



*Turtles All the Way Down? Psychometric Approaches to the Reduction Problem*

R.A. Kievit

## Nederlandse samenvatting

---

*Een beroemde wetenschapper (sommigen zeggen dat het Bertrand Russell was) gaf eens een lezing over sterrenkunde. Hij beschreef hoe de aarde om de zon draait, en hoe de zon, op haar beurt, rond het centrum van onze melkweg draait. Aan het eind van de lezing stond een kleine oude dame achterin de zaal op en zei: "Wat je ons hebt verteld is complete onzin. De wereld is plat, en balanceert op de rug van een reuzenschildpad". De wetenschapper glimlachte een tikkeltje arrogant en antwoordde: "En waar balanceert de schildpad dan op?" De oude dame sprak "Je bent een slimme jongeman, maar het is schildpad op schildpad, helemaal tot aan de bodem!" (Hawking, 1988, blz. 1)<sup>30</sup>*

U, de lezer, zit ergens in een stoel, en leest dit proefschrift uit nieuwsgierigheid, beleefdheid, of wellicht zelfs verveling. U bent een psychologische entiteit: met behulp van gestandaardiseerde tests en vragenlijsten kunnen psychologen u langs een verscheidenheid aan dimensies en schalen kwantificeren en beschrijven. Cognitieve wetenschappers kunnen een schatting maken van het aantal cijfers dat u in uw werkgeheugen kunt vasthouden (Unsworth & Engle, 2007), de snelheid waarmee u mentaal driedimensionale objecten kunt roteren (Borst, Kievit, Thompson, & Kosslyn 2011), in hoeverre uw respons verlangzaamt na het maken van een fout (Dutilh et al., 2012) en een schatting maken van uw algemene intelligentie, waarmee uw levensverwachting, verwacht inkomen en werkprestaties over tientallen jaren te voorspellen is (bij. Gottfredson & Deary, 2004). Door veranderingen in uw gemoedstoestand door de tijd te bestuderen is het mogelijk om uw emotionele welzijn te beschrijven en kwantificeren als een dynamisch netwerk van interacterende symptomen, en met behulp daarvan te voorspellen welke constellatie psychische klachten de kans verhogen, of verlagen, dat u in een depressie raakt (Bringmann et al., 2013). Kortom, u bent een psychologische entiteit, die op diverse manieren kan worden beschreven, gemeten en gekwantificeerd, en met zowel andere individuen als uzelf door de tijd heen vergeleken kan worden. Samen vormen deze beschrijvingen en voorspellingen een onvolmaakte, doch voorspellende en informatieve, beschrijving van u als persoon.

Tegelijkertijd bent u een biologisch wezen. U heeft ongeveer 1100 gram hersenweefsel (Allen, Damasio, & Grabowski, 2002), dat bestaat uit 86 miljard neuronen (Azevedo et al., 2009), verbonden door maar liefst 3 kilometer van axonen per kubieke millimeter (Braitenberg & Schüz, 1998) die zo energiek elektrische impulsen en neurotransmitters uitwisselen dat uw brein, hoewel het slechts 2 % van uw lichaamsgewicht opmaakt, maar liefst 25 % van alle energie in uw lichaam verbruikt (Mink, Blumenschine, & Adams, 1981). Met behulp van scanners kunnen we meten hoeveel grijze stof zich in diverse plekken in uw hersenen bevindt, in kaart brengen hoe witte stofbanen verschillende hersengebieden met elkaar verbinden, bestuderen hoe hersengebieden hun activiteit onderling synchroniseren, en gelokaliseerde veranderingen in neurotransmitterniveaus meten. Kortom, u bent ook een biologische wezen, gemaakt van miljarden continu interacterende cellen.

Een van de grootste wetenschappelijke uitdagingen van onze tijd is dat beide beschrijvingen in zekere zin waar zijn, maar op hetzelfde moment in vele

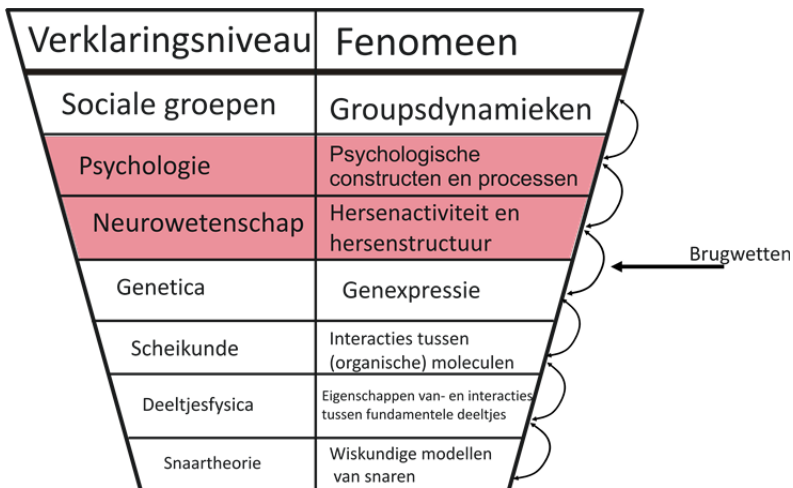
---

<sup>30</sup> In deze Nederlandse samenvatting vertaal ik de engelse citaten. Voor de originele citaten, zie de voorgaande hoofdstukken.

opzichten onverenigbaar verschillend lijken te zijn. Dit roept belangrijke vragen op: 'bestaan' psychologische processen, toestanden en eigenschappen op eenzelfde manier op lagere (biologische) verklaringsniveaus, en zouden we deze eigenschappen dus met behulp van hersenscanners kunnen meten? Of zijn de biologische en de psychologische verklaringsniveaus fundamenteel verschillend, en behoeven ze ieder een eigen set ontologische en epistemologische gereedschappen? De vraag hoe deze verschillende verklaringsniveaus samenhangen staat bekend als *het reductieprobleem*, en is de kernvraag van dit proefschrift.

## Het reductieprobleem

Het reductieprobleem behelst de vraag hoe verschillende verklaringsniveaus in de wetenschap aan elkaar gerelateerd zijn. Een invloedrijke kijk op deze uitdaging werd geschetst door Oppenheim en Putnam (1958), gebaseerd op hun hypothese van 'de eenheid van de wetenschap'. Vanuit dit perspectief, te zien in Figuur 9.1, bestaat de wetenschap uit een hiërarchie van verklaringsniveaus die variëren van 'grote' of 'hoge' (bijv. sociale groepen) tot 'kleine' of 'lage' (bijv. atomen of subatomaire deeltjes) verklaringen. Constructen of eenheden op een hoger niveau bestaan uit ('zijn gemaakt van') elementen op het lagere verklaringsniveau. Uiteindelijk bestaat de wetenschap dus slechts uit basale, fysieke bouwstenen die op complexe manieren samenkomen en zo de hiërarchie van wetenschappen vormen. Dit biedt een vereenvoudigde kijk op de (natuur)wetenschappen, zodanig dat de wetenschap een geheel vormt dat je kunt beschrijven van het hoogste niveau van abstractie tot helemaal 'tot aan beneden', naar het niveau van de fundamentele fysieke bouwstenen van het universum. De sociologie onderzoekt hoe mensen in groepen met elkaar interacteren, psychologen bestuderen het mentale functioneren van mensen, neurowetenschappers bestuderen de hersenen, de hersenen bestaan uit complexe biochemische bouwstenen, etcetera, tot aan de fundamentele bouwstenen van (sub)atomaire fysica. De vraag blijft echter hoe deze verklaringen aan elkaar kunnen, of moeten, worden gerelateerd.



**Figuur 9.1.** De eenheid van de wetenschap volgens Oppenheim en Putnam (1958). Nagel stelde brugwetten voor, die regelmatigheiden en wetten in de verschijnselen zouden vertalen naar lagere niveaus. Dit proefschrift richt zich op de relatie tussen psychologie en biologie, in rood gemarkeerd.

Oppenheim en Putnam poneerden twee manieren om na te denken over de eenheid van de wetenschap. De eerste is een geïdealiseerd, meta-theoretisch perspectief op de voorspelde uiteindelijke stand van zaken in een (hypothetisch) afgeronde wetenschap. De suggestie is dat uiteindelijk alle wetenschappelijke kennis binnen hetzelfde kader verenigd kan worden, zodanig dat elk beschrijfbaar patroon in de natuur vertaald kan worden naar wetten op lagere niveaus. Het tweede perspectief op de eenheid van de wetenschap is een beschrijvende, namelijk dat de eenheid van de wetenschap 'bestaat als een trend binnen wetenschappelijk onderzoek, ongeacht of unitaire wetenschap ooit bereikt wordt' (Oppenheim & Putnam, blz. 3). Dat wil zeggen, ongeacht of de uiteindelijke 'vereniging' van alle wetenschappen zal slagen, de trend binnen de wetenschap is om onderzoek te doen *alsof* die vereniging in principe mogelijk is. Oppenheim en Putnam formuleerden een definitie van de formele vereisten waaraan voldaan moet worden om te kunnen spreken van een reductie tussen twee verklaringsniveaus. Ze stellen (1958, blz. 5) dat theorieën op hogere verklaringsniveaus ('T2', bijvoorbeeld een psychologische bevinding) als succesvol tot een lager verklaringsniveau T1 gereduceerd kan worden beschouwd (bijvoorbeeld een neurowetenschappelijke bevinding) dan en slechts dan als:

- 1) De woordenschat van T2 termen bevat die niet in het vocabulaire van T1 voorkomen
- 2) Observaties die te verklaren zijn door middel van T2 zijn ook door T1 verklaarbaar zijn
- 3) T1 minstens zo goed gesystematiseerd is als T2

De uiteindelijke specificatie van deze voorwaarden door Oppenheim en Putnam liet veel te wensen over. De precieze betekenis van wat als een 'verklaring' gold werd bijvoorbeeld nergens helder gespecificeerd. Echter, hun voorstel vormde een inspiratiebron voor meer ontwikkelde reductionistische modellen, zoals uiteengezet door bijvoorbeeld Ernest Nagel (1961). Hij stelde een formalisering voor van de eisen waaraan moet worden voldaan om te kunnen zeggen dat een theorie op een bepaald niveau tot een lager niveau in de hiërarchie gereduceerd is. Nagel stelde voor dat er binnen elk verklaringsniveau regelmatigheden of wetten te vinden die de interactie tussen entiteiten *op dat verklaringsniveau* beschrijven, zoals bijvoorbeeld interacterende hersengebieden. Echter, als het zo is dat de entiteiten op een bepaald niveau (bijv. hersengebieden) bestaan uit onderdelen van een lager niveau (bijv. biochemische moleculen), dan moet het mogelijk zijn om wetenschappelijke uitspraken op het hogere niveau (interactie van hersengebieden) door middel van wetmatigheden *te vertalen* naar regelmatigheden op een lager niveau (eigenschappen van biochemische moleculen). Nagel noemde zulke wetten *brugwetten* (bridge laws), omdat ze een brug slaan tussen verklaringsniveaus. Uiteindelijk zou een volledige specificatie van deze brugwetten ons toestaan om alle ontologische claims, behalve het meest basale fysieke niveau, de deur uit te doen: De brugwetten zouden ons in staat stellen om bevindingen in welke wetenschappelijke laag dan ook 'omhoog' of 'omlaag' te vertalen. Hoewel er weinig of geen brugwetten zijn ontwikkeld is dit reductionistische uitgangspunt in het wetenschappelijk discours alomtegenwoordig. Zo is het niet ongevoel om de volgende redenering in papers aan te treffen: 'psychologische processen gebeuren in de hersenen, *dus* moeten we in staat zijn psychologische theorieën te vertalen naar neurale theorieën'. Echter, ondanks de aanvankelijke populariteit werd ook deze formele benaderingen al snel bekritiseerd.

De meest invloedrijke kritiek kwam in een serie papers door onder andere Putnam (bijv. 1967) en later Fodor (1974). Het argument dat zij ontwikkelden tegen reductionisme staat bekend als *meervoudige realiseerbaarheid*. Deze positie stelt dat een bepaalde mentale toestand zoals bijvoorbeeld 'pijn ervaren' door diverse fysieke toestanden gerealiseerd kan worden. Eén zo'n fysieke toestand die zou corresponderen met 'pijn ervaren' is bijvoorbeeld het vuren van een specifieke set neuronen, te weten 'C-fibers' (zie Hardcastle, 1997; Place, 1956; Smart, 1959; Rysewyk, 2013). Echter, zo beargumenteren Putnam en Fodor, er is a priori geen reden waarom alleen C-fibers dat doel zouden kunnen dienen: Andere typen zenuwcellen (bijvoorbeeld in andere dieren, zoals reptielen) of zelfs niet-biologische substraten (bijvoorbeeld silicium) kunnen eveneens deze functie hebben. Recente succesvolle hersenstamtransplantaten (gemaakt van silicium in plaats van zenuwweefsel, zie Hagan & Wilson, 2013) hebben er bijvoorbeeld voor gezorgd dat doof geboren kinderen kunnen horen. Een strikte reductionist in de Nageliaanse traditie zou in zo'n geval ofwel moeten concluderen dat wat deze patiënten doen niet 'echt' horen is (omdat de brugwet geen melding maakte van silicium), of dat de brugwetten op de een of andere manier moeten voorzien in het feit dat vooralsnog onbekende fysieke realisaties eenzelfde toestand zouden kunnen doen realiseren (omdat gehoorimplantaten in de toekomst wellicht van weer andere materialen worden gemaakt). Dit werd gezien als een verwoestende kritiek op het idee van de brugwet als formalisatie van reductie: als een mentale toestand door een groot aantal, mogelijk oneindig, stel fysieke realisaties kan worden gerealiseerd, dan is het onduidelijk hoe brugwetten deze meervoudige realisatie van psychologische fenomenen op fysieke realisaties zouden kunnen vangen.

Fodor beargumenteerde dat om deze en andere redenen fysicalisme (de opvatting dat het heelal uiteindelijk is gemaakt van materie, zie Stoljar, 2009) niet noodzakelijkerwijs hetzelfde is als reductionisme (de overtuiging dat hogere fenomenen formeel gereduceerd kunnen worden tot lagere fenomenen). In plaats daarvan betoogde hij dat de wetenschap inherent verdeeld is, en niet noodzakelijk een eenheid hoeft te vormen (1974, blz. 97). Volgens Fodor wordt de ontologische en epistemologische autonomie van de 'speciale wetenschappen' (disciplines zoals psychologie, sociologie en economie) niet in gevaar door fysicalisme als basisaanname te veronderstellen. Deze visie staat bekend als *niet-reductief fysicalisme* (of materialisme), en werd al snel populair onder filosofen (zie Block, 1997). Fodor betoogde verder dat of het mogelijk is wetenschappelijke disciplines te reduceren uiteindelijk een empirische vraag is, en dat er weinig empirische voorbeelden van succesvolle reductie zijn (maar zie Bickle, 1998, voor een andere kijk). Deze focus op de empirische kant van het reductieprobleem werd later benadrukt door Daniel Dennett (1995), die onderscheid maakt tussen reductionisme in het algemeen, namelijk de verdedigbare wetenschappelijke insteek om grotere eenheden te onderzoeken als bestaande uit kleinere onderdelen, en 'hebzuchtig reductionisme' (greedy reductionism). 'Hebzuchtig reductionisme' treedt op wanneer men poogt fenomenen te reduceren tot lagere verklaringsniveaus zonder de benodigde tussenstappen te definiëren. Dennett gebruikt de metafoer van een (conceptuele) hijskraan die we kunnen gebruiken om wetenschappelijke bevindingen op een bepaald verklaringsniveau stukje bij beetje te vertalen naar een niveau 'lager'. Als men bijvoorbeeld betoogt, omdat de hersenen nu eenmaal fysieke objecten zijn, dat psychologische toestanden 'eigenlijk' natuurkundige of scheikundige fenomenen zijn, zonder eerst de benodigde

tussenstappen vast te stellen (hijskranen die het 'vertaalwerk' tussen de niveaus kunnen doen), dan maakt men zich schuldig aan hebzuchtig reductionisme. Om dezelfde reden is het wetenschappelijk van weinig nut om te verkondigen dat de psychologie tot neurologie gereduceerd 'moet worden': De wetenschappelijke taak is het ontwikkelen van de theoretische en empirische hijskranen die deze reductie daadwerkelijk tot stand kan brengen.

Het is duidelijk dat zelfs binnen de aanname van het fysicalisme er veel onenigheid is over hoe wetenschappelijke wetmatigheden binnen verklaringsniveaus zich zouden kunnen laten vertalen naar andere verklaringsniveaus. Nergens is deze uitdaging zo groot als het reduceren van de geest naar het lichaam, of de psychologie naar het brein.

### **Het lichaam-geest probleem**

De relatie tussen lichaam en geest al eeuwen voor hoofdbreken onder zowel wetenschappers en filosofen. Sinds de psychologie aan het eind van de 19e eeuw als onafhankelijke wetenschappelijke discipline begon te opereren heeft ze zich beziggehouden met het vaststellen van de realiteit en de wetenschappelijke relevantie van de fenomenen die ze bestudeert: psychologische constructen en eigenschappen. De ontologische status van deze psychologische constructen is het onderwerp van veel discussie in de psychologische literatuur, zoals bijvoorbeeld de discussies over de status van 'algemene intelligentie' (Spearman, 1927). Dergelijke discussies kregen een nieuwe impuls in de tweede helft van de 20e eeuw, toen technologische vooruitgang het mogelijk maakte om psychologische eigenschappen en processen tegelijkertijd met biologische processen (bijvoorbeeld hersengolven) te bestuderen. De bestudering van mensen op biologische verklaringsniveaus, zoals de structuur en de activiteit van de hersenen, de invloed van hersenletsel op cognitie en de rol en functie van genen had vrijwel direct een grote invloed op de wetenschappelijke psychologie. Echter, ondanks de technische ontwikkelingen (non-invasief meten van hersenactiviteit), nieuwe paradigma's (het tijdelijk kunnen verstoren van hersenfunctie met magnetische pulsen) en een grote hoeveelheid onderzoek (165.000 hits in scholar.google.com door te zoeken naar "cognitive neuroscience"), is de precieze relatie tussen biologie en psychologie met betrekking tot de aard van psychologische constructen nog steeds onduidelijk.

Een van de meest beroemde, of beruchte, pogingen om de relatie tussen de geest en het lichaam te beschrijven is het dualisme van Descartes (1641). Hij concludeerde dat, omdat mentale toestanden andere eigenschappen dan fysieke toestanden hebben (alle fysieke objecten hebben 'extensie' - ze nemen driedimensionele ruimte in, terwijl dat niet voor 'gedachtes' geldt), ze wel andere soorten entiteiten moesten zijn. Hij betoogde dat het mentale uit een andere, niet-fysieke substantie bestaat die via de pijnappelklier met de fysieke hersenen interacteert. Aan het andere filosofische uiterste wordt betoogd dat de 'folk psychology', de alledaagse manier waarop we spreken over 'overtuigingen', 'herinneringen' of 'verlangens', fundamenteel onjuist is, en uiteindelijk door de neurowetenschappen vervangen zal worden. Dit standpunt staat bekend als eliminatief materialisme (Churchland, 1981), aangezien het doel is het hogere beschrijvingsniveau volledig weg te verklaren.

Tussen deze twee filosofische uitersten zijn een verscheidenheid van meer genuanceerde benaderingen ten opzichte van het lichaam-geest probleem ontwikkeld. Jaegwon Kim (1984, 1986) stelt dat mentale eigenschappen

*afhankelijk* zijn van fysieke eigenschappen. Dit idee, ook wel superveniëntie genoemd, houdt in dat concepten op hogere verklaringsniveaus worden bepaald door de 'bouwstenen' op een lagere verklaringsniveaus. Dit houdt in dat er slechts een verschil in het hogere construct (bijvoorbeeld 'jouw IQ is hoger dan dat van mij') kan zijn wanneer er ook een verschil in de hersenen bestaat ('jouw herseneigenschappen verschillen van mijn hersenen'). In hoofdstuk twee wordt in meer detail op dit idee ingegaan. Een andere theorie heet 'emergentie', en stelt dat complexe, hogere eigenschappen ontstaan wanneer constellaties van lagere entiteiten samenkomen. Vanuit dit perspectief bestaat een complexe eigenschap zoals 'intelligentie' simpelweg niet op het niveau van individuele neuronen: het is slechts wanneer vele miljarden neuronen samenkomen en interacteren dat intelligentie (of intelligent gedrag) kan ontstaan. In dit opzicht is de relevantie van deze theorieën van de cognitieve neurowetenschappen duidelijk: Kunnen we psychologische processen begrijpen door het zorgvuldig bestuderen van de bouwstenen van de hersenen (bijvoorbeeld neuronen, gliacellen en neurotransmitters), of bestaan de eigenschappen die ons interesseren (bewustzijn, intentie, pijn ervaren) alleen wanneer deze bouwstenen op de juiste manier samenkomen? Een laatste theorie over de relatie tussen lichaam en geest die voor ons relevant is is identiteitstheorie. Deze theorie suggereert dat een psychologische toestand identiek is aan de neurologische realisatie. Er zijn diverse interpretaties van identiteitstheorie. Een strenge vorm van identiteitstheorie, namelijk type-type identiteitstheorie, stelt dat een bepaald type neurale activiteit (bijvoorbeeld 'C-neuronen vuren') identiek is aan een bepaald type psychologische toestand ('pijn ervaren'). Minder strenge versies van identiteitstheorie stellen dat een bepaalde psychologische toestand *op een bepaald moment* identiek is aan één of andere neurale toestand op datzelfde moment (token-token identiteitstheorie). Deze interpretatie laat echter toe dat er geen regelmaat in de neurale realisatie hoeft te zitten: eenzelfde psychologische toestand (bijv. 'pijn ervaren') kan op verschillende momenten identiek zijn aan verschillende patronen van neurale activiteit. Een meer pragmatische conceptualisering van identiteitstheorie, Heuristische Identiteit Theorie (McCauley & Bechtel, 2001), stelt dat de *aanname* van een identiteitstheoretische relatie tussen hersenen en geest de meest productieve manier is om empirische vooruitgang te boeken. In hoofdstukken twee, drie, vier en zes gaan we in meer detail in op de verschillende vormen van identiteitstheorie.

## De gevaren van hebzuchtige (neuro)reductionisme

Hoewel bestudering van de precieze relatie tussen lichaam en geest een puur academische exercitie kan lijken, blijkt in de praktijk dat een verkeerd begrip van deze relatie directe gevolgen voor de samenleving kan hebben (Miller, 2010). Een wijdverbreide veronderstelling is dat het neurale verklaringsniveau, in zekere zin, meer tastbaar en dus 'echter' is. Deze manier van denken wordt ook wel 'neurorealisme' genoemd. Racine, Bar-Ilan en Illes (2005) onderzochten dit fenomeen door de bestudering van fMRI-gerelateerd nieuws, en de vaak kritiekloze rapportage van neurowetenschappelijke bevindingen. Ze beschrijven de valkuilen van neurorealisme, die volgens de auteurs in de ogen van de maatschappij een fenomeen kritiekloos 'echt', 'objectief' of 'effectief' kan doen lijken (blz. 160). Ze illustreren deze tendens door te wijzen op voorbeelden uit de populaire pers als "Vet eten brengt *echt* plezier" en "Een nieuwe vorm van neuroimaging *benijst* dat acupunctuur pijn

verlicht" (blz. 160). Een rapport van de NIMH (National Institute of Mental Health) geeft de volgende beschrijving van schizofrenie: 'Mentale ziekten zijn echte, diagnoseerbare, behandelbare hersenaandoeningen' (NIMH; Hyman, 1998, blz. 38, aangehaald in Miller, 2010). Verder evidentie voor neurorealisme werd gevonden in een enquête: Een meerderheid van ondervraagde mensen, gaven aan een zware depressie te zien als "een chemische onbalans" (80 %) en een "neurobiologisch probleem" (67 %; Pescosolido et al., 2010). Dit perspectief is uiteraard niet per definitie verkeerd, maar het is van groot belang dat overtuigingen die we hebben ten aanzien van de relatie tussen de hersenen en psychologie juist zijn, en gebaseerd zijn op solide theorieën en data. Zoals Pescosolido en collega's beargumenteren (citerend uit werk van Hinshaw, 2006): "Publieke opinie doet er toe: Ze vormen de brandstof voor de mythe dat een psychische aandoening per definitie levenslang, hopeloos, en afkeurenswaardig is" (blz. 1324).

Hoewel dit soort argumentaties het meest voorkomen in populair wetenschappelijke stukken, blijkt een zeer vergelijkbare heuristiek ook in wetenschappelijk schrijven veel voor te komen. Neem bijvoorbeeld de volgende beschrijvingen (cursief toegevoegd) over de rol van de biologische psychologie met betrekking tot psychologische constructen, zoals bijvoorbeeld persoonlijkheid: "temperamenten worden vaak beschouwd als *biologisch gebaseerde* psychologische tendensen met intrinsieke paden van ontwikkeling.", (McCrae et al., 2000, blz. 173), algemene intelligentie, "Dit *benijs* van biologische correlaten van *g* ondersteunt de theorie dat *g* geen methodologisch artefact maar *een feit van de natuur*" (Jensen, 1986, blz. 301) en "Ultiem begrip van *g* moet voortkomen uit de *meest diepgaande en gedetailleerde bestudering van het menselijk brein in zijn puur fysische en chemische aspecten*" (Spearman, 1927, blz. 403), en psychopathologische aandoeningen zoals schizofrenie, "Een doel van psychofysiologische onderzoek is zowel de diagnose als de symptomen *verankeren in biologische werkelijkheid*" (Ford, 1999, blz. 667). Onjuiste of overenthousiaste interpretatie van bevindingen uit de biologische psychologie kan grote gevolgen hebben. In India werd in 2008 bijvoorbeeld een man tot gevangenisstraf veroordeeld op basis van een EEG leugendetector (Deceiving the Law, 2008), ondanks diverse methodologische bezwaren ten opzichte van leugendetectie door middel van neuroimaging, zie bijv. Ganis et al., 2011). In Italië werd de straf van een voor moord veroordeelde man met 18 maanden verminderd omdat de verdediging beargumenteerde dat hij het 'agressie-gen' had (Feresin, 2009).

Het is duidelijk dat een coherente en verdedigbare visie ten aanzien van de relatie tussen lichaam en geest geen louter academische exercitie is. Het geven van 'ontologische voorrang' aan het biologisch verklaringsniveau als zijnde 'meer echt', zonder een coherent kader van reductie, is een duidelijk geval van hebzuchtige reductionisme, en kan aanzienlijke nadelige gevolgen hebben, niet alleen binnen de wetenschap, maar ook in de maatschappij. Om deze reden is een degelijk conceptueel kader voor de integratie van biologische en psychologische wetenschappen van essentieel belang. Een dergelijke aanpak zou beide verklaringsniveaus moeten integreren, en kunnen navigeren tussen de grotendeels mislukte uitersten van ongerechtvaardigd 'neurorealisme' enerzijds en de ontkenning van het belang van biologische processen voor de psychologie anderzijds. In de volgende paragraaf schets ik een dergelijk kader.

!



## Een psychometrische benadering van het reductieprobleem

Zoals blijkt uit het bovenstaande is het adequaat conceptualiseren van de precieze relatie tussen verklaringsniveaus in de wetenschap een verre van uitgemaakte zaak. Deze vraag is nergens dringender dan wanneer we de hersenen aan gedachten proberen te verbinden: hier dienen we de schijnbaar onverenigbare domeinen van ons subjectieve perspectief op de wereld enerzijds en miljarden vurende neuronen en neurotransmitters anderzijds te verenigen. De felheid van het debat heeft geleid tot de onverteerbare uitersten van het volledig ontkennen van het bestaan van psychologische toestanden (bijv. Churchland, 1981) enerzijds en de suggestie dat neurowetenschappen voor de psychologie irrelevant zouden zijn anderzijds (bijv. Fodor, 1999). Geen van deze uitersten sluit aan bij de empirische werkelijkheid. In dit proefschrift zal ik proberen dit vraagstuk op een nieuwe manier aan te pakken. Op basis van de filosofie van de geest, theorieën over het reductieprobleem en meettheorie zal ik betogen dat het reductieprobleem in feite, in elk geval ten dele, een *meetprobleem* is. Dat wil zeggen dat we, als we de relatie tussen de hersenen en de geest beter willen begrijpen, we zorgvuldig moeten weten wat we op ieder verklaringsniveau meten, en hoe we die metingen op het neurale niveau kunnen relateren aan psychologische metingen. Om die reden is het ontwikkelen en toepassen van theoretisch geïnspireerde statistische modellen essentieel.

Hier zullen we vooral focussen op latente variabele modellen. Latente variabele modellen zijn modellen die waarneembare variabelen (bijvoorbeeld reactietijdmetingen, de amplitude van een hersengolf, of het aantal goedgegemaakte items in een test) relateren aan gehypothetiseerde onderliggende constructen die ofwel de oorzaak (Borsboom, Mellenbergh & Van Heerden 2003) of het gevolg (Edwards & Bagozzi, 1991) van de geobserveerde metingen zijn. Voor ons doel betekent dit dat we het psychologische construct van belang (bijvoorbeeld 'intelligentie' of 'werkgeheugencapaciteit') conceptualiseren als een latente variabele, en pogen om het beste model te formuleren dat deze latente variabele relateert aan de waarneembare metingen van zowel neurale als psychologische aard. Daarmee bereiken we een aantal doelen. Ten eerste, de twee domeinen van neurale en gedragsmatige metingen worden, a priori, op gelijke voet gezet, waarmee we onnodige aannames over welk niveau 'belangrijker' is vermijden. Ten tweede, door de grafische voorstelling van hypothesen maken we impliciete causale veronderstellingen expliciet, en vergemakkelijken daarmee het wetenschappelijke debat over deze kwestie. Ten derde, de relatieve voorspellende waarde van neurale versus psychologische meetinstrumenten worden een puur empirische, in plaats van ideologische, kwestie: In zulke meetmodellen kunnen we de statistische eigenschappen (in termen van voorspellingen, meetfout en meer) direct vergelijken. Als laatste en wellicht meest belangrijk: Door het opstellen van specifieke modellen kunnen theoretische representaties direct kwantitatief vergeleken worden met andere kandidaten, of zelfs volledig verworpen worden.

### Overzicht

In dit proefschrift benader ik het reductieprobleem als een meetprobleem, met de nadruk op het gebruik van statistische modellen om de relatie tussen metingen van de hersenen en metingen van gedrag te verenigen. In hoofdstuk twee bestudeer ik de aard van het reductieprobleem in de cognitieve

neurowetenschappen. Ik betoog dat het reductieprobleem in feite een meetprobleem is, en dat om die reden de statistische relatie tussen de metingen van de neurale en psychologische domeinen ons kan informeren over de plausibiliteit van theorieën over het lichaam-geest probleem. Vanuit dit perspectief is het aloude lichaam-geest probleem, voor een niet onaanzienlijk deel, een kwestie die empirisch te vatten is. In dit hoofdstuk nemen we twee invloedrijke theorieën uit de filosofie van de geest, identiteitstheorie en superveniëntie, en vertalen deze naar psychometrische (structural equation) modellen. Deze twee modellen, een één-dimensionele reflectief meetmodel en een formatief, MIMIC-model, vertalen de theoretische aannames in deze twee theorieën (identiteitstheorie en superveniëntie) naar statistische voorspellingen. Wanneer deze modellen geformuleerd zijn, kunnen ze empirisch worden getoetst en de relatieve evidentie voor elk model voor een bepaalde dataset vergeleken worden. We vergelijken deze modellen voor twee datasets, één die de relatie tussen intelligentie en hersenomvang bestudeert en de ander die individuele verschillen in persoonlijkheid relateert aan grijze stof in verschillende plekken in het brein. Voor beide datasets blijkt het formatieve (MIMIC) model dat de theorie van superveniëntie vertegenwoordigt statistisch beter te passen. We bespreken de implicaties van deze bevinding en laten zien dat deze resultaten aansluiten bij de recente onderzoek naar de cognitieve neurowetenschap van emoties.

In hoofdstuk 3 ga ik in op een aantal commentaren van wetenschappers uit diverse disciplines met betrekking tot het kader dat in hoofdstuk twee uiteengezet wordt. Deze commentaren richten zich, onder andere, op de mechanistische interpretatie van modellen, de interpretatie van causaliteit wanneer hersenen aan gedrag worden gelinkt, de verschillen tussen psychologie in het lab en in een natuurlijke setting en het onderscheid tussen data enerzijds en de onderliggende fenomenen anderzijds. Om deze suggesties het hoofd te bieden breid ik het kader uit hoofdstuk 2 langs diverse lijnen verder uit. Ten eerste betoog ik dat goed ontwikkelde meetmodellen de beste manier zijn om het uiteindelijke doel van de cognitieve neurowetenschappen te bereiken: een beter inzicht in de neurale mechanismen die ten grondslag liggen aan psychologische processen. Ten tweede betoog ik dat de rechtvaardiging van de focus op de hersenen als het beste, of 'ware' verklarensniveau niet uitsluitend gebaseerd kan zijn op het feit dat de hersenen een fysiek object zijn, maar moet voortkomen uit andere criteria zoals de verklarende kracht, het voorspellend vermogen of inzicht in de onderliggende mechanismes. Ten derde laat ik zien hoe geobserveerd gedrag kan worden geïntegreerd in een hiërarchische versie van een MIMIC-model, om daarmee gedrag, psychologische processen en hersenstructuur in één model te kunnen vatten. Tot slot ga ik in op de psychometrische voorspellingen die voortvloeien uit diverse andere relevante theorieën uit de filosofie van de geest, zoals emergentie.

In hoofdstuk 4 breid ik de basisprincipes van hoofdstuk twee verder uit, en richt ik me specifiek op de relatie tussen intelligentie en de hersenen. Er zijn verschillende redenen voor de prominente rol van intelligentie in dit proefschrift (het komt empirisch aan bod in de hoofdstukken twee, drie, vier en zeven, en wordt als empirisch voorbeeld besproken in hoofdstuk vijf). Intelligentie is sinds het begin van de vorige eeuw van grote interesse in de wetenschappelijke psychologie. Dit heeft geleid tot vele empirische bevindingen, zoals het feit dat de prestaties op diverse algemene vaardigheden

over het algemeen positief correleren. Dat wil zeggen dat mensen die goed scoren op taak A, gemiddeld gesproken ook goed zullen scoren op taak B, C etcetera. Deze bevinding wordt ook wel de 'positive manifold' genoemd (Carroll, 1993). Ander werk heeft laten zien dat IQ-testscores gedurende de 20e eeuw gestaag zijn toegenomen (het Flynn-effect, Flynn, 1987); dat intelligentie gerelateerd is aan diverse aspecten van de sociaal-economische status zoals inkomen, onderwijs en gezondheidszorg (Gottfredson & Deary, 2004), en dat intelligentie gerelateerd is aan diverse meer fundamentele cognitieve vaardigheden zoals reactietijd en perceptuele snelheid (Jensen, 2006). Maar ondanks de voorspellende waarde van intelligentie zijn er nog steeds vele felle debatten over wat intelligentie nu eigenlijk is, hoe we het optimaal zouden moeten meten en meer. Deze discussies hebben betrekking op de veranderingen in IQ scores over generaties heen (Flynn, 1987), de interpretatie van latente variabelen (Kan, Kievit, Dolan, & Van der Maas, 2011), de rol van dynamische modellen (Van der Maas et al., 2006), de verschillen in factorstructuur tussen mensen die hoog en laag scoren op intelligentietests (Molenaar, Dolan, Wicherts, & Van der Maas, 2010) en vele andere kwesties. Naast de statistische en conceptuele discussies over de aard van intelligentie, bestuderen cognitieve neurowetenschappers mogelijke "biologische correlaten" van intelligentie, de biologische eigenschappen die mogelijk de onderliggende oorzaak, of essentie, van verschillen in intelligentie zouden kunnen vormen. Voorgestelde biologische correlaten zijn onder andere hersengrootte (McDaniel, 2005), de mate van connectiviteit van witte stofbanen, de grijze stof dichtheid in de hersenen (Jensen & Sinha, 1993), plaatselijke laesies (Woolgar et al., 2010), hersenactiviteit gedurende complexe redeneertaken (Gray, Chabris & Braver, 2003) en de gelijkensis in intelligentiescores tussen eenenige tweelingen en twee-eiige tweelingen (cf. Deary, Penke & Johnson, 2010). Sommigen stellen op zoek te zijn naar een 'neuro g', (Haier et al., 2009), een onderliggende neurale eigenschap die intelligentie volledig zou kunnen vatten. Ondanks de vele studies blijft de precieze relatie tussen intelligentie en de verschillende voorgestelde 'neurale correlaten' onduidelijk. In hoofdstuk vier probeer ik verscheidene mogelijke representaties van deze relatie in modellen te vatten om ze zo empirisch te vergelijken. Net als in hoofdstuk twee vind ik dat een formatief (MIMIC) model, waarbinnen een gewogen geheel van neurale eigenschappen samen (gedeeltelijk) algemene intelligentie verklaren het best door de data wordt ondersteund.

De eerste hoofdstukken focussen op hoe we het best, statistisch gezien, hersenmetingen aan psychologische metingen kunnen relateren, in een poging de relatie tussen de verklaringsniveaus beter te begrijpen. In hoofdstuk 5 bekijk ik in detail welke problemen kunnen ontstaan wanneer onzorgvuldig inferenties over verschillende verklaringen worden getrokken. Meer specifiek onderzoek ik het voorbeeld van Simpson's Paradox (Simpson, 1951). Simpson's Paradox treedt op wanneer, binnen dezelfde dataset, een statistische patroon op groepsniveau (bijvoorbeeld 'Er is een positieve relatie tussen de inname van koffie en neuroticisme') in de omgekeerde richting optreedt binnen all subgroepen (bijvoorbeeld 'voor zowel vrouwen als mannen bestaat er een negatieve relatie tussen de inname van koffie en neuroticisme'). Simpson's paradox treedt op wanneer onzorgvuldig conclusies worden getrokken die verklaringen niveaus overbruggen, zoals van grotere groepen naar subgroepen (bijv. man en vrouw) of van subgroepen naar individuen die door de tijd heen veranderen. Op basis van simulaties en experimenteel onderzoek (dat laat zien

dat mensen niet goed zijn in het herkennen van de paradox) en een overzicht van empirische voorbeelden uit de psychologie en de aangrenzende velden concludeer ik dat Simpson's Paradox, naar alle waarschijnlijkheid, ondergediagnosticeerd is. Simpson's paradox is nuttig als illustratie van het feit dat onze intuïtieve gevolgtrekkingen ons gemakkelijk op een dwaalspoor kunnen brengen, vooral wanneer hypothesen betrekking hebben op verschillende verklaringsniveaus. Ik probeer dit probleem aan te pakken door middel van een diverse suggesties voor het voorkomen, en indien nodig 'genezen', van gevallen van Simpson's Paradox. Om gevallen van Simpson's Paradox in nieuwe datasets op te sporen leveren we een software pakket, geschreven in R, (Kievit & Epskamp, 2012, zie bijlage B) bij die automatisch de aanwezigheid van subgroepen kan detecteren, en kan controleren of conclusies op groepsniveau generaliseren naar mogelijke subgroepen in de data.

In hoofdstuk 6 bespreek ik een nieuwe benadering om gedrag en hersenenactiviteit te integreren: De studie van representatieve geometrie. Vanuit dit perspectief kunnen we neurale of psychologische representaties zien als punten in een hoog-dimensionale ruimte. Deze ruimte kan direct worden gedefinieerd in termen van inherente eigenschappen van de meting (bijvoorbeeld het aantal verschillende plekken in het brein die we kunnen meten), of het aantal dimensies waarop een persoon een stimulus moet beoordelen), of indirect worden afgeleid door middel van technieken zoals multidimensional scaling (Kriegeskorte & Mur, 2012; Shepard, 1958). Wanneer deze ruimte eenmaal gedefinieerd is kunnen stimuli met elkaar worden vergeleken en gecontrasteerd in termen van diverse metingen zoals de mate waarin ze op elkaar lijken (similariteitsanalyse), of de mate waarin bepaalde typen stimuli samen clusteren binnen een subdomein van de multivariate ruimte. Door het definiëren van representaties in een dergelijke, abstracte ruimte kunnen we neurale en psychologische representaties direct met elkaar vergelijken. Dit maakt representatieve geometrie een veelzijdige manier om de hersenen aan mentale toestanden te relateren, aangezien het een manier geeft om weg te abstraheren van irrelevante details en in plaats daarvan te kijken naar de structuur van mentale representaties. In dit hoofdstuk bespreken we de theoretische basis van deze techniek, laten we zien hoe het in diverse cognitieve domeinen tot inzichten heeft geleid, en laten we zien hoe dit kader kan worden gebruikt om individuen en groepen met elkaar te vergelijken. Tot slot bespreken we kort hoe representatieve geometrie kan worden gebruikt om de identiteitstheorie op een nieuwe manier te testen. We laten zien hoe we met behulp van representatieve geometrie zowel type-type, type-token en token-token identiteitstheorie kunnen formaliseren en toetsen, en dat deze interpretatie van identiteitstheorie een rijker beeld van identiteitstheorie geeft dan meer traditionele interpretaties.

In hoofdstuk 7 onderzoek ik de relatie tussen hersenfunctie en abstract redeneervermogen (ook wel 'vloeibare intelligentie' genaamd). Wetenschappelijke studies van het redeneervermogen richten zich vaak op één van twee mogelijke dimensies: Verschillen tussen mensen (in hun redeneervermogen), en verschillen binnen mensen (in de moeilijkheid van de redeneertaak die moet worden uitgevoerd). Veel artikelen over redeneervermogen voegen deze twee dimensies samen of negeren het onderscheid, ondanks het feit dat ze in principe onafhankelijk van elkaar kunnen zijn: ee (neurale) eigenschappen die onderscheid maken tussen mensen die beter of minder goed scoren op redeneervermogen hoeven niet dezelfde

processen te zijn als het verschil in hersenactiviteit binnen individuen wanneer ze redeneertaken van toenemende moeilijkheid uitvoeren. Ik laat zien hoe twee dimensies van vloeibare intelligentie, de inter-individuele differentiatie tussen mensen en de intra-individuele differentiatie tussen meer of minder uitdagende redeneertaken tegelijk kunnen worden gemodelleerd met behulp van het zogeheten Rasch model (Rasch, 1960). Door beide dimensies als voorspellers mee te nemen tijdens het bestuderen van hersenactiviteit laat ik zien dat de beide dimensies in de hersenen zeer verschillend zijn. De twee dimensies, verschillen in vaardigheid en verschillen in moeilijkheid, vertonen ook gedeeltelijk overlap in een subset hersengebieden. Ik introduceer de term neurale ergodiciteit voor gebieden die zowel op de intra-individuele dimensie als op de inter-individuele dimensie grotere activiteit vertonen.

## Discussie

In de discussie, bespreek ik bestaande uitdagingen voor het kader dat geschetst is en de veelbelovende mogelijkheden voor verdere statistische modellering van reductieve theorieën. Allereerst bekijken we de centrale vraag: Kunnen we psychologische constructen op een 1-op-1 manier vertalen naar eigenschappen van het brein? Met andere woorden, kunnen we hersenscanners het best zien als een omslachtige manier om dezelfde eigenschappen te meten? Het centrale resultaat van deze dissertatie is dat, op basis van diverse empirische illustraties en theoretische overwegingen, dit *niet* het geval is. Of, in termen van de introductie: Het zijn niet ‘schildpadden, helemaal tot aan de bodem’. Met andere woorden, de relatie tussen de hersenen en gedrag is asymmetrisch, en de statistische modellen die we gebruiken om de twee verklaringsniveaus te linken dienen die aanname accuraat te representeren. Er dient een slag om de arm gehouden te worden aangezien het mogelijk is dat andere scantechnieken, andere psychologische constructen of andere methodes wel een op-één mapping tussen hersenen en cognitie toelaten. In hoofdstuk zes stellen we één zo’n andere methode voor, waar we niet naar de hersenactiviteit zelf kijken maar naar het patroon van activiteit ten opzichte van andere patronen.

De tweede kernconclusie is dat een vorm van ontologisch pluralisme essentieel is. Met andere woorden, het debat over het reductionisme is niet een vraag tussen hersenen OF gedrag, maar hersenen EN gedrag. Onze taak als wetenschapper is om modellen te ontwikkelen die de twee verklaringsniveaus op een informatieve wijze aan elkaar kunnen relateren. Om dit te bewerkstelligen dienen cognitieve neurowetenschappers zich, in elk geval ten dele, te laten informeren door theorieën uit de filosofie van de geest, of in elk geval stil te staan bij de aannames die vrijwel per definitie gedaan worden wanneer een analyse wordt uitgevoerd. Tegelijkertijd dienen filosofen van de geest geïnformeerd te zijn door recente wetenschappelijke ontwikkelingen en doorbraken, omdat die theorieën kunnen informeren, aanpassen en inspireren. Belangrijker nog, filosofen zouden als onderdeel van het ontwikkelen van een theorie stil moeten staan bij hoe de theorie, mogelijkerwijs, empirisch getoetst zou kunnen worden.

Uiteraard heeft de hier voorgestelde aanpak ook enkele beperkingen. Een versimpelende aanname die door deze hele these gemaakt wordt is dat het psychologische en het neurale niveau ieder als één eenheid gezien kunnen worden. In werkelijkheid is het binnen zowel de neuro- als de psychologische wetenschap gebruikelijk om de relevante constructen op diverse resolutieniveaus te bestuderen. Verder zijn alle neuroimaging technieken

imperfect: Zo heeft fMRI bijvoorbeeld uitstekende spatiële resolutie, maar tamelijk slechte temporele resolutie (en geldt voor EEG het omgekeerde). Het is dus mogelijk dat een andere constellatie van neurale metingen of technieken tot een andere conclusie ten aanzien van de plausibiliteit van identiteitstheorie en superveniëntie zou hebben geleid. Een meer praktisch probleem van de voorgestelde aanpak is dat Structural Equation Modellen vaak grote proefpersoon aantallen vereisen. Dit is moeilijk te rijmen met de kosten van neuroimaging: Een doorsnee studie heeft ongeveer 20 proefpersonen, wat een flinke beperking oplegt aan het type modellen dat überhaupt gefit zou kunnen worden. Het delen van data op grote schaal wordt echter steeds gebruikelijker, en zou dit probleem in de nabije toekomst grotendeels moeten kunnen oplossen. Een andere manier om dit aan te pakken is het gebruik van simulaties. Er zijn diverse ontwikkelingen gaande die het mogelijk maken om hersenactiviteit op diverse resolutieniveaus mogelijk te maken. Hoewel we ver af zijn van het creëren van kunstmatige intelligentie, is het inmiddels mogelijk om bepaalde empirische observaties, zoals het patroon van hersenactiviteit tijdens rust, met enige accuratesse te simuleren (zie bijv. Gosh et al., 2008). Een voordeel van deze aanpak is directe en volledige controle over de aannames die onderdeel uitmaken van een simulatie, waardoor deze invloeden in detail kunnen worden bestudeerd.

### Conclusie

Het is verleidelijk om te denken dat de verwarrende en imperfecte beschrijvingen uit de psychologie plotseling perfect helder worden wanneer we in plaats van slechts naar gedrag ook in de hersenen kijken. Echter, er is zowel geen a priori reden om er van uit te gaan dat de hersenen het 'juiste' verklaringsniveau zijn (waarom niet de genen, of biologische moleculen), en a posteriori weinig empirische evidentie dat de psychologie op korte of lange termijn vervangen zal worden door neurowetenschappen. Dit betekent niet dat de hersenwetenschappen (of de genetica) irrelevant zijn voor de psychologie, integendeel: Er zijn in relatief korte tijd immense doorbraken bereikt die ons begrip over de psychologie en de onderliggende mechanismen immens hebben vergroot en verdiept. Het is onze taak als cognitieve neurowetenschappers om theorieën en modellen te ontwikkelen die recht doen aan zowel de psychologie als de neurowetenschappen, om uiteindelijk op onderling inzichtelijke wijze deze twee verklaringsniveaus aan elkaar te koppelen. Alleen zo kunnen we grip krijgen op misschien wel de grootste wetenschappelijke uitdaging van onze tijd: de relatie tussen lichaam en geest.