**Het positivisme en zijn ongeldigheid (geschreven in 2006)**

dr. Ewald Vervaet

In dit artikel vat ik mijn kritiek op de positivistische methode of kortweg het positivisme, zoals dat onder meer in de Nederlandse psychologie wordt toegepast, samen. §1 zet uiteen wat ik onder het positivisme versta. Vervolgens laat ik zien dat positivisten hun werkwijze legitimeren door naar de natuurwetenschappen te verwijzen (§2) en dat die legitimering onterecht is (§3): natuurwetenschappelijke methode is één, de positivistische lezing daarvan is twee. Dat en waarom de statistische methode van het positivisme niet tot kennisverwerving leidt en ook niet kan leiden, is het onderwerp van §4. In §5 trek ik mijn algehele conclusie.

Met dit artikel maak ik aanspraak op een algemeen geldig betoog, maar ik zal me – als natuurkundige en als psycholoog – zoveel mogelijk beperken tot de psychologie.

**1. Positivisme**

Onder positivisme versta ik die wetenschapsfilosofische en methodologische stroming waarin men onder meer met puntsschalen, correlatiecoëfficiënten en regressieanalyse werkt.1 Dit positivisme is in feite een verbijzondering van het klassieke empirisme. Daarin gaat men ervan uit dat inzichten, wetten en theorieën uit waargenomen feiten zouden voortkomen. Het positivisme voegt aan die gedachte toe, dat men de waargenomen feiten bij voorbaat in getallen zou dienen te vatten om die via formules en berekeningen nader te verwerken en aldus inzichten, wetten en theorieën te verkrijgen. Zo stelt de testpsycholoog Drenth enerzijds: ‘[In de fysica] begint men met observatie en probeert men vanuit deze geobserveerde feiten te komen tot een meer algemeen geldende theorie’. Daarin brengt hij de empiristische gedachte tot uitdrukking. Anderzijds stellen Drenth en Sijtsma, eveneens een testpsycholoog: ‘Evenals het fysische begrip “lengte” kunnen ook psychologische eigenschappen gemeten worden’, terwijl ze een temperatuurschaal met een psychologische test vergelijken. Daarmee drukken ze het getalsmatige van het positivisme uit.2

Om de gedachte nader te bepalen volgen nu drie voorbeelden van positivistisch onderzoek.

Voorbeeld 1. Met de *Nederlandse PersoonlijkheidsVragenlijst* (NPV) kan men iemands score S voor verongelijktheid bepalen. Dat gebeurt met 19 beweringen waarop men ‘juist’, ‘?’ of ‘onjuist’ kan antwoorden, zoals ‘Er zijn maar weinig mensen, die mij begrijpen’ en ‘Ik vertrouw de mensen pas als ik ze goed ken’. Voor elke keer dat men ‘juist’ antwoordt, krijgt men 3 punten; voor elke keer dat men ‘onjuist’ antwoordt, krijgt men 1 punt; als men niet tussen ‘juist’ en ‘onjuist’ kan kiezen, dient men ‘?’ te antwoorden, waarvoor men 2 punten krijgt. Iemands totaalscore S ligt dus tussen 19 en 57. De formule die hierbij wordt aangewend, is:

S = ∑ (over i van 1 tot en met 19) si met si = 1, 2, 3.

Score S van persoon P is met een formule om te rekenen tot een score die P plaatst in de verdeling van de scores S in een aselecte steekproef.

Voorbeeld 2. Met de *Eindtoets Basisonderwijs*, beter bekend als de Citotoets, kan men voor ‘Rekenen-Wiskunde’ de score R van een leerling met 60 vraagstukken bepalen. Voor elk vraagstuk worden vier antwoorden gesuggereerd: één is juist en drie zijn onjuist. Daaruit dient hij (of zij) er één te kiezen. Als hij het juiste antwoord kiest, krijgt hij 1 punt; als hij één van de drie onjuiste antwoorden kiest, krijgt hij 0 punten. Totaalscore R ligt dus tussen 0 en 60. De formule die hierbij wordt aangewend, is:

R = ∑ (over i van 1 tot en met 60) ri met ri = 0, 1.

Score R van leerling L kan met een formule omgerekend worden tot een ‘standaardscore’ die L plaatst in de andere deelnemers binnen een bepaalde regio en die wordt uitgedrukt met een getal tussen 501 en 550.

Voorbeeld 3. De onderzoekers Dijkstra en Van Erven willen het inlevende vermogen van forensisch psychiatrische patiënten bepalen met de zogeheten empathieschaal. De coëfficiënten in de vergelijking waarmee die wordt uitgerekend, zijn verkregen in een regressieanalyse en afgerond op gehele getallen:

E = V1 + V2 + V3 + V4 + V5 + V6 – 4(V3A + V4B + V5B + V6C) – 6.3

In §4 zullen we zien dat de formules van deze drie voorbeelden ten onrechte niet empirisch zijn aangetoond.

**2. Positivisme en de natuurwetenschappen**

Men kan zich bij de drie voorbeelden van positivistische wetenschapsbeoefening afvragen waarom testpsychologen, het Cito en Dijkstra & Van Erven doen zoals ze doen: getalsmatig werken, formules aanwenden en berekeningen uitvoeren. Het antwoord is: men doet zo omdat men denkt dat ook de natuurwetenschappers sedert de zestiende en zeventiende eeuw zo te werk zijn gegaan en dat dat de grote successen van vooral de latere natuur-, sterren- en scheikunde verklaart. Daar hebben we in §1 al een voorbeeld van gezien. Immers, Drenth en Sijtsma vergelijken enerzijds het meten van het begrip ‘lengte’ met het meten van psychologische eigenschappen als intelligentie en persoonlijkheid en anderzijds een temperatuurschaal met een psychologische test.

Hier volgen nog zes voorbeelden uit de psychologie van het legitimeren van positivistisch onderzoek door te verwijzen naar de natuurwetenschappen.4

Voorbeeld A. De Franse pedagoog en psycholoog Binet en de Franse arts Simon, de grondleggers van de testpsychologie: ‘We zijn ervan overtuigd [...] dat men wetenschappelijke precisie en exactheid moet invoeren telkens als men het kan, en men kan dat bijna altijd’, terwijl ze een door hen met tests haalbaar geachte geünificeerde nomenclatuur rond de begrippen idiotie, debiliteit en imbeciliteit vergelijken met de geünificeerde nomenclatuur voor lengtematen en elektrische eenheden in de natuurkunde.5

Voorbeeld B. De Britse psycholoog Spearman: ‘[Gauss’ formules in de meetfouttheorie] hebben niet weinig bijgedragen aan de exactheid van modern natuurkundig en psychologisch onderzoek’ en ‘[Gebrek aan overeenstemming tussen de Gausscurve voor meetwaarden enerzijds en meetwaarden anderzijds] is van dag tot dag minder aanvaardbaar aan het worden daar de psychologie zich tot een steeds exactere wetenschap aan het ontwikkelen is’.6

Voorbeeld C. De Duits-Britse psycholoog Eysenck wil naar analogie van de geschiedenis van de natuurwetenschappen langs kwantitatieve weg een classificatie in de persoonlijkheidsleer introduceren. De elementen daarin zouden geen psychologische dimensies hoeven te hebben, precies zoals ‘snelheid’ in de natuurkunde als een dimensieloos getal te definiëren zou zijn, namelijk als een fractie van de lichtsnelheid.7

Voorbeeld D. De Nederlandse psycholoog en methodoloog De Groot: ‘Hoewel […] er in het kader van hypothesevorming géén strikte eisen ten deze kunnen worden gesteld, zal zonder meer duidelijk zijn, dat doelgerichte systematiek bijzonder nuttig kan zijn. Evidente voorbeelden zijn gemakkelijk te vinden: systematisch volgehouden waarnemingen en metingen (b.v. aan de sterrenhemel, vgl. Newton’s gebruik van Kepler’s en diens gebruik van Tycho Brahe’s waarnemingen, […] een systematische beschrijving en classificatie (b.v. in de zoölogie, vgl. Darwin’s theorievorming’ en ‘In het algemeen ligt trouwens aan het moderne streven naar toepassing van exacte methoden, bijvoorbeeld van een axiomatische opbouw en van mathematische modellen [...], tot op zekere hoogte de analogieredenering ten grondslag, dat werkwijzen die in de wiskunde en de natuurwetenschappen klaarblijkelijk tot grote successen hebben geleid, ook de gedragswetenschappen verder moeten kunnen brengen’.8

Voorbeeld E. Drenth stelt een kleurenblindheidsplaat enerzijds en IQ- en persoonlijkheidstests anderzijds op voet van gelijkheid. Hij vat ze alledrie namelijk op als ‘een systematische classificatie- of meetprocedure waarbij het mogelijk wordt een uitspraak te doen over één of meer empirisch-theoretisch gefundeerde eigenschappen van de onderzochte óf over specifiek niet-testgedrag, door uit te gaan van een objectieve verwerking van reacties van hem/haar, in vergelijking tot die van anderen, op een aantal gestandaardiseerde, zorgvuldig gekozen stimuli’.9

Voorbeeld F. De Nederlandse pedagoog en onderwijskundige Geurtz: ‘Een IQ-test verschilt methodologisch niet van een spectraal-analyse: alle ijzeren voorwerpen hebben een spectrum Y. Wat is ijzer? IJzer is die stof die bij spectraal-analyse spectrum Y vertoont’.10

**3. Vertekend versus correct beeld over wetenschap**

Als psycholoog kan ik me voorstellen dat de stand van zaken er voor een positivistische psycholoog oppervlakkig beschouwd uitziet als gesteld in de voorbeelden A-F van §2: zowel positivisten als de natuurwetenschappers bedienen zich van getallen, formules en berekeningen. Als natuurkundige weet ik echter dat één en ander er voor een natuurwetenschapper en meer nog voor een historicus van de natuurwetenschappen geheel anders uitziet. In de natuurwetenschappen vloeien getallen, formules en berekeningen voort uit een theorie, terwijl ze in positivistisch onderzoek voorop worden gezet. De legitimering door een positivist met ‘Wat ik doe leidt tot kennis, want zo is het ook in de natuurwetenschappen gegaan’ houdt dus geen stand.

We zullen in deze paragraaf zien dat de positivistische verklaring voor het succes van de natuurwetenschappen niet houdbaar is en dat kennis dus ook niet langs positivistische weg verworven kan worden. Daar het positivisme een vorm van empirisme is (§1), sta ik eerst stil bij de niet-houdbaarheid van het empirisme (I) en vervolgens bij die van het positivisme (II). Onderwerpen die meer de statistische methode betreffen, komen apart aan bod in §4.

I. Volgens het empirisme komen inzichten, wetten en theorieën uit waargenomen feiten voort. Dat proces wordt inductie genoemd. Er zijn echter ten minste drie soorten feiten en overwegingen die aantonen dat inductie een niet-bestaand proces is en dus dat het empirisme een niet houdbare verklaring is voor de totstandkoming van kennis.

In de eerste plaats laat de ontwikkelingspsychologie zien dat kennis niet uit de buitenwereld de binnenwereld in komt, maar dat voor kennisverwerving een psychologische structuur vereist is, waar een prikkel uit de buitenwereld in wordt opgenomen.

Voorbeeld 1. Hans van 9 maanden speelt met een poppetje (eraan sabbelen, erop slaan, enzovoort). U neemt het voorzichtig van hem af en verstopt het – langzaam en goed zichtbaar voor Hans – links van hem onder een gele doek. Anders dan twee maanden terug zoekt hij het poppetje door onder die doek te kijken. Hij pakt het en speelt er weer mee. Opnieuw neemt u het na een poosje van hem af en verstopt u het onder die gele doek. Weer zoekt en vindt hij het daar. De derde keer verstopt u het poppetje echter rechts van hem onder een blauw jasje. Hans zoekt weer links onder de gele doek! Welnu, als het waar zou zijn dat kennis inductief verworven zou worden (dus via de zintuigen uit de buitenwereld zou binnenkomen), dan had Hans de derde keer geweten dat het poppetje rechts van hem onder het blauwe jasje zou liggen. Die prikkels zijn echter wel op zijn netvliezen gevallen maar met 9 maanden is zijn psychologische structuur nog van dien aard dat hij die prikkels assimileert aan de opvatting ‘Daar zijn de dingen waar ik ze tot nu toe met succes heb gevonden’.11

Voorbeeld 2. Femke van 28 maanden speelt met een knikkerspel met drie banen. Na een poosje doet u voor hoe de knikker op de helft van baan 2 onderschept kan worden: u legt op die plaats een vinger voordat u de knikker bovenin loslaat. U neemt de onderschepte knikker en vraagt ‘Kun jij dat ook?’. Femke knikt van ja. U: ‘Waar ga je hem pakken?’. Femke, op ¼ van baan 3 wijzend: ‘Daar’. Ze legt daar echter pas een vinger als de knikker aan die plek is voorbijgerold! U doet het nog vele malen voor, ook op andere plaatsen dan de helft van baan 2. Toch blijft ze op de banen 2 en 3 haar vinger te laat leggen (op baan 1 haalt ze de knikker geregeld in door versneld met een vinger achter hem aan te gaan). Opnieuw: als het waar zou zijn dat men zich kennis inductief eigen maakt, dan zou Femke zien hoe u het al die keren voordoet, namelijk door een vinger vóór het rollen van de knikker neer te leggen op de plaats waar u de knikker wilt onderscheppen. Die prikkels vallen weliswaar in haar ogen maar met 28 maanden is haar psychologische structuur zodanig dat ze wel besef heeft van het verleden maar nog niet van de toekomst. Aan die structuur assimileert ze uw voordoen, zodat ze nog geen toekomstvoorspellinkje kan doen als ‘Als ik nu m’n vinger hier leg, kan ik er dadelijk de knikker pakken’.12

In de tweede plaats laat ook de wetenschapsgeschiedenis zien dat kennis niet inductief wordt verworven. Aannemende dat Aristoteles en Kepler op een heldere dag van de zon eenzelfde soort prikkels kregen, oordeelde de eerste ‘s avonds dat de aarde stil had gestaan en de zon eromheen had bewogen, terwijl de tweede dan oordeelde dat de zon was blijven stilstaan en dat de aarde om zijn eigen as was gedraaid. Ook zij beschikten elk over een psychologische structuur waaraan ze die prikkels assimileerden. Van die structuur maakten geocentrische gedachtes deel uit bij Aristoteles en heliocentrische bij Kepler.

In de derde plaats zijn er waarnemingsfysiologische redenen om de inductiehypothese te verwerpen. Zo wordt een visueel beeld op de netvliezen uiteengelegd in honderdduizenden visuele veldjes die elk door één cel in de visuele cortex worden waargenomen en pas in hogere corticale delen tot beelden worden samengesteld.13 Perspectief tekenen pleit eveneens tegen inductie. Immers, daar wij perspectivisch zien, is het beeld dat op het netvlies valt gelijkvormig met een perspectivische tekening. Welnu, indien inductie ruimtelijke kennis zou verklaren, zou de mensheid al vanaf de allereerste rotstekeningen perspectivisch hebben getekend en dan zouden de allereerste kindertekeningen eveneens perspectivisch van aard moeten zijn. Beide is stellig niet het geval. Pas rond 1413 ontdekte de Italiaan Brunelleschi het perspectief tekenen, terwijl het kind zijn eerste perspectivische tekeningen pas vanaf een jaar of 9, 10 maakt. Ook hier geldt dat er eerst een bepaalde psychologische structuur moet zijn, voordat de volwassene of het kind over het vermogen beschikt om perspectivisch te tekenen.14

Om ontwikkelingspsychologische, wetenschapshistorische en waarnemingsfysiologische redenen dient inductie dus verworpen te worden. Inductie is een niet-bestaand proces: kennis komt niet uit de buitenwereld binnen. Daarmee is het empirisme als correcte beschrijving voor kennisverwerving en wetenschapsbeoefening eveneens verworpen.

Hoe komt kennis dan wel tot stand? Kennis ontstaat in een wisselwerking tussen binnenvallende prikkels enerzijds en de kenner en zijn psychologische structuren anderzijds. Over die wisselwerking valt veel meer te zeggen. Nemen we Aristoteles en Kepler die elk voor zich op een heldere dag de verplaatsing van de zon aan de hemel hebben gevolgd. Beider gedachtes daarover stonden in de contekst van een theorie ter verklaring van onder meer het ‘dwalen’ van zon, maan en planeten tussen de vaste sterren en het niet op de aarde vallen van zon, maan en planeten. Aristoteles’ leermeester Plato nam aan dat de aarde in het middelpunt van het universum zou staan en dat de zon, de maan, de vijf planeten en de vaste sterren zich elk op een sfeer zouden bevinden. Plato’s leerling Eudoxos trok Plato’s 8-sferen-theorie na en vond haar houdbaar voor de zon, de maan, de vaste sterren en wellicht ook één of meer planeten. Voor (bijna) alle planeten vond hij echter onregelmatigheden ten opzichte van Plato’s theorie. Die verklaarde hij met 18 extra sferen. Kallippos trok Eudoxos’ 26-sferen-theorie na en vond haar goeddeels houdbaar, maar omdat hij voor Mars en Venus afwijkingen vond nam hij 9 extra sferen aan, ‘om de waargenomen feiten te kunnen verklaren’. Aristoteles trok Kallippos’ 35-sferen-theorie na en vond haar goeddeels houdbaar. Om gevonden afwijkingen te verklaren oordeelde hij het ‘noodzakelijk dat er voor iedere planeet andere sferen worden toegevoegd’. Ook wilde hij – anders dan zijn louter meetkundig denkende voorgangers – de bewegingsoverdracht tussen de sferen onderling verklaren. Aldus kwam hij op een totaal van 56 sferen.

Het algemene patroon is dus dat men ten opzichte van een houdbaar geachte achtergrond waarnemingen doet die daarvan afwijken, dat men die verrassingen tracht te verklaren, dat men die verklaringspoging natrekt, enzovoort want vroeg of laat doet zich een hogere-orde-verrassing voor, die een hogere-orde-verklaringspoging vergt, die ook weer wordt nagetrokken. Al met al ontstaat er dus een cyclus die uit drie stappen bestaat: … 🡪 verrassing V 🡪 verklaringspoging 🡪 verankering 🡪 verrassing V´ 🡪 …. Dit is de onderzoekscyclus. Voor enkele toepassingen van de onderzoekscyclus bij meer gecompliceerde aangelegenheden zie noot 15.

Ook na Aristoteles blijkt de onderzoekscyclus de geschiedenis van het denken over zwaartekrachtverschijnselen beschrijvend te verklaren. Om er twee momenten uit te lichten: Copernicus opperde zijn heliocentrische theorie onder meer naar aanleiding van de afwijkingen van de seizoenen ten opzichte van de Juliaanse kalender; terwijl Kepler op contradictievrije wijze kon aannemen dat een zwaardere massa een lichtere massa sterker zou aantrekken dan omgekeerd, stelde Newton een contradictie vast vanuit de wet van actie en reactie die hij op zeker moment had, en nam hij ter verklaring van die contradictie aan dat twee massa’s elkaar met eenzelfde kracht zouden aantrekken, wat leidde tot z’n zwaartekrachtwet met (m1×m2).15

De empirie speelt in de wetenschap dus een andere rol dan het empirisme stelt. De theorie komt niet uit feiten voort maar ter verklaring van een verrassing werpt men een theorie op, die men vervolgens met nieuwe feiten op haar empirische houdbaarheid natrekt. Dat proces heet bij Aristoteles apagogè en bij Peirce abductie. Zo men wil, niet langs inductieve weg ontstaat houdbare kennis maar langs abductieve weg. Drenths voorstelling van zaken dat men ‘[in de fysica] begint met observatie en vanuit deze geobserveerde feiten tot een meer algemeen geldende theorie probeert te komen’ houdt dus geen stand. Zo is het nooit in de natuurkunde gegaan en zo zal het dus ook nooit in de psychologie lukken. Het menselijke brein verwerft geen kennis langs inductieve weg en wel langs abductieve weg.

II. De inductieve component van het positivisme heeft dus geen feitelijke grond. Gaan we daarom over op zijn getalsmatige en rekenkundige component, en wel aan de hand van vijf fragmenten uit de voorbeelden van §2.

Ad voorbeeld A. Binets en Simons bewering dat men bijna altijd ‘wetenschappelijke precisie en exactheid moet invoeren’ is niet houdbaar: slechts in het kader van een verklaringspoging zijn getalsmatige precisie en exactheid, waar ze op doelen, zinvol. De grotere exactheid die sedert Newtons zwaartekrachtwet mogelijk is, is een relevant voorbeeld: die exactheid zetten natuurkundigen niet voorop zoals in de 3-puntsschaal van de NPV, maar vloeit voort uit die wet die een houdbare verklaring is gebleken voor de contradictie in Keplers zwaartekrachttheorie. Met andere woorden, wetenschappelijke getalsmatigheid krijgt men vanaf een zeker moment cadeau als men verklarend en toetsend voortwerkt. Een ander voorbeeld daarvan is de toekenning van golflengtegetallen aan spectraalkleuren. Dat is een reële getalsmatigheid die rond 1825 vooral door Fraunhofer is bevorderd, en wel geheel in het kader van Youngs golftheorie van 1801 voor het licht als verklaringspoging voor interferentieverschijnselen. Omgekeerd, in 2006 weten we zeker dat de optica niet verder was gekomen als Newton de kleuren van de regenboog een 7-puntsschaal had gegeven.16 De getalsmatigheid van intelligentietests is wel op zulke puntsschalen gebaseerd en staat niet in een verklarend kader. De geünificeerde nomenclatuur waar Binet en Simon in 1905 op hoopten, is er dan ook in 2006 nog steeds niet.

Ad voorbeeld C. Eysencks bewering dat de elementen in zijn stelsel geen psychologische dimensies zouden moeten hebben omdat dat in de natuurkunde met ‘snelheid’ ook het geval zou zijn als men die als een fractie van de lichtsnelheid definieert, berust op een misverstand. Immers, ‘De snelheid is 0,01’ betekent in zo’n natuurkundige stelsel nog steeds ‘De snelheid is 3.000 kilometer per seconde’. In de natuurwetenschappen bestaan er geen dimensieloze feiten. Wie ze dus daarbuiten toch wil invoeren, bedient zich vooralsnog van artefacten: zo gaat kennisverwerving niet in zijn werk.17

Ad voorbeeld D. De Groot duidt Tycho Brahes bepalingen van planeetposities als ‘systematisch volgehouden waarnemingen en metingen’. Dat waren ze stellig ook, maar ze stonden eerst en vooral in het kader van Brahes poging om Copernicus’ heliocentrische theorie te verwerpen en om zijn eigen half-geo- en half-heliocentrische theorie aan te tonen. Welnu, Brahes theorie was een verklaringspoging, onder meer voor het feit dat hij de ruimte tussen Saturnus’ baan en de vaste sterren in Copernicus’ theorie veel te groot achtte. ‘Systematisch volgehouden waarnemingen en metingen’, bijvoorbeeld met tests en vragenlijsten waarvan niet alle onderdelen in de contekst van een verklarende theorie staan, is dan ook een vorm van artefacten verzamelen en geen wetenschap. Zeker, ‘werkwijzen die in de wiskunde en de natuurwetenschappen klaarblijkelijk tot grote successen hebben geleid, moeten ook de gedragswetenschappen verder kunnen brengen’, maar dan dient men wel een houdbare voorstelling van zaken te hebben over de werkwijzen van wiskundigen en natuurwetenschappers. De Groots inductionistische methodologie doet dat niet; de abductionistische onderzoekscyclus wel.18

Ad voorbeeld E. Drenth stelt kleurenblindheidsplaten en psychologische tests ten onrechte op voet van gelijkheid. Tests zijn inderdaad ‘classificatie- of meetprocedures’ die echter niet gebaseerd zijn op ‘empirisch-theoretisch gefundeerde eigenschappen’. Dat is lippendienst aan theorievorming want ook in 2006 is er geen enkele persoonlijkheidstheorie waaruit voortvloeit dat er drie wetenschappelijk relevante reacties zijn op beweringen als ‘Er zijn maar weinig mensen, die mij begrijpen’ en ‘Ik vertrouw de mensen pas als ik ze goed ken’ en dat men daar dan 1, 2 of 3 punten zou voor dienen te krijgen, alsof de reacties in enig wetenschappelijk relevant opzicht onderling 2, 3 of 1½ keer zo veel waard zouden zijn. Laat staan dat die persoonlijkheidstheorie empirisch nagetrokken en houdbaar bevonden zou zijn. Omgekeerd, kleurenblindheidsplaten zijn wel een toepassing van een empirisch houdbaar bevonden verklaringspoging, namelijk voor de opmerkelijke constatering dat niet iedereen kleuren op gelijke wijze waarneemt. De oudste vermelding daarvan is uit 1684 van de Engelse oogarts Dawbeney die over een jongedame van 22 schrijft dat ze ‘zeer goed kon zien, maar geen andere kleuren dan zwart en wit’. Verdere stappen volgens de onderzoekscyclus worden pas vanaf 1777 door de Engelse oogarts Huddart gezet. Bovendien beschouwt men in kleurenblindheidsplaten iemands reacties niet ‘in vergelijking tot die van anderen’ zoals in tests; zie de kwantitatieve schalen van §4.19

Ad voorbeeld F. Geurtz’ methodologische vergelijking van een IQ-test met een spectraalanalyse gaat niet op. Immers, iemands totaalscore op een IQ-test vloeit niet voort uit een empirisch houdbaar gebleken verklaringspoging voor een of meer intelligentieverschijnselen. Spectraalanalyses gaan echter terug op afwijkende waarnemingen van Fraunhofer: het spectrum van vlammen bleek twee heldere gele lijnen te bevatten (1813); in plaats van heldere lijnen bleek het spectrum van de zon ‘bijna ontelbaar vele sterke en zwakke verticale lijnen, die echter donkerder zijn dan de overige delen van het kleurenbeeld’ te bevatten (1814).20

Kortom, wetenschappelijke getalsmatigheid en berekeningen zijn zeer wel mogelijk, maar dan wel vanuit een houdbaar gebleken verklaringspoging (zoals in Newtons zwaartekrachttheorie, in de optica, in kleurenblindheidsplaten en in spectraalanalyses) en niet door getalsmatigheid voorop te zetten in puntsschalen.

De natuurwetenschappelijke methode is kennelijk één en de niet-houdbare positivistische lezing daarvan twee. Natuurwetenschappelijke kennis wordt niet op positivistische wijze verworven en dus wordt er, omgekeerd, in de positivistische disciplines niet aan wetenschap gedaan vanwege de vertekende kijk op kennisverwerving, vanwaaruit daarin wordt gewerkt.

**4. De statistische methode en haar onhoudbaarheid**

Om aan te tonen dat positivistische getalstoekenningen oneigenlijk zijn, volstaat bovenstaand betoog vanuit de onderzoekscyclus en het begrippenpaar inductie-abductie: in de wetenschap zijn alleen getallen die vanuit een theoretisch kader zijn toegekend, eigenlijk. Hier vloeien ten minste twee vragen uit voort. Ten eerste kan men zich afvragen hoe de positivisten tot hun getalsmatigheid komen. Vervolgens is er de vraag hoe ze met hun getallen verder rekenen. Beide vragen vinden hun antwoord in de statistische methode en haar ontstaansgeschiedenis.

Vóór we bij die twee vragen stilstaan, merken we op dat de term ‘statistiek’ in ten minste zes betekenissen en/of conteksten wordt gebruikt. In vijf daarvan gaat men abductief te werk volgens de onderzoekscyclus: de beschrijvende statistiek, zoals van het Centraal Bureau voor de Statistiek; de kansrekening; de verzekeringswiskunde; de meetfouttheorie; de stochastische natuurkunde, zoals de statistische mechanica. De zesde betekenis is die van de statistische methode (of de inductieve statistiek; omdat die term al meteen tot uitdrukking brengt waar de zwakte zit, zullen we die vermijden).21

Na een aanloop van enkele decennia komt de statistische methode goed van de grond in 1845. Dan wendt de Belgische sterrenkundige Quetelet (1796-1874) de klokkromme (Gausscurve of normale verdeling22) oneigenlijk aan. Dat wil zeggen, hij gaat er terecht van uit dat als men van één persoon de lichaamslengte vele malen bepaalt, de meetresultaten een klokkromme vormen. Dan keert hij deze houdbare gedachte van de meetfouttheorie ten onrechte om. Terwijl er bij iemands lichaamslengte werkelijk een getal bestaat, waarvan men de waarde tracht te bepalen, zou er in een collectief een gefingeerd mens bestaan, die dat collectief door zijn lichaamslengte vertegenwoordigt en ten opzichte van wie de lichaamslengtes van de andere leden meetfouten zouden zijn. De lengte van die gefingeerde mens zou het gemiddelde zijn. Quetelet trekt dit onder meer bij 100.000 Franse recruten na. Hun lichaamslengtes vormen inderdaad ongeveer een klokkromme, rond het gemiddelde van 162 centimeter. Omdat de helft binnen 162 ± 4,9 centimeter ligt, noemt hij 4,9 centimeter de ‘waarschijnlijke meetfout’. Hij besluit: ‘Men kan verscheidene reeksen waarnemingen [onder de klokkromme] met elkaar vergelijken, zelfs als deze waarnemingen betrekking zouden hebben op heterogene kwantiteiten’.23

Het overplanten van de klokkromme naar een heterogene groep is om ten minste drie redenen niet terecht. Ten eerste, in de meetfouttheorie is er een ‘ware’ waarde waar de afzonderlijke metingen van afwijken, terwijl ze herhalingen van elkaar zijn. De metingen van de lichaamslengtes van de personen Q, R, … zijn echter geen herhalingen van de meting van de lichaamslengte van persoon P wiens lichaamslengte voor het collectief zou staan. Ten tweede, bij een lichaamslengte van 162 centimeter is 4,9 centimeter een veel te grote meetfout. Ten derde, het is de vraag wat in Quetelets procedure een gemiddelde betekent. Zo is het gemiddelde van twee rechthoekige driehoeken doorgaans geen rechthoekige driehoek. Hoe zouden positivisten, die immers zonder inhoudelijke theorie werken, kunnen bepalen of hun gemiddeldes nog relevant zijn voor hun kenveld?

Kortom, in navolging van Quetelet wendt men in de statistische methode de klokkromme die in de meetfouttheorie voor een homogeen object geldt, ten onrechte aan in het meten van een groep heterogene objecten.24

De Engelse wetenschapper Galton (1822-1911) erkent dat intelligentie in tegenstelling tot lichaamsmaten geen inhoudelijke metriek heeft. Om intelligentie daar toch van te voorzien verdeelt hij de Queteletiaanse klokkromm e in 100 gelijke delen. Als men dan 100 individuen rangschikt van de domste naar de slimste, zou men een metriek hebben met de ‘waarschijnlijke meetfout’ als eenheid. Hier komen niet alleen de IQ-metriek uit voort, maar ook allerlei puntsschalen zoals de 3-puntsschalen in de NPV en de 5- en 2-puntsschalen in de empathieschaal; zie §1 respectievelijk noot 3.25

Hiermee is de eerste vraag beantwoord: de positivistische getalsmatigheid komt voort uit een onjuist begrijpen van de meetfouttheorie in het algemeen en van de klokkromme in het bijzonder. Terwijl de grootte van een meetfout slechts iets zegt over de meetnauwkeurigheid bij een bepaalde theoretische en instrumentele stand van zaken, zouden de klokkromme en de standaarddeviatie, die uit Quetelets en Galtons ‘waarschijnlijke meetfout’ is voortgekomen, een metriek opleveren. Die metriek zou naar een bepaald kenveld verwijzen en dus relevant geacht worden te zijn voor de objecten van dat kenveld. In de niet-positivistische wetenschap wordt dat laatste echter bepaald door de theorieën die als empirisch houdbaar gebleken verklaringspogingen beschikbaar zijn.

Met andere woorden, in navolging van Galton wendt men in de statistische methode de klokkromme van de abductief werkende meetfouttheorie ten onrechte aan in het inductieve meten van heterogene objecten.26

Uit Galtons kwantitatieve schalen voor kwaliteiten zijn andere gedachtengangen en formules voortgekomen. Met name de formule voor de correlatiecoëfficiënt van de Engelse wiskundige Pearson van 1896 en de door de Engelse wiskundige Yule in 1899 geïntroduceerde regressieanalyses zijn belangrijke rekeninstrumenten van het positivisme.27 Hun ongeldigheid laat ik nu zien voor regressieanalyses.

In een regressieanalyse poneert men een vergelijking, doorgaans van het type X = aA + bB + cC + … als in het derde voorbeeld van §1, en tracht men met data voor X, A, B, C, … de coëfficiënten a, b, c, … te bepalen met de methode van de kleinste kwadraten.

In de niet-positivistische wetenschappen is de methode van de kleinste kwadraten er voor het bepalen van coëfficiënten in vergelijkingen die empirisch zijn nagetrokken. Zo weet men sedert 1609 (Kepler) dat de baan van een planeet ellipsvormig is. Dan is het zinnig om de coëfficiënten a en b te bepalen van X 2 + aY 2 = b (als men het vlak van de ellips al kent). Positivisten hebben zulke kennis per definitie niet en zijn daar in het bepalen van de coëfficiënten a , b, c, … juist langs inductieve weg naar op zoek. Echter, bekeken vanuit de onderzoekscyclus is X = aA + bB + cC + … op zijn best een verklaringspoging voor een eventueel verband tussen X, A, B, C, …, dat eerst empirisch aangetoond dient te worden. Die stap slaan positivisten over – ten onrechte gezien de status van de meetfouttheorie in het algemeen en van de methode van de kleinste kwadraten in het bijzonder. Spearmans bewering in voorbeeld B van §2 dat ‘[Gauss’ formules in de meetfouttheorie] niet weinig hebben bijgedragen aan de exactheid van modern natuurkundig en psychologisch onderzoek’, getuigt dan ook van een onjuist begrip van wetenschappelijke exactheid. Zoals we hebben gezien vloeit natuurkundige exactheid voort uit een houdbaar gebleken verklaringspoging. Omdat de positivistische psychologie per definitie niet van verklaringspogingen uitgaat, is haar exactheid dus vooropgezet en oneigenlijk.

Om dezelfde reden zijn de formules voor S, R en E in §1 op zijn best verklaringspogingen waarvan de houdbaarheid nog aangetoond dient te worden en die dus vooralsnog niet houdbaar geacht mogen worden.

Hiermee is ook de tweede vraag beantwoord. Immers, formules van niet-positivistische wetenschappen, zoals F=(m1×m2)/r2 voor Newtons zwaartekrachtwet en pV=C (bij constante T) voor de wet van Boyle28, vloeien enerzijds voort uit het stelselmatig verklarend en toetsend voortwerken volgens de onderzoekscyclus en leveren anderzijds de rekeninstrumenten. Het positivisme daarentegen beschikt vanwege het verklaring- en toetsingsloze meten per definitie niet over zo’n rekeninstrumentarium. In plaats daarvan meent men zijn rekeninstrumenten te kunnen ontlenen aan de meetfouttheorie en gedachtegangen die daar sedert Quetelet ten onrechte mee worden verbonden.29

**5. Conclusie**

Ik vat de gevolgde weg samen en trek mijn conclusie.

In positivistisch onderzoek tracht men uit getalsmatige data via formules en berekeningen inzichten, wetten en theorieën te verwerven. Daarbij bedient men zich in hoge mate van de statistische methode en haar rekeninstrumentarium. Het positivisme in het algemeen en de statistische methode in het bijzonder zijn echter geen vormen van kennisverwerving. Immers, de drie stappen van de onderzoekscyclus – verrassing, verklaringspoging en verankering – vormen een houdbaar geheel, met name in de geschiedenis van de natuurwetenschappen, maar ook in die van de taalwetenschappen (zie de noten 15 , 19 en 20). De onderzoekscyclus is dus niet slechts een verklaringspoging voor kennisverwerving, maar ook een empirisch houdbare verklaring daarvoor. Daarom kan hij op het positivisme toegepast worden om te zien of het aan die drie stappen voldoet of niet.

Ten eerste, het positivisme stoelt op de inductiehypothese van het empirisme, die niet houdbaar blijkt te zijn (§3, I). Ten tweede, de positivistische getalsmatigheid en berekeningen vloeien ten onrechte niet voort uit een houdbaar gebleken verklaringspoging, maar zijn vooropgezet (§3, II). Ten derde, in de statistische methode is de meetfouttheorie ten onrechte omgekeerd, onder meer naar het toepassen van de klokkromme op gegevens die geen uitkomsten zijn van herhaalde proefnemingen, naar kwantitatieve schalen voor kwaliteiten en naar regressieanalyses (§4).

Daar komt bij dat het positivisme nog geen van zijn beloftes heeft vervuld. Volgens Binet en Simon zouden tests tot een geünificeerde nomenclatuur rond idiotie en dergelijke leiden, die vergelijkbaar zou zijn met de nomenclaturen in de exacte wetenschappen. Die eerste is er niet gekomen, terwijl de positivistische definitie van intelligentie ook na een eeuw nog op het niveau staat van ‘Intelligentie is wat deze IQ-test meet’. Datzelfde geldt voor Eysencks woorden dat ‘veelbelovende tests ontdekt worden in termen van hun correlaties met [de persoonlijkheidsdimensies neuroticisme N, psychoticisme P of extraversie E]’.30 Ook 54 jaar later is het nog steeds ‘Neuroticisme is wat deze N-schaal meet’, ‘Psychoticisme is wat deze P-schaal meet’ en ‘Extraversie is wat deze E-schaal meet’.

Tenzij positivisten alsnog óf de onhoudbaarheid van de onderzoekscyclus en van abductie en de houdbaarheid van inductie, van vooropgezette getalsmatigheid en berekeningen en van de statistische methode aantonen óf aantonen dat het positivisme wél aan de onderzoekscyclus voldoet en dus abductief van aard is, dient het positivisme niet meer aan de universiteiten beoefend te worden en dienen positivistische onderzoeksprogramma’s óf in abductieve zin omgevormd te worden (in de cognitieve psychologie van IQ-tests naar Piagetiaans intelligentieonderzoek; van persoonlijkheidsvragenlijsten naar onderzoek vanuit de zelfkennistheorie; bij het Cito van meerkeuzetoetsen naar open vragen; van empathieschalen naar psychologische niveaus van toerekeningsvatbaarheid zoals in Brands benadering) óf van de universiteiten te verdwijnen.31

**Noten**

1 Over de definitie van de term positivisme bestaat allerminst overeenstemming. Historisch gaat het woord ‘positivisme’ terug op de Fransman Comte (1798-1857) en zijn *Cours de philosophie positive* (1830-1842). Dan verwijst het naar de wetenschapsopvatting volgens welke men van feiten die onbetwijfelbaar (‘positief zeker’) zijn, tracht te komen tot objectieve kennis over de empirie. Enerzijds hebben zich daaruit in de filosofie diverse richtingen ontwikkeld, zoals het ‘logisch positivisme’ en het ‘neopositivisme’. Anderzijds, in 2006 gebruiken niet-filosofische onderzoekers – met name in de gammawetenschappen, maar ook in de alfa- en soms zelfs in de betawetenschappen – methodes en technieken die als ‘positivistisch’ worden aangeduid. De statistische methode speelt daarin een centrale rol, zoals we nog zullen zien. Welnu, Comte was er nu juist zeer op tegen zijn positivisme te verbinden met de statistische methode zoals die bij de Belg Quetelet (1796-1874) gestalte aan het krijgen was. Comte spreekt zelfs van een ‘diep irrationele aberratie’.

Ook is er het rechtspositivisme. Dat heeft al helemaal niets met methodologische positivisme van het artikel te maken.

2 P.J.D. Drenth, *Inleiding in de testtheorie*, Deventer, Van Loghum Slaterus, 1975, p.276.

P.J.D. Drenth en K. Sijtsma, *Testtheorie*, Houten, Bohn, Stafleu & Van Loghum, 1990, p.26 en p.29. De auteurs noemen ‘lengte’ kennelijk een natuurkundig begrip. Met enige goede wil kan dat, maar dan is het mijns inziens aan te bevelen van ‘fysisch-geografisch begrip’ (namelijk in het geval van ‘geografische lengte’) en/of ‘sterrenkundig begrip’ (namelijk in het geval van ‘astronomische lengte’ en ‘galactische lengte’) te spreken. Echter, zelfs dan is ‘lengte’ toch eerst en vooral een geometrisch begrip. Prof. A.T. de Hoop bracht onder mijn aandacht dat de lengte in 1983 als fundamentele natuurkundige eenheid is afgeschaft. Dat verandert uiteraard niets aan mijn bezwaren tegen Drenths en Sijtsma’ voorstelling van zaken.

3 J.B. Dijkstra en T. van Erven, ‘De empathie van forensisch psychiatrische patiënten’, *Stator*, 2002, vol.5, p.23-28. In die vergelijking staan V1,2,3,4,5,6 voor de varianties in 5-puntsschalen voor ‘inleven in de belevingswereld van een ander’ (V1), ‘je kunnen verontschuldigen’ (V2), ‘rechtvaardige afweging van belangen’ (V3), ‘interesse hebben voor de ander’ (V4), ‘erkenning van het slachtoffer zijn van de ander’ (V5) respectievelijk ‘sympathie opbrengen voor de noden en behoeften van de ander’ (V6), terwijl V3A = (bij gunstige score op V3) variantie in een 2-puntsschaal voor ‘geloofwaardig versus alleen uit op eigen gewin’, V4B = (bij ongunstige score op V4) variantie in een 2-puntsschaal voor ‘is wel of niet in staat tot interesse hebben voor de ander’, V5B (bij ongunstige score op V5) variantie in een 2-puntsschaal voor ‘is wel of niet in staat tot erkenning van het slachtoffer zijn van de ander) en V6C = (bij gunstige score op V6) variantie in een 2-puntsschaal voor ‘brengt uit zichzelf sympathie op voor de noden en behoeften van de ander versus brengt die alleen op met sturing door een ander’. De term ‘-6’ brengt de waarde 0 in het centrum van de verdeling voor de controlegroep.

Hier volgt nog een voorbeeld van een econometrische regressievergelijking. Op 26 maart 1998 publiceerde het Centraal Planbureau (CPB) in *Keuzes in kaart* (Den Haag, CPB, 1998) de resultaten van zijn doorrekeningen van de programma’s van vijf politieke partijen, met het oog op de Kamerverkiezingen van 6 mei 1998. Die door rekeningen zijn uitgevoerd met drie econometrische modellen, namelijk het MIMIC-, het ATHENA- en het JADE-model. Formule 23 uit het JADE-model, de consumptievergelijking, bijvoorbeeld luidt: (c - c -1)/c -1 = A∆ 2lnw + B∆r l -1 + C{∆(Y L/p c)}/c -1 + D{∆(Y 0/p c)}/c -1 + E(∆Y Z/p c)/c -1 + F(∆A F/p c)/c -1 + G{ln(c -1/c \* -1)}, met c = reële consumptie in een bepaald jaar, c -1 = idem in het jaar ervoor, w = loonvoet, r l -1 = lange rente in het jaar ervoor, Y L = inkomsten uit loon, p c = prijs van de reële consumptie, Y 0 = inkomsten uit overdracht van vermogen, Y Z = inkomsten uit winst en A F = huizenvoorraad. Met de substituties (c - c -1)/c -1 = Z, Δ 2lnw = P, Δr l -1 = Q, {Δ(Y L/p c)}/c -1 = R, {Δ(Y 0/p c)}/c -1 = S, (ΔY Z/p c)/c -1 = T, (ΔA F/p c)/c -1 = U en {ln(c -1/c \* -1)} = V gaat de vergelijking over in Z = AP + BQ + CR + DS + ET + FU + GV. Dit is een regressievergelijking waarvan het zaak is de coëfficiënten A, B, …, G te bepalen. Het CPB vond: A = -0,253 ± 0,08, B = -0,458 ± 0,35, C = 0,881 ± 0,13, D = 0,871 ± 0,20, E = 0,225 ± 0,11, F = 0,071 ± 0,03 en G = -0,462 ± 0,16. De relatieve onzekerheidsmarges zijn dus 31,6% (A), 76,4% (B), 14,8% (C), 23,0% (D), 48,9% (E), 42,3% (F) respectievelijk 34,6% (G).

Deze econometrische regressievergelijking is om ten minste twee redenen minder ongunstig dan het psychologische voorbeeld in de hoofdtekst. 1. De gegevens waarmee men econometrische regressieanalyses uitvoert, zijn doorgaans feitelijk van aard; zie de grootheden als reële consumptie, loonvoet enzovoort. 2. Bij Keynes, Modigliani en anderen zitten achter de algebraïsche verbanden allerlei inhoudelijke noties. Het is echter zeer de vraag of de overgang van noties omtrent stijgingen en dalingen van factoren in afhankelijkheid van andere factoren voldoende aanleiding geven tot bedoelde algebraïseringen.

Zelfs als we aannemen dat de consumptievergelijking een reële (in plaats van overgealgebraïseerde) verklaringspoging is, dan nog lijkt die verklaringspoging alleen al om de grootte van de onzekerheidsmarges verworpen te moeten worden, Immers, A, B, …, G worden in de theorie als constanten aangemerkt. Bij relatieve onzekerheidsmarges tussen 14,8% (voor C) en 76,4% (voor B) is dat echter zeer problematisch.

4 Acht andere voorbeelden zijn:

Voorbeeld 1. ‘De wetenschappelijke methode om feiten te onderzoeken [...] is toepasbaar zowel op sociale als op fysische problemen’ en bestaat in ‘de ordelijke classificatie van feiten, gevolgd door de herkenning van hun relaties en terugkerende sekwenties’. Deze kennisoptiek zou ook voor de psychologie gelden en als exacte voorbeelden waarin ze al in tot uitdrukking zou zijn gekomen merkt Pearson aan: ‘De waarnemingen van Newton aan de relatie tussen de bewegingen van een vallende steen en de maan [...], van Kirchhoff aan bepaalde lijnen die voorkomen in het spectrum van het zonlicht; K. Pearson, *The grammar of science*, Londen, Black, 1900 (tweede druk), p.6, p.18v, p.19v en p.29.

Voorbeeld 2. ‘Het ideaal van de wetenschappelijke methode moet onzes inziens zijn [...]: een overdenking van de feiten die uit de eerste hand zijn verzameld’; een test zou te vergelijken zijn met een microscoop; A. Binet en Th. Simon: ‘Le développement de l’intelligence chez les enfants’, *L’année psychologique*, 1908, vol.14, p.1-94, met name p.1 en p.60v.

Voorbeeld 3. De natuurkunde en de psychologie zouden vier meetschalen hebben: nominale, ordinale, interval- en verhoudingsschalen. Net als de schaal voor hardheid van metalen zouden IQ-tests een ordinale schaal vormen, al zou soms met succes het intervalnivo benaderd worden, waar onze Celsius- en Fahrenheitthermometerschalen en onze kalender op staan; S.S. Stevens, ‘On the theory of scales of measurement’, *Science*, 1946, vol.103, p.677-680; ‘Mathematics, measurement, and psychophysics’, *Handbook of experimental psychology* (red. Stevens), New York, Wiley, 1951, p.21-30.

Voorbeeld 4. ‘Wetenschappelijke vooruitgang [...] hangt af van de introductie van meting in het onderzoeksveld’; zo zou Newtons zwaartekrachttheorie alleen mogelijk zou zijn geweest als gevolg van Keplers metingen in de sterrenkunde; R. Cattell, *Personality and motivation, structure and measurement*, New York, World book, 1957, p.4; Kepler heeft vanwege zijn slechte ogen nauwelijks planeetstanden opgemeten. Hoogstwaarschijnlijk doelt Cattell op Tycho Brahe’s metingen.

Voorbeeld 5. De Amerikaanse psycholoog Guilford vergelijkt zijn systeem voor ‘psychomotorische vaardigheden, temperamentsfactoren en psychopathologische factoren’ met ‘het scheikundige periodieke systeem, geïntroduceerd door Mendeleev’; J.P. Guilford, *The nature of human intelligence*, New York, McGraw-Hill, 1967, p.47.

Voorbeeld 6. De Amerikaanse psycholoog Cronbach vergelijkt de testpsychologie met Shannons stochastische communicatietheorie: ‘Het signaal (ware score) is de informatie die we graag zouden hebben – en zouden kunnen verkrijgen als er perfecte ontvangst was. Maar in plaats daarvan krijgen we een signaal dat is verstoord door interferentie of ruis. De “sneeuw” op het televisiescherm [...], de afwijking van de ware score die het gevolg is van het testen van een leerling op een bijzonder goed of slecht ogenblik – ze vormen allemaal “ruis”. Hoe groter de ruis, hoe minder zeker we kunnen zijn van het onderliggende signaal’ en poneert de formule: ‘signaal : signaal-plus-ruis = variantie van ware scores : variantie van gemeten scores’; L.J. Cronbach, *Essentials of psychological testing*, New York, Harper & Row, 1970, p.156.

Voorbeeld 7. ‘Heeft de psycholoog eenmaal zijn data leren onderscheiden, dan is er in de verdere gang van zaken geen principieel verschil met de natuurkunde. Evenals de fysicus gebruikt de psycholoog instrumenten en kwantificeert hij. […] Meten is een gestandaardiseerde en gesystematiseerde versie van een normaal menselijke activiteit. Als de psycholoog zijn meetlat – de test, de vragenlijst – gebruikt, doet hij eigenlijk niets anders. Om dat duidelijk te maken gaan wij twee begrippen met elkaar vergelijken: het begrip oplosbaarheid in de natuurkunde en het begrip defensiviteit in de psychologie [...] zijn op een niet principieel verschillende wijze te meten’; C. Sanders, L.K.A. Eisenga en J.F.H. van Rappard, *Inleiding in de grondslagen van de psychologie*, Deventer, Van Loghum Slaterus, 1976, p.40.

Voorbeeld 8. Drenth en Sijtsma: ‘[Bij validiteit] betreft het een gedachtengang van algemeen wetenschappelijke aard’, *Testtheorie* (op.cit), p.174.

5 A. Binet en Th. Simon, ‘Sur la nécessité d’établir un diagnostic scientifique des états inférieurs de l’intelligence’, *L’année psychologique*, 1905, vol.11, p.163-190, met name p.164 en p.168.

6 Ch. Spearman, ‘The method of “right and wrong cases” (“constant stimuli”) without Gauss’s formulae’, *British journal of psychology*, 1906-1908, vol.2, p.227-242, met name p.227 en p.228.

7 H.J. Eysenck, *The scientific study of personality*, Londen, Routledge & Kegan Paul, 1952, p.25-43.

8 A.D. de Groot, *Methodologie*, Den Haag, Mouton, 1961, p.48 en p.51. Het tweede citaat wordt vervolgd met: ‘Reeds door Fechner, en later o.a. door Thurstone en Stevens, werd bijvoorbeeld in de argumentatie ten gunste van de ontwikkeling van metrische schalen in de psychologie [...] met zoveel woorden naar de natuurwetenschappen en haar resultaten verwezen [...]. Hetzelfde geldt voor het gebruik van de strikt “hypothetico-deductieve” methode van theorieontwikkeling’.

Andere voorbeelden in *Methodologie* zijn: ‘Aan het proces van verscherping van (de grenzen van) een begrip zijn gewoonlijk twee aspecten te onderscheiden: de uitwerking van het nomologische net van het begrip, in theoretische relaties en deducties, empirische specificaties, onderzoek-bevindingen [...] en het gaandeweg afslijten van de surplus-betekenis. Fraaie voorbeelden van het verloop van dit proces zijn in de natuurkundge te vinden, in de geschiedenis van begrippen als “kracht”, “energie”, maar ook: “atoom”, “molecuul”, “lichtgolven”, en dergelijke’ (p.121) en ‘De onderscheiding tussen distortie en ruis uit de informatietheorie correspondeeert met die tussen systematische en toevallige fouten in de statistiek’ (p.282).

9 *Inleiding in de testtheorie* (op.cit), p.92.

10 J. Geurtz, ‘”Onwetenschappelijke” theorieën in psychologie blijken bruikbaar’, *de Volkskrant*, 4 april 1987.

11 J. Piaget, *La construction du réel chez l’enfant*, Neuchâtel, Delachaux & Niestlé, 1937, p.9-85, met name p.42-60.

12 E. Vervaet, ‘Het ontstaan van het zelfgevoel – VI’, *Struktuur en genese*, 2000, vol.13, p.4-58, met name p.31-34; E. Vervaet, ‘Het ontstaan van het zelfgevoel – VIII’, *Struktuur en genese*, 2003, vol.16, p.4-53, met name p.21v; E. Vervaet, *Groeienderwijs*, Amsterdam, Ambo, 2002 (vierde druk 2006), p.144-146 en p.159-161.

13 Zie onder meer D.H. Hubel en T.N. Wiesel, 'Receptive fields of single neurons in the cat’s striate cortex’, *Journal of physiology*, 1959, vol.48, p.574-591; D.H. Hubel en T.N. Wiesel , ‘Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex’, *Journal of physiology*, 1962, vol.160, p.106-154; D.H. Hubel en T.N. Wiesel, ‘Receptive fields and functional architecture of monkey striate cortex’, *Journal of physiology*, 1986, vol.195, p.215-243.

E. Vervaet, *Geheugenmodellen en leeralgorithmen* (scriptie), Amsterdam, Vrije Universiteit, 1976.

14 S.Y. Edgerton, *The Renaissance rediscovery of linear perspective*, New York , Basic books, 1975. Edgerton spreekt van ‘herontdekking’ door Brunelleschi. Daarmee geeft hij mijns inziens te veel eer aan de Romeinen die perspectivisch tekenen al eerder gevonden zouden hebben.

J. Piaget en B. Inhelder, *La représentation de l’espace chez l’enfant*, Parijs, PUF, 1948, hoofdstuk 6.

15 E. Vervaet, *Strukturalistische verkenningen in kennisleer en persoonlijkheidsleer* (proefschrift Universiteit van Amsterdam), Amsterdam, Vervaet, p.13-124, met name p.33-38 (onderzoekscyclus) en p.103-124 (geschiedenis van het denken over zwaartekrachtverschijnselen); E. Vervaet, ‘ Newton’s construction of the law of gravitation’, in *Newton's scientific and philosophical legacy* (red. P.B. Scheurer & G. Debrock), *International Archives of the history of ideas* (123), Dordrecht , Kluwer, 1988, p. 281-288.

Publicaties met wetenschapsgeschiedenissen die de onderzoekscyclus ondersteunen en niet in het vervolg van het artikel worden aangehaald zijn (samenvattingen zijn te vinden via www.stichtinghistos.nl/struktuur.htm): E. Brand, ‘De ontdekking van aardolie’, *Struktuur en genese*, 1989, vol.2, p.25-28; over de stoommachine: E. Vervaet, ‘ Theorie en praktijk (2)’, *Maandblad Geestelijke volksgezondheid*, september 1989, p.970-974; E. Vervaet, ‘AIDS-mysterie’, *Struktuur en genese*, 1990, vol.3, p.7; E. Vervaet, ‘De ontdekking van Amerika – I’, *Struktuur en genese*, 1992, vol.5, p.24-28; J. Stelleman en E. Vervaet, ‘De ontdekking van de kernramp bij Tsjernobyl’, *Struktuur en genese*, 1996, vol.9, p.28-33; E. Vervaet, ‘Taalvoorbeelden van de onderzoekscyclus’, *Struktuur en genese*, 1996, vol.9, p.34-38 en 1997, vol.10, p.4-30.

Een historisch voorbeeld van de onderzoekscyclus is het ontstaan van de hoofsheid van de twaalfde eeuw. Gezien het feit dat er bijvoorbeeld in de tiende eeuw in het geheel geen hoofsheid was, is dat epistemologisch gezien een ‘verrassing’ die om een verklaring vraagt. Een verklaringspoging is dat ze in de elfde eeuw is ontstaan in het klerikale milieu rond de Duitse keizer en diens Rijnlandse keurvorsten (Jaegers vermoeden) en een andere verklaringspoging dat ze in de twaalfde eeuw is ontstaan onder de Zuid-Franse aristocratie (Duby’s vermoeden); vóór beide verklaringspogingen zijn er feiten (‘verankerd’), maar beide kennen ook hun verrassingen (‘hogere-orde-verrassingen’) die dus nog om een hogere-orde-verklaringspoging vragen; F. van Oostrom, *Stemmen op schrift*, Amsterdam, Bert Bakker, 2006, p.134vv.

Men lette op het verschijnsel van het wijken van het verklaren. Bijvoorbeeld, Newtons zwaartekrachttheorie moge dan in 1687 een empirisch houdbare verklaringspoging geweest zijn voor de vrije val, de getijden, het dwalen van zon, maan en planeten tussen de vaste sterren, en dergelijke, maar daarmee was het nog een open vraag waarom twee massa’s elkaar zouden aantrekken, en wel volgens hun product en omgekeerd evenredig met het kwadraat van de afstand tussen hun zwaartepunten.

Watts stoommachine van 1765 wordt wel eens als voorbeeld genoemd van iets dat zonder theorie tot stand gekomen zou zijn. Carnots theoretische verklaring ervoor stamt immers uit 1824. De geschiedenis van de stoommachine wordt echter geheel volgens de onderzoekscyclus beschreven. De oerverrassing is het dansen van een deksel op een kokende pan water. De eerste die daar een houdbare verklaring voor heeft, is Hero van Alexandrië (ongeveer 100 v.Chr.). Vervolgen op Hero worden gegeven door onder meer Torricelli, Papin en Newcomen, de uitvinder van de eerste atmosferische stoommachine in 1712. Ze worden echter allemaal geleid door bepaalde inzichten en theorieën. Het verschil met Carnot is dan ook niet dat deze wel en zij geen theorieën gehad zouden hebben, maar dat zij vanuit concrete handelingen, waarnemingen en gedachtes met stoom en stoommachines opereerden terwijl Carnot dat in zijn theorie van 1824 formeel deed; ‘Theorie en praktijk (2)’ (op.cit).

De onderzoekscyclus moet als verklaringspoging voor de totstandkoming van geldige en betrouwbare kennis ook voor zichzelf gelden. Dat blijkt het geval te zijn; zie R. Kooijman en E. Vervaet, ‘Is de onderzoekscyclus zelfgeldig?’, *Struktuur en genese*, 1997, vol.10, p.32-42; E. Vervaet en R. Kooijman, ‘Deze zin is onjuist’, *Kunst en wetenschap*, herfst 1998, 15v.

16 Newton geeft de kleuren van de regelboog een getal, van 1 voor rood tot en met 7 voor violet; I. Newton, *Opticks*, Londen, Smith&Walford, 1704, propositie VI, probleem II. Die getallen hebben echter geen andere betekenis dan een plaatsnummer. Met dank aan prof.dr. R.A. Crone die me op deze passage bij Newton attendeerde.

17 *Strukturalistische verkenningen* (op.cit), p.144-148.

18 *Strukturalistische verkenningen* (op.cit), p.110; ‘ Newton’s construction’ (op.cit).

19 E. Vervaet, ‘De eerste kleurenblindheidsproeven’, *Tijdschrift voor de geschiedenis der geneeskunde, natuurwetenschappen, wiskunde en techniek*, 1991, vol.14, p.74-95; E. Vervaet, ‘Pseudo-isochromatische kleurenblindheidsproeven’, *Struktuur en genese*, 1990, vol.3, p.30-39 en 1991, vol.4, p.18-34; E. Vervaet, ‘Zicht op kleur’, *Natuur en techniek*, 1990, vol.58, p.636-647.

20 E. Vervaet, ‘De geboorte van de spectraalanalyse’, *Intermediair*, 23 oktober 1987, p.43, p.45 en p.47.

21 E. Vervaet, ‘Statistiek en de statistieken’, *Struktuur en genese*, 2004, vol.17, p.26-54; E. Vervaet, ‘Statistisch supplement – I’, *Struktuur en genese*, 2005, vol.18, p.7-24.

22 De term ‘Gausscurve’ is onjuist omdat Gauss de klokkromme niet heeft gevonden, maar slechts in de meetfouttheorie heeft geïntroduceerd. De term ‘normale verdeling’ is terminologisch twijfelachtig omdat er niets normaals is aan de klokkromme en niets abnormaals aan welke andere verdeling dan ook en leunt inhoudelijk te zeer tegen de statistische methode die nu juist ter discussie staat. Zie verder ‘Statistiek en de statistieken’ (op.cit), A.IV en B.IV (De Moivre en de klokkromme) en noot 34.

23 A. Quetelet, ‘Sur l’appréciation des documents statistiques, et en particulier sur l’appréciation des moyennes’, *Bulletin de la Commission Centrale de Statistique*, 1845, vol.2, p.205-286, met name p.250-272; A. Quetelet, *Lettres à S.A.R. le duc régnant de Saxe-Cobourg et Gotha, sur la théorie des probabilités*, Brussel, Hayez, 1846, p.124-148 (brieven 19-21).

‘Statistiek en de statistieken’ (op.cit), p.36.

24 ‘Statistiek en de statistieken’ (op.cit), p.44v.

25 F. Galton, 'On a proposed statistical scale’, *Nature*, 1874, vol.9, p.348v; F. Galton, ‘Statistics by intercomparison, with remarks on the law of frequency of error’, *Philosophical magazine*, vierde serie, 1875, vol.49, p.33-46.

‘Statistiek en de statistieken’ (op.cit), p.36v.

26 ‘Statistiek en de statistieken’ (op.cit), p.45v.

27 K. Pearson, ‘Mathematical contributions to the theory of evolution – III. Regression, heredity and panmixia’, *Philosophical transactions*, 1896, p.253-318, met name p.261-268.

G.U. Yule, ‘An investigation into the causes of changes in pauperism in England , chiefly during the last two intercensal decades, I’, *Journal of the Royal Statistical Society*, 1899, vol.62, p.249-286.

‘Statistiek en de statistieken’ (op.cit), p.37.

Allerlei formules van de statistische methode wortelen in Darwins pangenetische en Galtons en Pearsons voorouderlijke erfelijkheidsleer, twee erfelijkheidstheorieën die verworpen dienen te worden in het licht van Mendels in 1900 herontdekte erfelijkheidsleer. Beide verworpen erfelijkheidstheorieën bevatten echter vele elementen die in de hedendaagse statistische methode doorwerken. Met name vervallen formules die Galton, Pearson en anderen hebben afgeleid op basis van Gal­tons uitgangspositie in Darwins pangenetische theorie. Galton redeneert dat Darwins pangenetische theorie voorlopig is, maar dat we haar, houdbaar of niet, dienstbaar kunnen maken voor erfelijkheidsonderzoek, onder meer door er wiskundige formules mee af te leiden: ‘Als de pangenetische theorie klopt [...] brengt ze alle invloeden die op erfelijkheid betrekking hebben, in een vorm die geschikt is voor de macht van wiskundige analyse’; F. Galton, *Hereditary genius*, Londen, MacMillan, 1869; tweede druk van 1892, p.358. Darwins pangenetische theorie klopt echter niet zodat eruit voortvloeiende formules in verder onderzoek niet meer zinvol aangewend kunnen worden. Dat geldt ook voor formules die uit de voorouderlijke erfelijkheidsleer voortkomen.

28 ‘Statistisch supplement – I’ (op.cit), paragraaf D.VI.

29 Achteraf bekeken is de positivistische denkwijze in zekere zin begrijpelijk aangezien de meetfouttheorie net als het positivisme en de statistische methode geen vakinhoudelijke basis hebben. Dat maakt het positivisme en de statistische methode echter nog niet tot vormen van wetenschapsbeoefening. Daarvan zijn het trachten te verklaren van verrassingen in het kennisbestand dat men op goede gronden houdbaar acht, en het empirische natrekken van verklaringspogingen de twee hoofdkenmerken. Beide ontbreken, zowel in het positivisme als in de statistische methode.

30 *Scientific study* (op.cit), p.298.

31 Voor de zelfkennistheorie zie *Strukturalistische verkenningen* (op.cit), p.183-336.

E.J.P. Brand, *Het persoonlijkheidsonderzoek in het strafrecht*, Gouda, Quint, 2001. In de economische wetenschappen kunnen econometrische berekeningen (als in noot 2) worden omgevormd naar berekeningen over reële economische grootheden.