronderstelling en geen effect vind, is dat geen grote sprong voorwaarts. Maar als uit datzelfde onderzoek blijkt dat de interventie een zeer groot positief effect heeft, is dat een belangrijker bevinding.’

Deze publicatiebias in de richting van positieve resultaten vloeit ook voort uit de aard van conventionele statistische analyses in de psychologie. Bij gebruik van standaardmethodes ontwikkeld door Neyman en Pearson

leiden positieve resultaten tot verwerping van H_0 ten gunste van de alternatieve hypothese (H_1). Met deze statistische benadering – nulhypothese-significantietoetsing (NHST) genoemd – maakt men een inschatting van de waarschijnlijkheid (p) dat er een effect van dezelfde of grotere omvang wordt bereikt als de nulhypothese juist zou zijn. Wat dus niet wordt ingeschat, is de waarschijnlijkheid dat de nulhypothese zelf juist is; p -waarden geven een inschatting van de waarschijnlijkheid van een bepaald effect of een extremer resultaat gezien de hypothese, niet van de waarschijnlijkheid van een specifieke hypothese gezien het effect. Dit betekent dat een statistisch significant resultaat (standaard $p < .05$) de onderzoeker weliswaar in staat stelt om H_0 te verwerpen, maar dat een niet-significant resultaat ($p > .05$) de onderzoeker niet in staat stelt om H_0 te aanvaarden. Het enige wat de onderzoeker uit een niet-significante uitkomst kan concluderen is dat H_0 juist zou kunnen zijn, of dat de data wellicht ongevoelig zijn. De interpretatie van statistisch niet-significante effecten is daardoor inherent niet doorslaggevend.

Bedenk eens welk gedachteproces dit oplevert in de hoofden van onderzoekers. Als we niet rechtstreeks kunnen toetsen of er geen verschil is tussen experimentele condities, heeft het weinig nut om een experiment op te zetten waarin de nulhypothese het brandpunt van de belangstelling vormt. In plaats daarvan worden psychologen getraind om experimenten op te zetten waarin belangwekkende bevindingen altijd positief zijn. Deze bias bij het opzetten van experimenten zorgt er op zijn beurt voor dat psychologiestudenten aan hun onderzoekscarrière beginnen met in hun achterhoofd het mantra ‘nooit de nulhypothese voorspellen.’ Maar als onderzoekers nooit de nulhypothese kunnen voorspellen, en als positieve resultaten interessanter zijn voor tijdschriften dan negatieve resultaten, dan is het onvermijdelijke gevolg een bias waarbij de intercollegiaal getoetste literatuur gedomineerd wordt door positieve bevindingen die de H_0 verwerpen ten faveure van de H_1 , en waarbij het grootste deel van de negatieve of niet-significante resultaten ongepubliceerd blijft. Om ervoor te zorgen dat ze punten blijven scoren in de academische wereld worden onderzoekers gepusht om positieve resultaten te vinden die overeenkomen met hun verwachtingen – een mechanisme dat confirmatiebias stimuleert en beloont.

Dit alles zou in theorie mogelijk kunnen zijn, maar werkt het in de praktijk ook echt zo? Psychologen weten sinds de jaren vijftig van de twintigste eeuw dat tijdschriften eerder geneigd zijn om positieve resultaten te publiceren, maar het is van oudsher lastig om de rol van publicatiebias in de psychologie te kwantificeren.¹³ Een van de meest overtuigende analyses kwam in 2010 van psycholoog Daniele Fanelli van de University of Edinburgh.¹⁴

Fanelli redeneerde, zoals hierboven, dat elk terrein van de wetenschappelijke literatuur dat onderhevig is aan publicatiebias gedomineerd zou moeten worden door positieve resultaten die de opgestelde hypothese (H_1) ondersteunen. Om dit idee te toetsen, bekeek hij een willekeurige steekproef van meer dan 2000 gepubliceerde tijdschriftartikelen binnen het hele wetenschappelijke spectrum, van ruimtewetenschap tot scheikunde, natuurkunde, biologie, psychologie en psychiatrie. De resultaten waren opvallend. Voor alle wetenschappen gold dat positieve uitkomsten vaker voorkwamen dan negatieve. Zelfs binnen de ruimtewetenschap, met het hoogste percentage gepubliceerde negatieve bevindingen, ondersteunde 70 procent van de artikelen uit de steekproef de opgestelde hypothese. Cruciaal was dat deze bias het duidelijkst naar voren kwam in de psychologie, met een recordpercentage van 91. Het is ironisch dat de psychologie – de discipline die het eerste empirische bewijs van confirmatiebias leverde – er tegelijkertijd het meest gevoelig voor is.

De drang om positieve resultaten te publiceren is een belangrijke oorzaak van publicatiebias, maar het verklaart slechts de helft van het probleem. De andere verklaring vormt de behoefte aan nieuwe resultaten. Bij veel tijdschriften geldt dat artikelen eerder in aanmerking komen voor publicatie als ze ofwel een nieuwe methodologie beschrijven ofwel een nieuwe bevinding opleveren – en bij voorkeur allebei. De meeste tijdschriften die psychologisch onderzoek publiceren beoordelen de merites van manuscripten gedeeltelijk op de nieuwe elementen die zij bevatten. Sommige benoemen dit zelfs expliciet als een randvoorwaarde voor publicatie. Het tijdschrift *Nature* stelt dat resultaten ‘nieuw’ en ‘boeiend’¹⁵ moeten zijn om in aanmerking te komen voor collegiale toetsing, en het tijdschrift *Cortex* geeft aan dat empirische onderzoeksrapporten ‘belangrijk en nieuw materiaal moeten bevatten.’¹⁶ Het tijdschrift *Brain* waarschuwt auteurs dat ‘sommige [manuscripten] zonder collegiale toetsing worden afgewezen als gevolg van gebrek aan nieuwe elementen.’¹⁷ *Cerebral Cortex* gaat nog een stap verder: zelfs na collegiale toetsing ‘is de uiteindelijke acceptatie van een paper niet alleen afhankelijk van de technische merites, maar ook van subjectieve beoordeling van de nieuwe elementen die het bevat.’¹⁸ Binnen de psychologie zelf geeft *Psychological Science*, naar eigen zeggen het hoogst aangeschreven tijdschrift in de psychologie, prioriteit aan papers die ‘adembenemende’ bevindingen bevatten.¹⁹

Je zult je nu misschien afvragen: wat is er mis met de eis dat bevindingen nieuw moeten zijn? Iets wat als een ontdekking wordt aangeduid kan toch ook nog niet eerder geobserveerd zijn (dus moet het wel een nieuw resultaat

zijn)? En is het niet redelijk om aan te nemen dat onderzoekers die nieuwe resultaten proberen te produceren misschien ook nieuwe methodes moeten hanteren? Met andere woorden, hechten tijdschriften die 'nieuw' hoog in het vaandel hebben staan, niet gewoon waarde aan ontdekking? Het probleem van dit argument is de onderliggende aanname dat elke observatie binnen het psychologische onderzoek een ontdekking kan worden genoemd – dat elk paper een duidelijk en onherroepelijk feit beschrijft. Net als bij alle andere wetenschappelijke disciplines is dit ver bezijden de waarheid. De meeste onderzoeksbevindingen in de psychologie zijn eerder probabilistisch dan deterministisch; conventionele statistische tests communiceren in termen van waarschijnlijkheden en niet van bewijzen. Dit betekent dat geen enkel op zichzelf staand onderzoek of paper aanspraak kan maken op een ontdekking. Een ontdekking is volledig en zonder uitzondering afhankelijk van in hoeverre de oorspronkelijke resultaten door andere wetenschappers herhaald of gerepliceerd kunnen worden, en niet maar één keer, maar steeds weer opnieuw. Het is bijvoorbeeld niet genoeg om maar één keer te melden dat een bepaalde cognitieve therapie effectief is bij het verminderen van depressieve verschijnselen; dat resultaat moet een heleboel keer bij verschillende groepen patiënten en door verschillende groepen onderzoekers worden herhaald voordat de nieuwe interventie in de gezondheidszorg wordt geïntroduceerd. Wanneer een resultaat eenmaal een bevredigend aantal keren met dezelfde experimentele methode is gerepliceerd kan het als repliceerbaar worden beschouwd en kan het, in combinatie met ander repliceerbaar bewijs, een betekenisvolle bijdrage leveren aan het theoretische of toegepaste raamwerk waaronder het valt. Na verloop van tijd kan grootschalige accumulatie van repliceerbaar bewijs vanuit verschillende vakgebieden ervoor zorgen dat theorieën algemeen aanvaard worden en kunnen het in sommige gevallen zelfs wetmatigheden worden.

In de wetenschap staat de nadruk op het nieuwe ontdekking eerder in de weg dan dat het er bevorderlijk voor is, omdat het de waarde van (bijna exacte) directe replicatie ondermijnt. Zoals we hebben gezien, zijn journalisten de poortwachters van academische carrières. Dus als die vooral waarde hechten aan bevindingen die positief en nieuw zijn, waarom zouden wetenschappers dan proberen elkaar te repliceren? Binnen een structuur die het nieuwe beloont, wordt directe replicatie afgedaan als saai, niet creatief en een intellectueel zwaktebod.

Maar zelfs in een onderzoekssysteem dat gedomineerd wordt door positieve resultaten en liefde voor het nieuwe hebben psychologen nog enig besef dat reproduceerbaarheid belangrijk is. Daarom heeft de gemeenschap

onaantrekkelijke directe replicatie vervangen door een alternatieve vorm van validering waarbij men kan zeggen dat het ene experiment het kernconcept of het thema van een ander experiment repliceert maar waarbij een andere (nieuwe) methode wordt gebruikt. Dit proces wordt conceptuele replicatie genoemd. Zo op het eerste gezicht lijkt deze herdefinitie van replicatie tegemoet te komen aan de behoefte om eerdere bevindingen te valideren en tegelijkertijd voor nieuwe elementen te zorgen. Helaas levert dit alleen maar een heel nieuwe schadelijke vorm van confirmatiebias op.

Replicatie van concepten, niet van experimenten

Begin 2012 lanceerde John Bargh, hoogleraar psychologie aan Yale University, een scherpe publieke aanval op een groep onderzoekers die een van zijn eerdere bevindingen niet had weten te repliceren.²⁰ Volgens het onderzoek in kwestie, dat in 1996 door Bargh en collega's was gepubliceerd, liepen proefpersonen die er onbewust toe werden aangezet om te denken aan concepten die te maken hadden met ouderdom (bijv. via woorden als 'gepensioneerd', 'rimpel' en 'oud'), langzamer als zij aan het eind van het experiment het laboratorium verlieten.²¹ Op basis van deze bevindingen beweerde Bargh dat mensen opmerkelijk vatbaar zijn voor automatische priming door sociale constructen.

Het paper van Bargh was direct een succes en is al meer dan 3800 keer geciteerd. Binnen de sociale psychologie heeft het de aanzet gegeven tot een hele reeks onderzoeken naar sociale priming, die sindsdien in allerlei verschillende contexten zijn toegepast. Vanwege de impact van het paper lijkt het redelijk om aan te nemen dat de centrale bevindingen ervan vele keren gerepliceerd en bevestigd zijn. Maar schijn kan bedriegen.

Verschillende onderzoekers hebben aangegeven dat zij het oorspronkelijke onderzoek van Bargh niet hebben weten te repliceren, maar slechts weinig van deze niet gelukte replicaties zijn gepubliceerd. Dit is te wijten aan het feit dat tijdschriften (en reviewers) niet van negatieve bevindingen houden en directe replicaties vaak weigeren te publiceren. Eén zo'n poging tot replicatie, in 2008 door Hal Pashler en collega's van de University of California in San Diego, is bijvoorbeeld nooit in een academisch tijdschrift gepubliceerd; het paper is alleen te vinden in een online archief dat Psych-FileDrawer heet.²² Hoewel zij de omvang van de in het oorspronkelijke onderzoek gebruikte steekproef meer dan verdubbeld hadden, troffen Pashler en zijn team geen bewijs aan van sociale priming – ze vonden eerder nog het tegenovergestelde effect.