

recte afstammeling van De Groots AET – is de meting van de vaardigheden van basisschoolleerlingen nog steeds een van de drie onderzoeklijnen. Dit onderzoek betreft nu web-applicaties waarmee leerlingen steeds op hun eigen niveau, met dezelfde mate van uitdaging en succes, kunnen oefenen. Een voorbeeld is Reken tuin.nl, een systeem dat al door bijna tweeduizend scholen wordt gebruikt. Het is gebaseerd op nieuwe psychometrische inzichten en combineert onderwijs met toetsing. Er is geen momentopname meer, Reken tuin wordt gebruikt in alle schooljaren. Dit geldt overigens ook voor het leerlingvolgsysteem van het Cito dat een steeds grotere rol gaat spelen in de advisering. De effectiviteit van Reken tuin in schooladviezen moet

nog bepaald worden, maar het is een van de wegen naar een modernisering van de eindtoets.

Rede als richtsnoer

Het kritische werk van De Groot sloot aan bij dat van de Amerikaanse psycholoog Paul Meehl, die in zijn *Clinical versus statistical prediction* uit 1954 liet zien dat menselijk gedrag meestal beter wordt voorspeld met een statistische combinatie van gegevens dan met het klinisch oordeel. De Groot heeft voor verschillende onderdelen van de toegepaste psychologie aangetoond dat het klinisch oordeel het meestal aflegt tegen objectieve en gestandaardiseerde meetprocedures. Hij was echter geen harde kwantitatieve methodoloog, integendeel. Hij hield zelf meer van

kwalitatieve analyses van onder andere theoretische begrippen, ‘hardopdenk-protocollen’ en zelfrapportages. Hij was echter wél de voorvechter van de rede. Wetenschapsbeoefening moet rationeel zijn: theorieën moeten consistent zijn, hypothesen toetsbaar, meetprocedures deugdelijk, analyses correct, onderzoek openbaar en resultaten controleerbaar. Het kwam fraai tot uiting in de titel van het *liber amicorum* dat hem werd aangeboden bij zijn afscheid van de Universiteit van Amsterdam in 1979, *Rede als richtsnoer*.

Don Mellenbergh was hoogleraar psychologische methodenleer aan de Universiteit van Amsterdam tot 1999, Han van der Maas bekleedt die leerstoel sinds 2005.

De verleidingen van foponderzoek

Ruim een halve eeuw geleden beschreef De Groot het strakke keurslijf waarin de wetenschap zich moet dwingen wil zij tot betrouwbare resultaten komen. Zijn voorschriften beginnen nu pas een beetje door te dringen.

DOOR
ERIC-JAN WAGENMAKERS
EN VITTORIO BUSATO

ADRIANUS Dingeman de Groot kan met recht beschouwd worden als de vader van de Nederlandse methodologie. De Groot had uitgesproken ideeën over de manier waarop onderzoekers op een verantwoorde manier conclusies kunnen trekken uit hun empirische bevindingen. Deze ideeën presenteerde De Groot systematisch en met veel overtuiging in het 421 pagina's tellende boekwerk *Methodologie: Grondslagen van onderzoek en denken in de gedragswetenschappen* (eerste druk 1961; laatste en twaalfde druk 1994; Engelse vertaling: 1969). Skeptici

in binnen- en buitenland vinden in *Methodologie* een reeks tijdloze inzichten over wetenschappelijke kennisvergarig. Zijn oude inzichten beginnen de laatste jaren nieuw elan te verwerven.

Taaie kost

De Groot was zijn tijd ver vooruit, en de inzichten in *Methodologie* zijn van blijvend belang voor alle empirische disciplines. Niettemin is er in Nederland nauwelijks meer een student te vinden die *Methodologie* heeft gelezen – het staat natuurlijk ook niet op de lijst boeken in het curriculum. En in

het buitenland is het nog erger; daar wordt De Groot vooral geassocieerd met zijn schaakonderzoek, terwijl zijn bijdrage aan de methodologie nooit de erkenning heeft gekregen die het verdiende.

Nu was De Groot hier in zekere zin zelf schuldig aan. Hij publiceerde amper in internationale tijdschriften, en *Methodologie* is taaie kost. De Groot biedt een gedetailleerde systematiek, maar geeft weinig concrete voorbeelden en presenteert geen enkele inzichtelijke figuur. Als gevolg daarvan biedt hij de lezer nauwelijks tijd om op adem komen. Ter illustratie slaan wij zijn boek open op een willekeurige pagina en lezen de eerste twee zinnen (De Groot hield erg van cursief):

De empirische relaties van het type B omvatten allereerst uitspraken over feitelijke bevindingen, resultaten, uitkomsten van onderzoekingen. Maar tussen deze beide uitersten in vindt men een grote categorie van uitspraken, die men op verschillende wijze, d.w.z. of als A of als B kan opvatten.

Wanneer een dergelijke stijl vierhonderd pagina's stug wordt volgehouden, zal zelfs de meest gemotiveerde lezer op enig moment de vlag strijken.

Wij zijn grote bewonderaars van het werk van De Groot — een van ons heeft onlangs met collega's een artikel van De Groot in het Engels vertaald, de ander verzorgde een biografische schets over hem — maar zelfs ons lukt het niet *Methodologie* van kaft tot kaft te lezen. Eigenlijk zou het werk moeten worden gemoderniseerd en kernachtig worden samengevat voor beginnende wetenschappers.

Voorspellen

Gelukkig is het niet allemaal kommer

en kwel. Soms stopt De Groot even met categoriseren en systematiseren en wordt hij oprecht enthousiast. Op deze momenten begint de tekst te leven.

De centrale vraag waarover De Groot zich in *Methodologie* buigt, is hoe wetenschappers betrouwbare kennis kunnen verwerven. Hiertoe moeten we dit soort kennis eerst definiëren. De Groot gebruikt een aantrekkelijke en algemene definitie gebaseerd op de mate van voorspellend vermogen:

Het criterium bij uitstek voor 'ware kennis' is namelijk gelegen in het kunnen voorspellen van de uitkomst van een toetsingsprocedure. Als ik iets weet, kan ik iets voorspellen; kan ik niets voorspellen dan weet ik niets.

Maar hoe doe je zuivere voorspellingen? Welnu, in ieder geval niet door eerst naar je data te kijken! Een voorspelling moet dus van tevoren worden opgesteld. Dit klinkt kinderlijk eenvoudig, maar deze regel wordt in het wetenschappelijke metier zelden nageleefd, althans niet op rigoureuze en transparante wijze (een uitzondering betreft wellicht het huidige onderzoek naar de effectiviteit van medicijnen, waar het analyseplan vooraf moet worden gepubliceerd — maar zie het Compare-project van Ben Goldacre, <http://compare-trials.org>).

De Groot is zich ervan bewust dat het in onderzoek niet alleen maar gaat om voorspellen — in bepaalde stadia is het van belang om iets uit te proberen, dat wil zeggen om te exploreren en hypothesen te genereren in plaats van hypothesen te toetsen. Daarom presenteert De Groot de 'empirische cyclus', een *perpetuum mobile* van wetenschappelijke activiteiten: als dat wordt toegepast, leidt het noodzakelijkerwijs tot een toename in kennis.



Bem dacht dat hij had aangetoond dat mensen paranormaal in de toekomst kunnen kijken.

Empirische cyclus

De beschrijving van de empirische cyclus neemt het grootste deel van *Methodologie* in beslag. De figuur op pagina 28 laat de belangrijkste componenten zien. De cyclus begint bovenaan, bij 'bestaande kennis en bestaande data'. Vervolgens begint er een exploratieve exercitie, die resulteert in een nieuwe hypothese en een daarop gebaseerde nieuwe voorspelling. Deze nieuwe voorspelling wordt vervolgens getoetst in een nieuw experiment – daarmee begint de confirmatieve fase van de cyclus. Het is hier van essentieel belang dat de nieuwe hypothese niet getoetst mag worden op dezelfde data die tot deze hypothese hebben geleid (de rode lijn in de figuur).

Een zelfde punt is uiteraard gemaakt door andere wetenschappers, zoals bijvoorbeeld al in 1878 de Amerikaanse filosoof Charles Peirce en in 1938 door Hans Reichenbach, door de natuurkundige Richard Feynman in zijn befaamde essay 'Cargo cult science' uit 1974, de reeds genoemde Ben Goldacre en de laatste jaren zelfs door redacteurs van wetenschappelijke tijdschriften zoals Chris Chambers van *Cortex* in 2013.

Precognitie

Laten wij als concreet voorbeeld nemen het werk van Daryl Bem, een vermaarde sociaal-psycholoog. Hij publiceerde in 2011 een artikel in het *Journal of Personality and Social Psychology*, het vlaggenschip van de sociale psychologie. Bem dacht dat hij had aangetoond dat mensen precognitie hebben ('paranormaal in de toekomst kunnen kijken'). In Bems experiment moesten proefpersonen gokken aan welke kant van een computerscherm er een plaatje zou verschijnen. Nadat de proefpersoon een keuze had gemaakt werd de locatie van het plaatje

random bepaald. Nu bleek dat proefpersonen beter presteerden dan puur op grond van het toeval mocht worden verwacht. Maar dat gold alleen voor erotische plaatjes (en niet voor romantische, positieve, negatieve en neutrale plaatjes). Het effect was zelfs bijzonder sterk voor extraverte vrouwen. Op basis van deze beschrijving alleen al is het duidelijk dat de hypothese ('extraverte vrouwen hebben precognitie maar alleen voor erotische plaatjes') achteraf is opgesteld, nadat de data waren bekeken. Dat is op zich prima, maar deze hypothese is dus niet door de data getoetst: zij is enkel door de data gesuggereerd. Voor het toetsen van de gesuggereerde hypothese is – indachtig De Groot – een nieuw experiment nodig.

Top of fop

Helaas is het zeer verleidelijk voor onderzoekers om een groot gedeelte van de empirische cyclus af te snijden (de rode lijn, 'jezelf foppen', in de figuur). De onderzoeker kijkt naar de data met een open blik (zoals De Groot zegt: 'laten wij eens zien wat wij kunnen vinden'), ziet een verband, en test datzelfde verband dan middels een statistische toets op dezelfde data. Maar een dergelijke toets is betekenisloos, en de onderzoeker houdt zichzelf hier voor de gek. De Groot stelt:

Men 'mag' gerust statistische toetsen toepassen in een exploratie-onderzoek, als men er zich maar rekenschap van geeft, dat zij geen bewijskracht hebben. Zij hebben ook geen dwingend karakter meer, hoe zij ook uitvallen.

Om te voorkomen dat toponderzoek vervalt tot foponderzoek moet de wetenschapper zich dan ook dwingen de gehele empirische cyclus te voltooien, en geen bochten af te snijden door



Foto: Klaas Koppe.

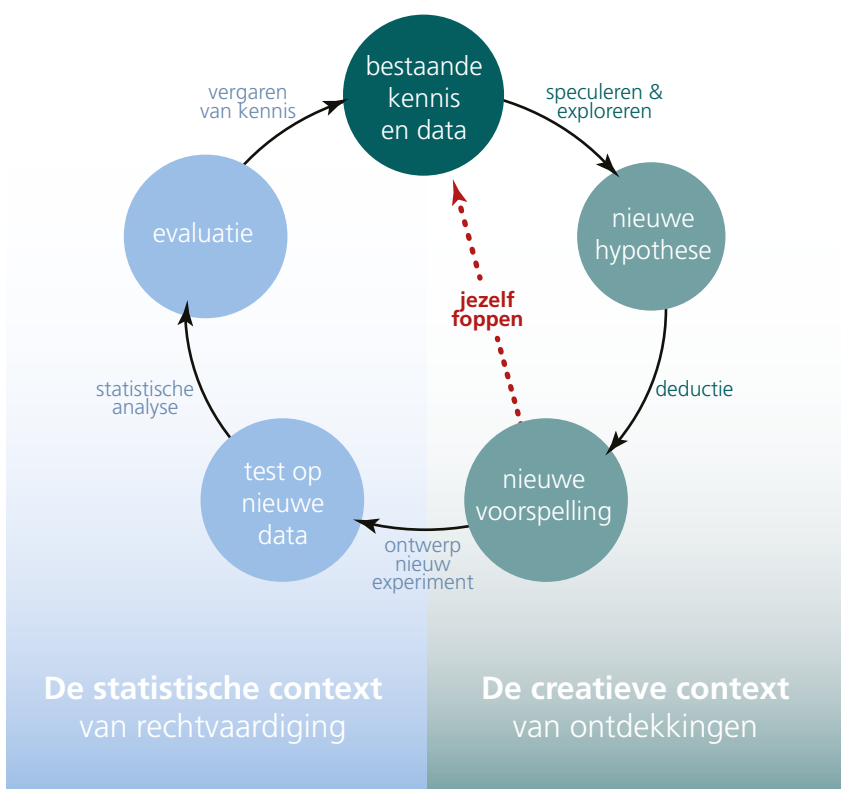
een toets uit te voeren op dezelfde data die deze toets hebben geïnspireerd. De Groot biedt ook hiervoor een uitkomst: het vooraf registreren van analyseplannen. De Groot begint met een voorstel:

Willen wij een onderzoek naar bepaalde consequenties van een theorie of hypothese inderdaad als toetsingsonderzoek (en niet

van het gehele onderzoek op de vorm van een verifieerbare voorspelling wordt gebracht.

Ten slotte, in een sectie ‘Het belang van de analyse vooraf’ weidt De Groot uit over wat dit inhoudt:

Een zo volledig mogelijke uitwerking op papier van de toetsings- (c.q. experimentele) opzet vooraf, is in ieder



als exploratie) opzetten, dan moet een scherpe *formulering vooraf* ter beschikking staan, die afleiding van toetsbare consequenties mogelijk maakt.

Vervolgens verduidelijkt De Groot de zin van die scherpe formulering vooraf voor de onderzoeker:

Daarmee bereikt hij enerzijds, dat hij zichzelf niet in de verleiding brengt, resultaten achteraf ‘goed te praten’, anderzijds, dat de uitkomst

geval sterk aan te bevelen. Deze moet dan bevatten: een korte expositie van de theorie, een formulering van de te toetsen hypothese(n); (...) een precieze beschrijving van de wijze waarop men steekproeven wil trekken of samenstellen; een vastlegging van de confirmatie-criteria, inclusief formulering van eventueel gebruikte nulhypothese(n), keuze van statistische toets(en), significantieniveau en resulterende confirmatie-intervallen.

Tegenvallend

Het is interessant dat er nu, meer dan een halve eeuw na het verschijnen van *Methodologie*, weer steeds meer wordt gepleit voor het pre-registreren van analyseplannen, precies zoals het De Groot voor de geest stond

Dit werpt natuurlijk wel de vraag op waarom het zo lang heeft moeten duren eer deze methode serieus werd genomen. Een andere vraag is hoe we statistische toetsen moeten interpreteren van bestaand wetenschappelijk onderzoek waarvoor de analyseplannen niet zijn onderworpen aan een ‘scherpe formulering vooraf’. De tegenvallende resultaten van recente replicatieonderzoeken doen vermoeden dat dergelijke toetsen inderdaad geen dwingend karakter hebben.

Eric-Jan Wagenmakers is hoogleraar methoden van de psychologie aan de Universiteit van Amsterdam.

Het boek *Methodologie* van De Groot is geheel online te lezen: www.dbnl.org/tekst/groo004metho1_01.

De Groot AD. The meaning of ‘significance’ for different types of research. (translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angélique Cramer, Dora Matzke, Don Mellenbergh, and Han L. J. van der Maas). *Acta Psychol (Amst)*. 2014;148:188-94. PMID: 24589374.

Busato VV, Adriaan de Groot: Meester in de psychologie. In: V. V. Busato, M. van Essen en W. Koops (red.): *Van fenomenologie naar empirisch-analytische psychologie. Vier grondleggers van de psychologie*. Amsterdam: Bert Bakker, 2014.

Wagenmakers EJ, Wetzels R, Borsboom D, van der Maas HL, Kievit RA. An agenda for purely confirmatory research. *Perspect Psychol Sci*. 2012;7:632-8. PMID: 26168122.