

Dualisme, reductionisme en the hard problem of consciousness

Bart Klink

Augustus 2019

De dieprode kleur van een roos, de bittere smaak van een glas bier, het prachtige kleurenspeel van een zonsondergang en de tinteling op je arm van een koel briesje op een warme zomerdag. Dergelijke bewuste ervaringen hebben een bepaald subjectief kwalitatief aspect dat moeilijk te omschrijven valt. Filosofen noemen dergelijke kwalitatieve aspecten van ervaringen *qualia*, waarvan het enkelvoud *quale* is.

Sinds de opkomst van de moderne neurowetenschappen weten we dat de hersenen een belangrijke rol vervullen bij dergelijke ervaringen en al ons andere geestelijk leven (denken, herinneren, voelen enz.). Hoe kan de elektrochemische communicatie tussen neuronen (zenuwcellen) ervoor zorgen dat wij ervaringen hebben met een bepaalde kwaliteit, de *qualia*? Zorgen ze daar wel voor? Onze intuïtie zegt dat dergelijke ervaringen iets fundamenteels anders zijn dan hersenactiviteit. Hersenactiviteit is iets materieels, te beschrijven met chemische en fysische wetmatigheden, zij het zeer complex. De ervaring lijkt echter niet materieel te zijn, maar iets onstoffelijks, zonder massa, ruimte en vorm. Dit lijkt ook te gelden voor andere mentale toestanden¹: gedachten hebben geen vorm, gevoelens geen gewicht en een herinnering heeft niet een bepaalde (fysieke) vorm. We lijken hier te maken te hebben met twee fenomenen van een fundamenteel andere orde: mentale toestanden en breinactiviteit.

Vormen van dualisme

Het idee dat dit ook daadwerkelijk twee verschillende toestanden zijn, wordt dualisme genoemd. De meest rigoureuze vorm hiervan is *substantiedualisme*, wat inhoudt dat het geestelijke (mentale) en het stoffelijke (brein, lichaam) onafhankelijk van elkaar zijn en dus ook los van elkaar kunnen bestaan, zoals in een leven na de dood of een uittredingservaring. Dit idee gaat al terug tot de oude Grieken, maar is vooral bekend geworden door René Descartes (1596-1650). Dit idee wordt in de moderne neurowetenschappen niet meer serieus genomen en is onder moderne filosofen een minderheidspositie, zeker onder filosofen van de geest².

Er is echter een subtielere vorm van dualisme, waarbij lichaam en geest niet als onafhankelijk van elkaar gezien worden, maar toch wezenlijk anders zijn. Bij dit *eigenschapsdualisme* (*property dualism*) worden twee verschillende eigenschappen onderscheiden van dezelfde (fysieke) substantie: mentale en fysieke eigenschappen. Een populaire vorm hiervan is het niet-reductief fysicalisme: het mentale (de geest) is niet te reduceren tot het fysieke (brein),

¹ Het mentale kan onderverdeeld worden in 1) ervaringen, 2) cognitieve toestanden, toestanden die ergens over gaan (intentionaliteit), zoals gedachten of overtuigingen, en 3) emoties (vaak een combinatie van 1) en 2) (Dooremalen et al., 2010). Dit onderscheid is vooral van belang omdat volgens het eliminatief materialisme ervaringen gereduceerd en niet geëlimineerd worden, maar (bepaalde) cognitieve toestanden mogelijk wel.

² Zoals een veelgebruikt studieboek neurowetenschappen aangeeft: “In neuroscience, there is no need to separate mind from brain; once we fully understand the individual and concerted actions of brain cells, we will understand our mental abilities.” (Bear et al., 2015:24). Zie verder Schouten & Looren de Jong (2012); Dehaene (2014); Blackmore & Troscianko (2018) en Gazzaniga (2018). Verderop zal ik het substantiedualisme verder bekritisieren.

maar is daar wel een eigenschap van. Dit staat tegenover de positie dat het mentale wel te reduceren valt tot het brein (reductief fysicalisme): *the mind is what te brain does*. Op dit neuroreductionisme zal hieronder verder worden ingegaan. Volgens de eigenschapsdualist kan de geest niet los bestaan van het lichaam, volgens de substantiedualist wel. De substantiedualist maakt dus een sterkere claim en heeft daarmee ook een zwaardere bewijslast.

In het dagelijks leven zijn we dualisten: we denken en spreken over lichaam en geest als gescheiden. Zo verstaan we onder lichamelijke klachten wat anders dan onder geestelijke klachten en kunnen we spreken over een jonge geest in een oud lichaam. Dit dualisme is intuïtief: we nemen het aan zonder erover na te denken, het *voelt* natuurlijk. Kleine kinderen denken ook van nature dualistisch (Bloom, 2005) en het idee van een leven na de dood (hemel/hel of reïncarnatie) is wijdverbreid. Dit dualisme dat we vanzelf aannemen is onderdeel van onze volkspychologie: de manier waarop we van nature denken over onze psychologie en die van onze medemensen (en eventueel dieren) (Churchland, 2013).

De moderne psychologie en neurowetenschappen hebben deze volkspychologie echter stukje bij beetje ontmanteld. Onze psychologie blijkt heel anders te werken dan we nature lijken te denken, zoals in elke moderne inleiding tot de psychologie na te lezen valt. Een vergelijkbare ontwikkeling heeft zich voltrokken in de astronomie. In de (geocentrische) volksastronomie staat de aarde stil in het middelpunt van het universum en draait de zon daaromheen. In de moderne astronomie draaien we met ruim 100.000 km/h om de zon, en ook nog eens met ruim 1.000 km/h om onze eigen aard-as (in Nederland). Dit is lastig voor te stellen en nog lastiger (of niet) te ervaren, maar toch is het waar. Onze intuïtieve volksastronomie is vervangen door een tegen-intuïtieve wetenschappelijke astronomie.

Reductionisme

Reductie en reductionisme zijn (vooral negatief) beladen termen. Ze zijn daarnaast onderwerp van complexe en vaak technische discussies in zowel wetenschap als filosofie, en ik pretendeer geenszins die hier te beslechten. In de literatuur worden verschillende definities gegeven, en onderscheiden verschillende auteurs ook verschillende vormen van reductionisme³. Wie de woorden ‘reductie’ en ‘reductionisme’ tegenkomt, moet zich altijd eerst afvragen wat de auteur eronder verstaat⁴.

In dit stuk betoog ik vooral *ontologisch*⁵ reductionisme van bewustzijn: het bewustzijn *is* een breintoestand⁶. In de filosofie van de geest wordt dit ook wel de identiteitstheorie⁷ genoemd omdat het mentale, waaronder bewustzijn, identiek is aan hersentoestanden: *the mind is what the brain does*. In de biologie (waaronder de neurowetenschappen) komt reductie vaak neer op het ontdekken van mechanismen (Bechtel, 2012). Zo’n mechanisme kan simpel of

³ Zie voor beknopte inleidingen Schouten & Looren de Jong (2012) en Brigandt & Love (2017).

⁴ In de eerste aflevering (*Wezens op zoek naar zingeving*) van de EO-serie *Waarom zijn we op aarde?* (2017) lopen verschillende vormen van reductionisme door elkaar, wat voor verwarring zorgt.

⁵ Ontologie gaat over wat bestaat. Volgens de substantiedualist bestaan lichaam en geest apart (ze hebben een aparte ontologie); volgens de ontologisch reductionist bestaat slechts het lichaam/brein.

⁶ Hierbij dient wel opgemerkt te worden dat het brein functioneert in een lichaam en in een omgeving (*embodied, embedded cognition*). Voor dit betoog maakt dit echter weinig verschil: het brein staat hier voor een brein in een lichaam, in de wereld.

⁷ Op de verschillen tussen type- en token-identiteit en het vermeende probleem van de ‘meervoudige realiseerbaarheid’ zal ik hier niet verder ingaan omdat het voor mijn betoog irrelevant is, zie verder Dooremalen et al. (2010) en Churchland (2013).

complex zijn, lineair of niet-lineair, statisch of dynamisch. Een reductie is dan geslaagd als de eigenschappen, functies en causale vermogens verklaard kunnen worden door mechanismen op een lager niveau. Op die manier zoeken neurowetenschappers naar de mechanismen achter leervermogen, cognitie, geheugen, besluitvorming, aandacht, waarneming, aansturing van beweging, en ook bewustzijn. In de biologie verklaren we tegenwoordig op die manier ook leven: leven is gereduceerd als we de moleculaire mechanismen en processen in de cel begrijpen. Er is geen extra, niet-fysisch, ingrediënt nodig naast de werking van de relevante mechanismes, niet voor leven en niet voor bewustzijn (Churchland, 2002).

Dit ontologisch reductionisme is gemeengoed in de neurowetenschappen, impliciet of expliciet (Bear et al., 2015). Het is zo evident en empirisch goed onderbouwd dat dit het uitgangspunt vormt voor vrijwel al het moderne onderzoek naar bewustzijn, waarneming, cognitie, emoties, lichaamsbesef, besluitvorming, leren en ziektebeelden, maar ook onderzoek naar computersimulaties en kunstmatige intelligentie. Het is daarmee een buitengewoon vruchtbaar onderzoeksprogramma waarmee vele toetsbare voorspellingen gedaan kunnen worden⁸. Wie dit niet inziet, is niet op de hoogte van de relevante literatuur.

Het substantiedualisme daarentegen kampt al eeuwen met dezelfde problemen en is qua onderzoeksprogramma steriel. Het kan niet verklaren waarom bijvoorbeeld drugs het bewustzijn veranderen, hoe en waarom illusies ontstaan, wanneer we ons wel of juist niet van iets bewust worden, waarom het doorsnijden van de hersenbalk leidt tot twee onafhankelijke ‘bewustzijnen’ in dezelfde persoon⁹, hoe we *brain-computer interfaces*¹⁰ kunnen aansturen, waarom de linkerhand van iemand met *alien hand syndrome*¹¹ wat anders wil dan zijn rechter, waarom hersenziektes tot mentale achteruitgang leiden en hoe we kunnen leren van ervaringen. De neurowetenschappers komen hiervoor echter met steeds betere reductionistische verklaringen, en praktische toepassingen. Het substantiedualisme wordt door onderzoekers niet meer serieus genomen (Dehaene, 2014; Gazzaniga, 2018). Een voorbeeld van een aannemelijke huidige reductionistische verklaring voor bewustzijn is de *global neuronal workspace theory* (Dehaene, 2014)¹².

Ontologisch reductionisme is vaak onderdeel van een breder fysicalisme (ook wel ‘naturalisme’ of ‘materialisme’ genoemd), het idee dat alles wat bestaat materie-energie in ruimte-tijd is en zich gedraagt volgens natuurwetten¹³. Bovennatuurlijke zaken als geesten, zielen, goden, demonen en een leven na de dood bestaan dus niet. Dit fysicalisme is niet

⁸ Een voorbeeld hiervan is het Hurvich–Jameson-model van kleurwaarneming. Met dit reductionistische model kunnen niet alleen allerlei fenomenologische eigenschappen van kleurwaarneming verklaard worden, maar kunnen ook precieze voorspellingen gedaan worden die blijken te kloppen. Zie verder Churchland (2005b).

⁹ De zogenaamde *split-brain patients*, waarbij de verbinding tussen beide hersenhelften (*corpus callosum* of hersenbalk) verbroken is. Bij zo’n patiënt ervaart en weet de linkerhersenhelft iets anders dan de rechterhersenhelft. Zie verder: <https://en.wikipedia.org/wiki/Split-brain>

¹⁰ Dit zijn systemen waarbij het brein direct communiceert met een computer. Zo kan bijvoorbeeld een verlamd persoon met door de computer uitgelezen gedachten een robotarm aansturen. Zie verder: https://en.wikipedia.org/wiki/Brain-computer_interface

¹¹ Zie verder: https://en.wikipedia.org/wiki/Alien_hand_syndrome

¹² Aangezien het wetenschappelijk onderzoek naar bewustzijn nog in de kinderschoenen staat, is deze theorie ongetwijfeld onvolledig en mogelijk ook deels onjuist, ook al is ze goed empirisch gefundeerd, doet ze toetsbare voorspellingen en zorgt ze voor een vruchtbaar onderzoeksprogramma. Mijn punt is hier dat de (ontologisch) reductionistische benadering empirisch sterk staat en wetenschappelijk vruchtbaar is. Mijn voorspelling is dat alle toekomstige succesvolle theorieën over bewustzijn ook (ontologisch) reductionistisch zullen zijn.

¹³ Over de definitie van fysicalisme is een hoop discussie in de filosofische literatuur. Soms wordt onder de begrippen ‘(metafysisch) naturalisme’ en ‘materialisme’ net iets anders verstaan. Voor de huidige discussie is dit verder irrelevant.

verenigbaar met substantiedualisme, maar wel met eigenschapsdualisme (als niet-reductief fysikalisme), menen de aanhangers hiervan.

Zelfs als de ontologisch reductionist gelijk heeft, wil dit niet zeggen dat we bewustzijn alleen neurowetenschappelijk moeten gaan onderzoeken en analyseren. De werking van het brein – en dus ook bewustzijn – kan op verschillende manieren of niveaus onderzocht worden, van (wetenschappelijke) psychologie tot moleculaire biologie. Dit zijn verschillende analyseniveaus die elkaar kunnen voeden en beïnvloeden met nieuwe inzichten, waardoor ze co-evolueren¹⁴. Voor sociale interactie – uiteindelijk de interactie tussen breinen – lijkt een nog hoger analyseniveau zinvoller: dat van de sociologie. Deze verschillende analyseniveaus vereisen geen aparte ontologie en zijn dus geen probleem voor de ontologisch reductionist.

Onder reductionisme wordt soms ook het idee verstaan dat we de werking van het geheel kunnen begrijpen door alleen naar de werking van de delen te kijken, of het geheel te zien als een eenvoudige (lineaire) optelsom van de delen. Deze vorm van reductionisme is terecht bekritiseerd als te simplistisch, zeker bij complexe fenomenen en functies als in de biologie, waaronder de neurowetenschappen. De neuronale netwerken waaruit het brein opgebouwd is, vormen buitengewoon complexe en niet-lineaire dynamische systemen, waardoor het gedrag van het geheel (brein) anders en onvoorspelbaarder is dan het gedrag van de delen (neuronen, of zelfs de onderdelen daarvan). Het gedrag van neuronen bestuderen is zeker zinvol, maar onvolledig om de werking van het gehele brein te begrijpen (Krakauer et al., 2017). Dit geldt vermoedelijk ook voor een complexe hersentoestand als bewustzijn. De functies van het geheel zijn *emergente*¹⁵ eigenschappen die niet terug te vinden zijn in de onderdelen van het geheel. Dit is echter geen probleem voor ontologisch reductionisme; (zwak) emergente eigenschappen behoeven geen aparte ontologie.

Eliminatief materialisme en illusionisme

Binnen het reductionisme zijn er filosofen, zoals de Paul en Patricia Churchland, die menen dat al ons natuurlijke spreken over het mentale (volkspychologie) uiteindelijk goddeels geëlimineerd kan en moet worden (eliminatief materialisme¹⁶) omdat dit een achterhaalde en onjuist theorie¹⁷ is. Over mentale toestanden kan het beste gesproken worden in neurowetenschappelijke termen¹⁸. Hoewel ik meen dat dit *in principe* mogelijk is, is het in veel gevallen niet praktisch, zeker niet in het dagelijks leven. De volkspychologie

¹⁴ Of er in een vergevorderd stadium van de neurowetenschappen nog ruimte is voor een autonome psychologie (en sociologie), of dat de beste verklaringen op neuronaal of moleculair niveau zullen plaatsvinden, is lastig te voorspellen. De toekomst zal dat moeten uitwijzen. Mijn eigen idee is dat de complexiteit op lagere niveaus te groot is om het gedrag op hogere niveaus goed te analyseren en te verklaren, ook al is dit *in principe* mogelijk. Daarom hoeven de psychologie en sociologie niet te vrezen voor hun bestaansrecht, mits ze consistent blijven met de lagere niveaus.

¹⁵ Dit is een zwakke vorm van emergentie, er ontstaat niet iets van een ontologisch andere orde. De discussie over emergentie is complex en uitgebreid, zie verder: O'Connor & Wong (2015); zie ook de klassieker Anderson (1972).

¹⁶ Reductie en eliminatie geven de uitersten van een spectrum aan. Reductie is een identificatie (licht is elektromagnetische straling, ervaringen *zijn* hersentoestanden), waarbij de fenomenen blijven bestaan, maar anders worden uitgelegd. Bij eliminatie verdwijnt een theorie of concept omdat ze door voortschrijdend inzicht nergens naar blijken te verwijzen (zoals 'flogiston', 'heks' en 'onstoffelijke geest'). Het concept van een atoom zit tussen deze uitersten in: atomen bestaan echt, maar anders dan aanvankelijk werd gedacht (onsplitsbare deeltjes).

¹⁷ Met een theorie bedoel ik hier het conceptuele kader over hoe de werkelijkheid in elkaar zit en werkt, en waardoor we onze ervaringen interpreteren.

¹⁸ Patricia Churchland gaf in 2013 aan hier wat milder in te zijn geworden (persoonlijke communicatie).

functioneert daarin te goed om plaats te maken voor een neurowetenschappelijk discours. Een *conceptueel dualisme* (Papineau, 2002) kan ook best zinnig zijn, waarbij we verschillende concepten gebruiken voor zaken waarvan we weten dat ze in werkelijkheid hetzelfde zijn (zoals geest en brein). Wie echter dieper wil graven, zoals psychologen, neurowetenschappers en filosofen doen, zal de volkspychologie in meer of mindere mate overboord moeten zetten omdat deze beperkt, onjuist en misleidend is. Veel van hen blijven onbewust deze volkspychologie (of delen daarvan) aanhangen, wat tot een hoop misvattingen en verwarring leidt. Ik kom hier later nog op terug.

Een veelgemaakt misverstand over het eliminatief materialisme is dat het stelt dat ervaringen niet bestaan en daarom geëlimineerd moeten worden. Wat volgens deze materialisten geëlimineerd moet worden, is evenwel *de manier waarop we denken en spreken* over het mentale, inclusief ervaringen (de volkspychologie dus). Het gaat dus om eliminatie van concepten, noties, theorieën en verklaringen, niet om eliminatie van de ervaringen of fenomenen zelf. Een voorbeeld hiervan is de verklaring voor leven. Lange tijd is gedacht dat wat iets levend maakt, een onstoffelijke *élan vital* is, een idee dat ook bekend staat als vitalisme. In de moderne biologie wordt leven echter gezien als een complex van moleculairbiologische mechanismen en biochemische processen. Hierbij is het *concept* van *élan vital* geëlimineerd, niet het *fenomeen* leven zelf. Op vergelijkbare wijze willen eliminatief materialisten niet het bewustzijn zelf elimineren, maar ons denken en spreken erover in volkspychologische termen. Voor *concepten* als ‘geest’ en ‘ziel’ en het *idee* dat bewustzijn iets onstoffelijks is, is dan geen plaats meer, maar voor het *fenomeen* bewustzijn (of de *ervaring* daarvan) nog wel.

Het is ook goed mogelijk dat ons huidige idee van bewustzijn als één bepaald verschijnsel te simpel of onjuist is. Mogelijk bestaan er meerdere vormen van bewustzijn die door verschillende hersengebieden gerealiseerd worden, zoals ook het geval bleek te zijn met het fenomeen geheugen. Mogelijk hebben verschillende dieren verschillende vormen van bewustzijn (zeker als hun zenuwstelsel anders werkt dan dat van zoogdieren, zoals bij octopussen) en mogelijk zal kunstmatige intelligentie in de toekomst nog een andere vorm van bewustzijn krijgen. Door dergelijke ontwikkelingen zullen onze opvattingen over bewustzijn over een eeuw net zo veranderd zijn als wat gebeurd is met onze opvattingen over leven, materie, ruimte en tijd in de afgelopen eeuw.

Een verwant misverstand gaat over het illusionisme, het idee dat bewustzijn een illusie is¹⁹. Daniel Dennett is al decennia een bekende verdediger van dit idee. Het misverstand komt voort uit het verwarren van twee verwante betekenissen van het woord ‘illusie’, namelijk: 1) iets bestaat niet, en 2) iets is niet wat het lijkt. Als we zeggen dat een gebroken rietje in een glas water (vanaf de zijkant gezien) een illusie is, zeggen we niet dat het rietje niet bestaat, maar dat het *lijkt alsof* het rietje gebroken is, ook al is het dat in werkelijkheid niet. Wanneer mensen als Dennett beweren dat bewustzijn een illusie is, zeggen ze niet dat bewustzijn niet bestaat, maar *dat het niet is wat het lijkt*²⁰. Het is een *psychologisch* vraagstuk waar deze

¹⁹ Ook deze discussie is complex en komt voor een groot deel neer op wat verstaan wordt onder begrippen, zoals blijkt uit een recente bundel van Keith Frankish (2017). Frankish zelf betoogt dat *fenomonaal* bewustzijn niet bestaat, maar ervaringen wel. Niemand zal ontkennen dat er een verschil is tussen het zien van blauw en groen, of tussen de pijn van een brandwond en die van een snee met een mes. Ik vraag me dan ook af wat het woord ‘fenomonaal’ hier zou moeten betekenen als het niet onderdeel van de ervaring zelf is. Zie voor overtuigende kritiek op Frankish de bijdragen van onder andere Pete Mandik en Jesse Prinz in dezelfde bundel.

²⁰ Zoals hij onder andere zegt in zijn gesprek met Nick Bostrom: "I'm not saying that consciousness doesn't exist. I'm just saying it isn't what you think it is." (op 32:20, <https://www.youtube.com/watch?v=NZwmrtq7tdI>)

illusie vandaan komt, net als dat het een natuurkundig-psychologisch vraagstuk is waarom het rietje gebroken lijkt. Hoe deze illusie ontstaat, is het onderwerp van veel interessant wetenschappelijk onderzoek waarop ik later nog zal terugkomen.

Kunnen we qualia dan reduceren of elimineren? Dat hangt af van hoe qualia precies gedefinieerd worden. Als we qualia zien als het kwalitatieve aspect van een ervaring, dan kunnen we qualia reduceren: een quale *is* een bepaalde hersentoestand. Als qualia echter *per definitie* ontologisch verschillend zijn van een hersentoestanden (onstoffelijke mentale fenomenen), dan zullen qualia geëlimineerd moeten worden²¹.

Mijn opvatting is dat we bewustzijn als ervaring kunnen reduceren tot hersenactiviteit (ontologisch reductionisme, identiteitstheorie) en dat we het volkpsychologisch denken en spreken erover goeddeels kunnen elimineren. In serieus neurowetenschappelijk en filosofisch discours zouden we dat ook (grotendeels) moeten doen om misverstanden en verwarring te voorkomen. In het dagelijks leven is dat niet nodig omdat onze volkpsychologie daar goed genoeg werkt vanuit praktisch oogpunt, ook al is ze onjuist. Op vergelijkbare wijze blijven we in het dagelijks leven spreken over een zon die opkomt en ondergaat, ook al weten we dat dit onjuist is en het in serieus astronomisch discours niet gebruikt kan worden.

The hard problem

Dan kunnen we ons nu gaan richten op *the hard problem of consciousness*.

Neurowetenschappers kunnen misschien steeds beter beschrijven hoe het brein werkt en nagaan welke hersenactiviteit samengaat met een bepaalde ervaring, maar daarmee hebben ze nog niet verklaard hoe die bepaalde hersenactiviteit *leidt tot* die bepaalde ervaring. Deze stap van hersenwerking naar bewuste ervaringen is volgens de filosoof David Chalmers het moeilijke probleem van bewustzijn: *the hard problem of consciousness*. Dit probleem lijkt zo evident en groot dat velen er het hoofd al over gebogen hebben, en het volgens sommigen onoplosbaar is. Als dit probleem werkelijk onoplosbaar zou zijn, is dat een argument tegen het fysicalisme (aldus de verdedigers), want dan moet er naast het fysieke nog iets anders bestaan: een onstoffelijk domein waartoe ook onze ervaringen (en eventueel de geest) behoren. Ik zal echter betogen dat *the hard problem* een schijnprobleem is. Er is helemaal geen probleem, al lijkt dat wel zo. Het probleem is een illusie, gevoed door onze intuïtie en volkpsychologie. Dat maakt de schijn van dit schijnprobleem lastig te doorzien. Aan deze intuïtieve plausibiliteit ontleent dit probleem (en het daarop gebaseerde argument tegen fysicalisme) zijn kracht, ook al is het onjuist. Het échte moeilijke probleem (*hard problem*) is dan ook om te verklaren hoe deze zeer krachtige illusie ontstaat. Hier kom ik later nog op terug

Chalmers (1995:201) formuleert *the hard problem* zelf als volgt:

“It is undeniable that some organisms are subjects of experience. But the question of how it is that these systems are subjects of experience is perplexing. Why is it that when our cognitive systems engage in visual and auditory information-processing, we have visual or auditory experience: the quality of deep blue, the sensation of middle C? How can we explain why there is something it is like to entertain a mental image,

²¹ De vraag is waar begrippen als qualia, *fenomenaal* bewustzijn en *what-it-is-likeness* (Nagel, 1974) precies naar verwijzen, en of ze niet-circulair gedefinieerd kunnen worden. Er lijkt hier eerder sprake van obscuur filosofisch jargon dat voor meer verwarring zorgt dan helderheid en vooruitgang scheidt in de discussie. Zie verder de bijdrage van Pete Mandik in Frankish (2017).

or to experience an emotion? It is widely agreed that experience *arises* from a physical basis, but we have no good explanation of why and how it so *arises*. Why should physical processing *give rise to* a rich inner life at all? It seems objectively unreasonable that it should, and yet it does.” [cursivering BK]

Uit dit citaat blijkt waar Chalmers de mist in gaat. De cruciale woorden heb ik gecursiveerd. Door woorden als *arise* (voortkomen uit) en *give rise to* (leiden tot) te gebruiken, gaat Chalmers impliciet al uit van een vorm van dualisme, waardoor er een ‘ontologisch gat’ *ontstaat* tussen het mentale en het fysieke. Zo komen ervaringen volgens hem *voort uit* een fysieke basis en *leiden* fysieke processen *tot* een innerlijk leven. Blijkbaar *bestaan* ervaringen niet uit een fysieke basis en ons innerlijk leven niet uit fysieke processen, maar komen ze daaruit voort. Zijn formulering van het probleem bevat dus al een vorm van dualisme, terwijl het probleem dat juist moet aantonen! Het probleem bestaat bij gratie van zijn definitie. Voor wie met zijn definitie van het probleem niet akkoord gaat – het *gedefinieerde* ontologische gat dus niet accepteert – is hier helemaal geen probleem. Dit punt is al eerder gemaakt in de filosofische literatuur (Dennett, 1996; Papineau, 2011; Carruthers & Schier, 2017). Desalniettemin blijft *the hard problem* de gemoederen bezighouden en zien sommigen er een argument tegen reductionisme of zelfs fysicalisme als geheel in. Hoe kan dat?

Chalmers heeft het voordeel dat zijn probleemstelling *intuïtief* plausibel klinkt: het *voelt* alsof er een ontologisch gat bestaat tussen het mentale en het fysieke. Deze intuïtie is sterk en ik kan de aantrekkingskracht ervan voelen. Het is dus niet vreemd dat veel mensen met Chalmers meegaan in deze intuïtie, en dan volgt *the hard problem* vanzelf. Maar wat bewijst deze intuïtie? Onze intuïties zitten er wel vaker naast. Intuïtief is het ook plausibel dat de aarde stilstaat en de zon daaromheen draait, zeker als je geen kennis hebt van de moderne astronomie. Dat dit een illusie is, was dan ook moeilijk voor te stellen in de dagen van Copernicus en Galilei. Het voelt niet alsof de aarde om de zon draait (met ruim 100.000 km/h), en ook nog eens om zijn eigen as (met ruim 1.000 km/h in Nederland). Een beroep op intuïties alleen is dus niet erg overtuigend, zeker niet als er goede redenen zijn om daaraan te twijfelen, zoals verderop zal blijken.

De reductionistische visie die meent dat Chalmers ons een schijnprobleem voorschotelt, wordt ook gedeeld door bijvoorbeeld de topbewustzijnsonderzoeker Stanislas Dehaene: “Once our intuition is educated by cognitive neuroscience and computer simulations, Chalmers’s hard problem will evaporate. The hypothetical concept of qualia, pure mental experience detached from any information-processing role, will be viewed as a peculiar idea of the prescientific era, much like vitalism—the misguided nineteenth-century thought that, however much detail we gather about the chemical mechanisms of living organisms, we will never account for the unique qualities of life.” (Dehaene, 2014:262).

Helaas zijn er ook neurowetenschappers die wel (reductief) fysicalisme voorstaan, maar soms onbewust blijken te geven van een vorm van dualisme. Zo schrijft de hersenwetenschapper Dick Swaab in de inleiding van zijn beroemde boek *Wij zijn ons brein*²²: “Zoals de nier urine

²² De titel van dit boek is een geval van wat ik ‘conceptuele contaminatie’ noem: een verhaspeling van concepten uit twee verschillende denkkaders. De voornaamwoorden ‘wij’ en ‘ons’ zijn hier volkpsychologische concepten en ‘brein’ een neurowetenschappelijk concept. Als de ‘wij/ik’ bestaat uit hersenprocessen (wat volgens de reductionist het geval is), komt de titel neer op de tautologie *Het brein is het brein*. Dit lijkt muggenzifterij, maar kan een hoop verwarring geven in deze discussie. Gelukkig maakt Swaab aan het begin van de inleiding duidelijk wat hij hiermee bedoelt: “Alles wat we denken, doen en laten gebeurt door onze hersenen. De bouw van deze fantastische machine bepaalt onze mogelijkheden, onze beperkingen en ons karakter; *wij zijn onze hersenen*.” (p. 23, cursivering in origineel).

produceert, produceert het brein de geest.” (Swaab, 2001, p. 24). Maar urine en de nieren bestaan onafhankelijk van elkaar, brein en geest niet. Een vergelijkbaar probleem zien we in de literatuur over de *neural correlates of consciousness* (NCC's), de (minimale) neuronale activiteit die correleert met bewustzijnstoestanden. De relatie tussen deze neuronale activiteit en de bewustzijnstoestanden wordt vaak opgevat als causaal: er is een correlatie omdat neuronale activiteit de bewustzijnstoestanden *veroorzaakt*. Dit impliceert echter dat neuronale activiteit iets anders is dan de bewustzijnstoestanden die het veroorzaakt, een impliciet dualisme dus. Volgens de reductionist is dit een denkfout: de bewustzijnstoestanden *zijn* de neuronale activiteit en is er slechts een correlatie in perspectieven (eerste- versus derdepersoons). De beste verklaring voor de empirische gevonden correlatie is identiteit, niet causaliteit (Polák & Marvan, 2018).

Ondermijning van de intuïtie

Een beroep doen op intuïtie is vaak terecht. Als ik een rode appel voor mij zie, concludeer ik intuïtief dat daar ook werkelijk een rode appel is. Dit is *prima facie* terecht, tenzij er goede redenen zijn om die intuïtieve conclusie te betwijfelen. Door de *ervaring* dat de aarde stil staat en de zon langs de hemelkoepel beweegt, hebben vele mensen voor de tijd van Copernicus geconcludeerd dat de aarde ook *daadwerkelijk* stil staat en de zon daaromheen draait. Dit was *prima facie* gerechtvaardigd omdat er geen goede reden was daaraan te twijfelen. Nu hebben we die reden uiteraard wel en snappen we ook hoe de geocentrische illusie ontstaat. De *prima facie* gerechtvaardigde intuïtieve conclusie is dus ondermijnd door voortschrijdend wetenschappelijk inzicht.

We hebben ook dergelijke goede redenen om te denken dat onze intuïtieve kijk op het mentale niet hoeft overeen te stemmen met de werkelijkheid: het is evolutionair nutteloos. Als het al mogelijk zou zijn dat we bijvoorbeeld bewustzijn kunnen ervaren als wat het volgens de reductionist werkelijk is (hersensactiviteit), welk evolutionair voordeel zou dat ons dan opleveren? Welk nut zou het voor ons hebben als we continu een stroom van informatie binnen zouden krijgen over elektrische signalen in zenuwcellen en chemische interacties tussen zenuwcellen? Evolutie heeft ons daarentegen uitgerust met een nuttig *model* van het bewustzijn waarmee we ons prima kunnen redden in de wereld om ons heen. Dennett noemt dit een *user illusion* (gebruikersillusie), net zoals dat onze computer ons een ‘bureaublad’ voorschotelt en doet alsof onze bestanden in ‘mappen’ georganiseerd zijn, terwijl het in werkelijkheid allemaal complexe fysieke toestanden van de hardware zijn. Dit is een nuttige manier om de werking van de computer voor te stellen omdat de werkelijke werking veel te complex is.

Voor wie een breder evolutionair perspectief op onze epistemologie heeft, is het niet vreemd dat onze meeste vertrouwde opvattingen over de werkelijkheid vaak niet stroken met de werkelijkheid zelf. Onze intuïties en beeld van de werkelijkheid zijn ontstaan om in sociale verbanden te overleven op de savanne, niet om de complete werkelijkheid te doorgronden. Ons wereldbeeld en onze intuïties zijn de afgelopen eeuwen regelmatig fundamenteel bijgesteld door nieuw wetenschappelijk inzicht. Het voorbeeld van een draaiende aarde die om een ster draait, is hierboven al besproken, maar ook de relativiteit van tijd, de kromming van ruimte, een universum van 13,8 miljard jaar oud, vaste stof die toch voor veruit het grootste deel leeg blijkt te zijn, wit licht dat uit vele kleuren blijkt te bestaan en dat we doodziek kunnen worden van (voor het blote oog) onzichtbare organismen zijn nog meer voorbeelden die tegen-intuïtief zijn, maar toch waar. Waarom zou onze intuïtieve kijk op het bewustzijn wel overeenstemmen met de werkelijkheid? Waarom zouden we ook niet hier een

tegen-intuïtieve wetenschappelijke correctie mogen verwachten, zoals de reductionist claimt? Het feit dat we dit psychologisch niet kunnen vatten – het ons geen ‘snapgevoel’ (De Regt & Dooremalen, 2015) oplevert – wil niet zeggen dat een verklaring onjuist is (Vlerick & Boudry, 2017).

De cognitiewetenschapper Michael Graziano heeft met zijn *attention schema theory* een interessante theorie ontwikkeld over hoe dit (illusionaire) model ontstaat en werkt (Graziano, 2013; Graziano & Webb, 2015). Ook de filosoof Brian Fiala werkt een psychologische hypothese uit over hoe de intuïtieve illusie ontstaat (Fiala, z.d.). Dergelijke neurowetenschappelijke en psychologische verklaringen voor het ontstaan van de dualistische intuïtie werken ondermijnend voor de dualistische intuïtie, en daarmee ook voor *the hard problem*. Als dergelijke verklaringen immers juist zijn, hebben we een goede reductionistische verklaring voor het ontstaan van de illusie, en is het *hard problem* zelf geen probleem meer.

Dat wij de reductionistische identiteitstheorie zo lastig kunnen begrijpen en niet intuïtief kunnen voelen, is een psychologisch feit over ons, niet een metafysisch feit over de werkelijkheid. Het is een radicale theorie die indruist tegen sterke intuïties, maar dat geldt voor meer wetenschappelijke theorieën, waaronder die van Copernicus en Galilei. Misschien hebben critici van reductionisme ook wel een te beperkt beeld van wat materie kan. Als het ervoor kan zorgen dat totaal verschillende fenomenen als water, rotsen en levende dieren mogelijk zijn, waarom dan niet ook de hersentoestand die wij bewustzijn noemen? De meest recente kennis uit de neurowetenschappen ondersteunt niet alleen de identiteitstheorie, maar kan ook verklaren waarom die zo tegen-intuïtief is. Mijn eigen ervaring is dat hoe beter je vertrouwd raakt met de neurowetenschappen, hoe gemakkelijker het wordt om deze intuïties los te laten en ze te zien voor wat ze zijn: evolutionair nuttige illusies. Dit blijft vooral een cognitieve stap in begrip. Intuïtief blijft het evenwel vreemd aanvoelen – net als het idee dat we om de zon draaien overigens.

Argumenten tegen dualisme

We hebben dus goede argumenten die de dualistische intuïtie ondermijnen, maar er zijn ook argumenten tegen het dualisme, nog sterker tegen het substantiedualisme dan tegen het eigenschapsdualisme. De argumenten tegen het substantiedualisme heb ik elders al besproken (Klink, 2007; zie ook Churchland, 2002; Dooremalen, 2010; Churchland, 2013; Martin & Augustine, 2015). De voornaamste argumenten zijn het interactieprobleem en het argument van causale geslotenheid.

Het eerste argument is het probleem hoe de onstoffelijke geest invloed kan hebben op het stoffelijke brein en vice versa. Dit probleem is al meer dan twee millennia oud en er is nog nooit een overtuigend antwoord op gevonden, en het ziet er ook niet naar uit dat dat nog gaat komen. Er moet hier sprake zijn van *mentale* veroorzaking (*mental causation*). Zo zou bijvoorbeeld mijn gedachte (mentaal) mijn brein (fysiek) aan moeten sturen als ik een glas oppak. De dualist moet echter eerst aannemelijk maken dát een dergelijke vorm van veroorzaking überhaupt bestaat en hoe dit te rijmen is met de behoudswetten uit de natuurkunde (zie hieronder). Dit is echter nog nooit overtuigend gedaan. De dualist kan nog tegenwerpen dat het geen probleem is dat we niet weten hoe het mentale en het fysieke elkaar beïnvloeden, want dat komt wel vaker voor in de wetenschap. Ook Newton wist bijvoorbeeld niet hoe zwaartekracht werkte. Deze tegenwerking is echter niet overtuigend. Het was vanaf het begin duidelijk dát de zwaartekracht er is en dat dit meetbaar is. Newtons claims waren

toetsbaar. Dit geldt allemaal niet voor de vermeende kracht achter mentale veroorzaking. Daarnaast is de werking van zwaartekracht de afgelopen eeuwen steeds verder opgehelderd en is dit consistent met de rest van de natuurkunde, terwijl dit geenszins het geval is bij mentale veroorzaking. Het interactieprobleem is en blijft dus een groot probleem voor de substantiedualist.

Het argument van causale geslotenheid (Papineau, 2002) houdt in dat de natuurwetenschappen laten zien dat elke oorzaak een fysieke oorzaak is en de natuurwetten (vooral de behoudswetten) geen ruimte laten voor niet-fysiek ingrijpen, zoals nodig is voor invloed van het onstoffelijke mentale. Causale geslotenheid is geen aanname, zoals nogal eens wordt beweerd, maar de meest waarschijnlijke conclusie gezien onze beste wetenschappelijke kennis. Ook de natuurkundige Sean Carroll (2016) laat zien dat er voor een onstoffelijke ziel of geest die ingrijpt op het brein geen ruimte is in de moderne natuurkunde. De reductionist heeft de problemen met interactie en causale geslotenheid niet. Ook laat het substantiedualisme al eeuwen vele vragen onbeantwoord en is het een steriel onderzoekprogramma, zoals we hierboven reeds hebben gezien. Bedenk ook dat de waarheid van het substantiedualisme gemakkelijk aangetoond had *kunnen* worden als mentale vermogens, waaronder bewustzijn, los blijken te kunnen bestaan van een functionerend brein, zoals telepatie of waarneming tijdens hersendood. Dit is uitgebreid onderzocht, maar nog nooit overtuigend aangetoond (Martin & Augustine, 2015; Musolino, 2015). Het idee dat de hersenen slechts een ontvanger zijn voor bewustzijn, een idee vooral populair in new-agekringen, is ook om tal van redenen niet geloofwaardig (Duch, 2017).

De eigenschapsdualist (zoals Chalmers) zit ook met het probleem van mentale veroorzaking: als het mentale een eigenschap is van het fysieke brein, maar daar niet aan gelijk is (ertoe kan worden gereduceerd), hoe kan het mentale dan een effect hebben op het fysieke? Volgens het epifenomenalisme heeft het mentale ook geen effect op het fysieke, het is er slechts een bijeffect van, net zoals de stoom uit een locomotief. Hiermee ontstaat echter een probleem: hoe kunnen we beschrijven wat er zich in het mentale afspeelt, bijvoorbeeld de quale van een rode roos? Onze spraakspieren moeten worden aangestuurd door de juiste gebieden in het brein, maar hoe zou dat kunnen als de onstoffelijke quale geen causaal vermogen heeft op het fysieke brein? Daarnaast is er de vraag aan wie of wat de ervaringen zich openbaren. In een onstoffelijke ziel gelooft de eigenschapsdualist niet, maar wie of wat neemt de onstoffelijke qualia dan waar en hoe? Ook snijdt dit evolutionair geen hout: wat is het nut van de mentale eigenschap bovenop de fysieke eigenschap? Hoe kan deze eigenschap überhaupt via evolutionaire mechanismen ontstaan zijn? De reductionist heeft deze problemen niet.

Argumenten tegen reductionisme en voor een *hard problem*?

Hierboven hebben we gezien dat de identiteitstheorie sterke papieren heeft, dualisme grote problemen heeft en dat het bestaan van een *hard problem* gebaseerd is op een illusionaire intuïtie, die op zichzelf goed te verklaren valt. Zijn er dan verder geen argumenten voor een *hard problem* en tegen het ontologisch reductionisme in te brengen? Die zijn de afgelopen decennia gegeven in de literatuur, maar ik meen dat ze allemaal falen. Deze argumenten zijn reeds uitgebreid bekritiseerd in de literatuur (zie o.a. Churchland, 2013), dus ik bespreek ze hier slechts kort.

Zombies

Chalmers (1996) probeert met een beroemd gedachte-experiment zijn *hard problem* kracht bij te zetten: de filosofische zombie. Deze zombie is, in tegenstelling tot die uit horrorfilms en

-series niet van een gewoon mens te onderscheiden. Qua uiterlijke verschijning, gedrag en in gesprekken is hij niets anders – molecuul voor molecuul hetzelfde – maar hij ervaart bij dit alles niets. Hij heeft geen innerlijk leven, geen qualia. Zijn hersenen verwerken informatie, maar hebben daarbij geen ervaringen. Omdat we ons zo'n zombie voor kunnen stellen en hij dus mogelijk is, is dit – aldus Chalmers - een argument tegen het reductief fysicalisme. Het probleem is echter dat deze mogelijkheid slechts een logische mogelijkheid is, en een logische mogelijkheid niet ook een werkelijke mogelijkheid hoeft te zijn. Zo is het ook logisch mogelijk dat een mier een trompet bespeelt, al is dit niet in werkelijkheid mogelijk. Het is aan Chalmers om aan te tonen (of ten minste aannemelijk te maken) dat filosofische zombies ook *werkelijk* kunnen bestaan. Daar is hij vooralsnog niet in geslaagd, en we hebben geen enkele reden om te denken dat ze bestaan. Zijn argument werkt dus alleen als je erin meegaat, en niets dwingt daartoe. In wezen is zijn argument een cirkelredenering: als je *aanneemt* dat zombies kunnen bestaan, is er inderdaad een *hard problem*, maar dat is nou juist wat aangetoond moet worden!

De vraag is zelfs of zo'n zombie voorstelbaar is. Daarover zijn de meningen verdeeld²³. De enige reden hiervoor lijkt te zijn dat sommigen die intuïtie hebben, maar ik heb die intuïtie zelf niet, en met mij vele anderen niet. Voor mij klinkt een volledig werkend brein zonder bewustzijn net zo onzinnig als H₂O dat geen water is of licht dat geen elektromagnetische straling is. Beide zijn volgens mij alleen voorstelbaar als je niet thuis bent in respectievelijk de neurowetenschappen en de schei- en natuurkunde. Op dergelijke dubieuze intuïties valt dus geen argument te bouwen, waardoor het zombie-argument faalt.

Chalmers vraagt zich ook af waarom de informatieverwerking van het brein gepaard moet gaan met bewuste ervaringen. Kunnen zombies niet net zo goed overleven en voortplanten? De vraag is of dat het geval is. Als weefselchade niet gepaard zou gaan met een bewuste negatieve ervaring van pijn (zoals bij patiënten die lijden aan pijn-asymbolie (Wegner & Jacobs, 2011)), is de reden voor een organisme om iets aan de oorzaak van die weefselchade te doen en het in het vervolg te voorkomen minder groot. Als seks niet lekker zou zijn (in ieder geval voor het dominante geslacht), zou de drang tot voortplanting waarschijnlijk minder groot zijn. Het gevoel dat met bepaalde informatieverwerking gepaard gaat, de affectieve component, heeft dus waarschijnlijk een evolutionair voordeel. Evolutionair lijken zombies dus minder succesvol.

De verklaringskloof

Een vergelijkbaar idee met Chalmers *hard problem* is Joseph Levines *explanatory gap* (verklaringskloof) dat hij in 1983 introduceerde (Levine, 1983). Deze kloof bestaat tussen de neurowetenschappelijke verklaringen en het kwalitatieve aspect van de ervaring, want deze qualia blijven volgens Levine onverklaarbaar. Zelfs als we een goede neurowetenschappelijke verklaring hebben voor pijn, dan blijft er iets onverklaard: “why the pain should feel the way it does!” (p. 357), de quale van de pijn dus. Volgens de reductionist is dit echter hetzelfde als je afvragen waarom de verbinding tussen twee waterstofatomen en één zuurstofatoom water oplevert, en niet iets anders. Dit is een onzinnige vraag. De vraag *lijkt* zinnig omdat ook hier weer de dezelfde intuïtie opspeelt dat het kwalitatieve aspect van de ervaring niet een fysieke toestand *kan* zijn. Levine ziet dit zelf ook in en geeft aan het einde van zijn artikel toe dat er een uitweg is: “one must either deny, or dissolve, the intuition which lies at the foundation of

²³ Uit een groot onderzoek onder (analytische) filosofen (Bourget en Chalmers, 2014) blijkt dat 35,6% zombies voorstelbaar, maar niet metafysisch mogelijk acht, 23,3% ze metafysisch mogelijk acht, 16% ze onvoorstelbaar acht en 25,1% er nog anders over denkt.

the argument” (p. 360). Dit is ook precies wat ik meen dat we moeten doen: inzien dat de intuïtie een illusie is.

Mary de kleurenwetenschapster en vleermuizen

Een ander argument dat tegen het fysicalisme is ingebracht, is het gedachte-experiment van Frank Jackson (1982). Hierin weet de kleurwetenschapster Mary alles wat er te weten valt over kleuren. Ze is alleen opgesloten in een zwart-witruimte en heeft dus zelf nooit kleur ervaren, ook al heeft ze er wel alle mogelijke wetenschappelijke kennis van. Ze heeft daarmee alle kennis *over* qualia, maar nooit qualia zelf *ervaren*. Op het moment dat ze deze zwart-witruimte verlaat en de kleurenwereld instapt, ervaart ze dus voor het eerst zelf kleur²⁴. Ze leert hierbij iets nieuws, iets wat ze op grond van haar wetenschappelijke kennis niet wist. Deze nieuwe ervaring is dus méér dan wetenschappelijke kennis, wat volgens Jackson een argument tegen het fysicalisme is.

Hij kwam hier later zelf op terug, en terecht. Dit argument is ook door anderen op verschillende manier bekritiseerd (Churchland, 2013). De sterkste kritiek is volgens mij dat een ervaring zelf wat is anders dan een beschrijving van die ervaring. Het gaat om verschillende perspectieven op hetzelfde. De ervaring is vanuit een eerstepersoonsperspectief, de beschrijving vanuit een derdepersoonsperspectief. Dit is een epistemologisch verschil (een verschil in hoe we kennen), niet een ontologisch verschil (een verschil in wat bestaat) of een metafysisch verschil (een verschil in hoe de werkelijkheid in elkaar zit en werkt). Beide perspectieven hoeven elkaar ook niet uit te sluiten. Zo kan ik subjectief voelen dat mijn elleboog gebogen is (eerstepersoonsperspectief) én kan een ander dat objectief beschrijven (derdepersoonsperspectief). Ook binnen het (reductief) fysicalisme is dit verschil in perspectief te verwachten omdat een ervaring een andere toestand is van het brein dan een beschrijving van die ervaring, ofschoon het beide breintoestanden zijn. Een beschrijving of verklaring van een ervaring is wat anders dan de verklaring zelf, net als dat een beschrijving of verklaring van zwangerschap wat anders is dan zwanger zijn.

Bijna een decennium eerder maakte Thomas Nagel een vergelijkbaar argument onder de ondertussen beroemde titel *What is it like to be a bat?* (Nagel, 1974). Ook al zouden we alles weten over hoe vleermuizen met sonar waarnemen, weten we dan *wat het is om als een vleermuis te zijn?* Het antwoord is natuurlijk ‘nee’: *kennis over* waarneming staat niet gelijk aan de waarneming *zelf*, zoals we ook bij Mary hebben gezien. De enige manier om te weten hoe het is om een vleermuis te zijn, is de anatomie en fysiologie van een vleermuis te hebben, maar dat is voor een mens per definitie onmogelijk. Jackson en Nagel hebben interessante gedachte-experimenten opgeworpen, maar ze zeggen slechts wat over epistemologie, niet over ontologie of metafysica. Als argument tegen reductionisme of fysicalisme falen ze dan ook. Ook Nagel geeft in zijn stuk overigens toe dat uit zijn punt niet volgt dat fysicalisme onjuist is.

Wet van Leibniz

De Wet van Leibniz houdt in dat x en y alleen identiek kunnen zijn als alle uitspraken die voor x gelden ook voor y gelden. Volgens de dualist is dit niet het geval voor hersentoestanden en mentale toestanden (waaronder bewustzijn), waardoor mentale toestanden geen hersentoestanden *kunnen* zijn. Mentale toestanden hebben bijvoorbeeld geen massa, volume, dichtheid en ruimtelijke vorm, terwijl breintoestanden die wel hebben, aldus de dualist.

²⁴ De vraag is of dit neurofysiologisch überhaupt wel mogelijk is, maar dat maakt voor het gedachte-experiment niet uit.

Dit is echter wederom een cirkelredenering. De dualist *veronderstelt* immers op grond van zijn intuïtie dat mentale toestanden geen massa, volume, dichtheid en ruimtelijke vorm hebben, en dat dit *dus* geen hersentoestanden kunnen zijn. Het is echter precies deze intuïtieve veronderstelling die de dualist moet rechtvaardigen. Immers, alleen als die veronderstelling waar is, is het dualisme gerechtvaardigd. Maar als we deze veronderstelling verwerpen – en dat is precies wat de reductionist doet – komt dit argument niet van de grond.

Verder is er nog een probleem met het toepassen van Leibniz' Wet. Die werkt alleen als x en y onderdeel zijn van dezelfde theorie. Met een theorie bedoel ik hier het conceptuele kader over hoe de werkelijkheid in elkaar zit en werkt, en waardoor we onze ervaringen interpreteren. Waarnemingen zijn dus altijd *theoriegeladen*: we interpreteren onze waarnemingen vanuit de theorie die we hebben, bewust of onbewust.

Lange tijd is bijvoorbeeld gedacht dat de Avondster en de Morgenster twee verschillende hemellichamen waren omdat ze op verschillende momenten zichtbaar zijn (respectievelijk vlak na zonsondergang en vlak voor zonsopkomst). Volgens Leibniz' Wet zouden de Avondster en Morgenster dus niet hetzelfde object kunnen zijn. Door ontwikkelingen in de astronomie weten we nu dat deze twee verschillende 'sterren' in werkelijkheid een en dezelfde planeet zijn, namelijk Venus. Wat in de premoderne astronomie (de oude theorie) dus twee verschillende objecten waren, is in de moderne astronomie (nieuwe theorie) dus een en hetzelfde object. Zo kunnen veranderende theoretische inzichten gebaseerd op wetenschappelijke bevindingen tot nieuwe identificaties leiden.

Hetzelfde punt kunnen we maken met waarnemingen vanuit een geocentrisch en heliocentrisch wereldbeeld (theorie). 'De koperen ploert die langs de hemelkoepel beweegt' (oude theorie) heeft niet dezelfde eigenschappen als 'de ster in ons zonnestelsel' (nieuwe theorie). Toch zijn beide objecten identiek (de zon), alleen benaderd vanuit verschillende theorieën, waardoor de eigenschappen verschillend *lijken*, maar het in werkelijkheid niet zijn.

Ik meen dat het met mentale toestanden en hersentoestanden net zo werkt. In onze volkpsychologie (oude theorie) lijken (onstoffelijke) mentale toestanden inderdaad fundamenteel anders dan (stoffelijke) hersentoestanden, maar dat is een illusie. Door psychologische en neurowetenschappelijke ontwikkelingen (nieuwe theorie) weten we dat mentale toestanden en hersentoestanden naar hetzelfde verwijzen, alleen vanuit een ander conceptueel kader. De Wet van Leibniz kan dus niet als argument gebruikt worden omdat x en y onderdeel zijn van twee verschillende theorieën (conceptuele kaders).

Correlatie is geen identiteit

Een ander argument is dat correlatie geen identiteit hoeft te zijn. Mentale toestanden kunnen immers *perfect* gecorreleerd zijn met hersentoestanden. Het probleem hiermee is dat een perfecte correlatie niet van identiteit te onderscheiden is. Het is een niet te toetsen claim die bovendien een extra entiteit (het mentale) veronderstelt. Je zou analoog kunnen stellen dat licht geen elektromagnetische straling *is* (zoals de reductionist meent), maar er slechts perfect mee correleert. Dit verklaart echter niets meer of beter dan wat de reductionist beweert, en we hebben geen goede reden om licht als iets *naast* elektromagnetische straling te postuleren. Als we vanuit de werking van het brein alles kunnen verklaren wat wordt toegeschreven aan het mentale, hebben we het mentale gereduceerd tot hersenfunctie, en is er geen reden om het mentale te postuleren *naast* hersenfunctie. Identiteit is dan een (ontologisch) simpelere verklaring en verdient daarom de voorkeur (Ockhams scheermes).

Daarnaast roept het idee van een perfecte correlatie een hoop vragen op die de reductionist niet heeft: wat zorgt voor deze perfecte correlatie? Hoe wordt mijn geest gecorreleerd met mijn brein en niet met dat van een ander? Wanneer ontstaat die correlatie in de evolutie en tijdens de embryonale ontwikkeling? Wat gebeurt er met de correlatie op het moment van sterven? Waarom blijft de correlatie bestaan bij hersenziektes, zoals Alzheimer? Is er wel of geen interactie? Waarom is die perfecte correlatie er überhaupt? Dit zijn vragen waar nog nooit een overtuigend antwoord op gegeven is, en niets wijst erop dat die antwoorden nog gaan komen. Als we het idee van de *perfecte* correlatie opgeven, en de ziel of geest los kan bestaan van het lichaam, komen we op de eerder besproken problemen met het substantiedualisme. De reductionist heeft al deze problemen niet.

Metafysica, geen wetenschap?

Een verwant argument aan het voorgaande is dat de reductionistische identiteitsclaim een metafysische claim is en geen wetenschappelijke. De wetenschap kan helemaal geen metafysische uitspraken doen, aldus de verdediger van dit argument. Dit veronderstelt een helder onderscheid tussen metafysica en wetenschap (fysica in het bijzonder) dat volgens mij onhoudbaar is. De geschiedenis van wetenschap en filosofie laat zien dat beide met elkaar verweven zijn: vraagstukken die ooit als zuiver metafysisch werden gezien, zoals de aard van ruimte, tijd en oorzakelijkheid, zijn nu onderdeel van wetenschappelijk onderzoek. Metafysische discussies zijn sterk beïnvloed door empirische bevindingen en conceptuele revoluties in de wetenschappen. Een metafysica die volledig losgezongen is van onze beste wetenschappelijke kennis, kan niet voorkomen dat metafysische luchtkastelen worden gebouwd.

De relatie tussen lichaam en geest bevindt zich ook op het raakvlak tussen metafysica en wetenschap (vooral de psychologie, neurowetenschappen en fysica). Wie hier iets zinnigs over wil zeggen, zou dus rekening moeten houden met wat deze wetenschappen over de kwestie zeggen, en dan is het meest voor de hand liggende antwoord ontologische reductionisme. De dualist kan evenwel triomfantelijk blijven beweren dat allerlei dualistische opties logisch of metafysisch *niet uitgesloten* zijn, maar hij gaat hierbij volledig voorbij aan het feit dat dit niets zegt over *aannemelijkheid*. Wetenschappers en (mijns inziens) verstandige filosofen hechten niet veel waarde aan puur logische mogelijkheden en zuiver metafysische speculaties, maar aan wat waarschijnlijk ook *daadwerkelijk* het geval is²⁵. Zij zullen hun schouders ophalen bij zuiver metafysisch tegengespuiter – en terecht. Voor wie de wetenschappelijke inzichten serieus neemt, is ontologisch reductionisme de meest aannemelijke metafysische conclusie.

Is de reductionist conceptueel verward?

Er is tot slot nog een vorm van kritiek op (ontologisch) reductionisme en eliminatief materialisme die vooral *conceptueel* van aard is. Deze kritiek is het meest rigoureuus uitgewerkt door neurowetenschapper Max Bennett en filosoof Peter Hacker (Bennett & Hacker, 2003). Deze kritiek komt erop neer dat de manier waarop de reductionisten spreken over het brein en de dingen die het doet, *conceptueel verward* is omdat ze woorden voor het

²⁵ Dergelijke modale intuïties zijn interessant voor gedachte-experimenten, maar beweren dat iets logisch of metafysisch mogelijk is of juist niet uitgesloten kan worden, zegt vrijwel niets over de werkelijke wereld waarin wij leven. Wetenschappers kijken met argwaan naar dergelijke filosofie die losgezongen is van de feitelijke wereld en empirische toetsbaarheid – en terecht. Gelukkig zijn er ook filosofen die hun filosofische vaardigheden integreren met onze beste wetenschappelijke kennis op een bepaald gebied, en zo wél vooruitgang in inzicht kunnen boeken.

brein gebruiken die alleen zinvol toegepast kunnen worden op personen. Deze kritiek is echter alleen geldig als men ervan uitgaat dat woorden voor hun betekenis gebonden zijn aan de betekenis die ze hebben in het dagelijks leven. In andere discoursen dan het dagelijks leven kunnen deze woorden echter wel degelijk een andere zinvolle betekenis krijgen. Technische literatuur in allerlei vakgebieden staat hier vol mee en gebruikers kunnen daarmee over het algemeen prima zinvol communiceren, en de neurowetenschappen vormen daarop geen uitzondering. Zie verder ook de reacties op Bennett & Hacker van Paul Churchland (2005a) en Daniel Dennett (2007).

Implicaties

Als de reductionist gelijk heeft, heeft dit ook godsdienstfilosofische implicaties. Zo is een leven na de dood niet mogelijk als bewustzijn verdwijnt met de hersendood. Dit is overigens al een oude gedachte, die ook te vinden is in de Hebreeuwse Bijbel: “stof zijt gij en tot stof zult gij wederkeren” (Genesis 3:19)²⁶. Ook zijn er implicaties voor het bestaan van de ultieme Geest. Als een lichaamloze geest immers niet kan bestaan, maakt dit het bestaan van God ook onwaarschijnlijk.

Conclusie

Dualisme is intuïtief plausibel. Hierdoor *lijkt* er inderdaad een *hard problem* van het bewustzijn te bestaan, bewustzijn *lijkt* iets fundamenteel anders dan hersentoestanden, het *voelt* intuïtief anders. Daarom *lijkt the hard problem of consciousness* inderdaad een erg moeilijk probleem, en daarmee een goed argument tegen (ontologisch) reductionisme of zelfs fysicalisme.

Toch meen ik dat dit probleem slechts schijn is. Er is slechts een probleem als je *aanneemt* dat bewustzijn iets fundamenteel anders, ontologisch verschillend, is van hersenfunctie. Maar die aanname is juist wat de reductionist ter discussie stelt. Die aanname is namelijk strijdig met goed bevestigde neurowetenschappelijke, evolutionaire en natuurkundige kennis. Het enige waar die aanname op gebaseerd is, is een dualistische intuïtie. Hierboven hebben we gezien dat die intuïtie echter een illusie is, een illusie die evolutionair goed te verklaren valt. Wie deze illusie doorziet, ziet ook dat Chalmers' *hard problem* slechts een schijnprobleem is.

Daarnaast is het (ontologisch) reductionisme alleen maar sterker komen te staan door nieuwe ontwikkelingen in de neurowetenschappen omdat het de beste verklaring biedt voor de gevonden correlaties. Dualisme wordt niet verworpen omdat met zekerheid aangetoond kan worden dat het onjuist is, maar omdat het veel meer vragen en problemen oproept dan reductionisme, en reductionisme geen extra, niet-fysieke ontologie nodig heeft. De identiteitstheorie, die het bewustzijn ziet als een hersentoestand, blijft daarmee nog steeds de meest plausibele positie.

Referenties

Anderson, P. W. (1972). More is different. *Science*, 177(4047), 393-396.

Bear, M. F., Connors, B., & Paradiso, M. (2015). *Neuroscience: Exploring the brain*. LWW

²⁶ In de Bijbel zijn verschillende opvattingen te vinden over wat er na de dood gebeurt, van niets tot een schimmig bestaan in de onderwereld (beide in het Oude Testament), en ook (onder invloed van de Griekse filosofie) een voortbestaan in de hemel (alleen in het Nieuwe Testament en vooral latere theologie).

- Bechtel, W. (2012). Reducing psychology while maintaining its autonomy via mechanistic explanations. In M. Schouten & H. Looren de Jong (Eds.), *The matter of the mind: Philosophical essays on psychology, neuroscience and reduction*. John Wiley & Sons.
- Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience* (Vol. 79). Oxford: Blackwell.
- Blackmore, S., & Troscianko, E. T. (2018). *Consciousness: An Introduction*. 3rd New edition edition. Abingdon, Oxon.
- Bloom, P. (2005). *Descartes' baby: How the science of child development explains what makes us human*. Random House.
- Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe?. *Philosophical studies*, 170(3), 465-500.
- Brigandt, I., & Love, A. (2017). Reductionism in biology. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Geraadpleegd via <https://plato.stanford.edu/entries/reduction-biology/>
- Carroll, S. B. (2016). *The big picture*. Dutton
- Carruthers, G., & Schier, E. (2017). Why are we still being hornswoggled? Dissolving the hard problem of consciousness. *Topoi*, 36(1), 67-79.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of consciousness studies*, 2(3), 200-219.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford university press.
- Churchland, P. S. (2002). *Brain-wise: Studies in neurophilosophy*. MIT press.
- Churchland, P. M. (2005a). Cleansing science. *Inquiry*, 48(5), 464-477.
- Churchland, P. (2005b). Chimerical colors: Some phenomenological predictions from cognitive neuroscience. *Philosophical psychology*, 18(5), 527-560.
- Churchland, P. M. (2013). *Matter and consciousness*. MIT press.
- Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.
- Dennett, D. C. (1996). Commentary on Chalmers' Facing Backwards on the Problem of Consciousness". *Journal of Consciousness Studies*, 3(1), 4-6.
- De Regt, H., & Dooremalen, H. (2015). *Het snapgevoel: Hoe de illusie van begrip ons denken gijzelt*. Boom

Dennett, D. C. (2007). Philosophy as naive anthropology: Comment on Bennett and Hacker. In Bennett, M. R., Dennett, D., Dennett, D. C., Hacker, P., & Searle, J. (2007). *Neuroscience and philosophy: Brain, mind, and language*. Columbia University Press.

Dooremalen, H., de Regt, H., & Schouten, M. (2010). *Stof tot denken: filosofische aspecten van brein en bewustzijn*. Boom.

Duch, W. (2017). Why minds cannot be received, but are created by brains. *Scientia et Fides*, 5(2), 171-198.

Fiala, B. J. (z.d.). Toward a debunking account of the explanatory gap. Geraadpleegd via http://www.u.arizona.edu/~fiala/files/Fiala_WritingSample.pdf

Frankish, K. (2017). *Illusionism: as a Theory of Consciousness*. Andrews UK Limited.

Gazzaniga, M. S. (2018). *The Consciousness Instinct: Unraveling the mystery of how the brain makes the mind*. Farrar, Straus and Giroux.

Graziano, M. S. (2013). *Consciousness and the social brain*. Oxford University Press.

Graziano, M. S., & Webb, T. W. (2015). The attention schema theory: a mechanistic account of subjective awareness. *Frontiers in psychology*, 6, 500.

Jackson, F. (1982). "Epiphenomenal Qualia". *Philosophical Quarterly*. 32 (127): 127–136.

Klink, B.J. (2007). Stof zijt gij en tot stof zult gij wederkeren. Geraadpleegd via <http://www.deatheist.nl/downloads/StofZijtGij.pdf>

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, 93(3), 480-490.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific philosophical quarterly*, 64(4), 354-361.

Martin, M., & Augustine, K. (2015). *The myth of an afterlife: The case against life after death*. Rowman & Littlefield.

Musolino, J. (2015). *The Soul Fallacy: What Science Shows We Gain from Letting Go of Our Soul Beliefs*. Prometheus Books

Nagel, T. (1974). What is it like to be a bat?. *The philosophical review*, 83(4), 435-450.

O'Connor, T., & Wong, H. Y. (2015). Emergent properties. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Geraadpleegd via <https://plato.stanford.edu/entries/properties-emergent/>

Papineau, D. (2002). *Thinking about consciousness*. Oxford University Press.

Papineau, D. (2011). What exactly is the explanatory gap?. *Philosophia*, 39(1), 5-19.

Polák, M., & Marvan, T. (2018). Neural Correlates of Consciousness Meet the Theory of Identity. *Frontiers in psychology*, 9.

Schouten, M., & Looren de Jong, H. (2012). Mind matters: the roots of reductionism. In M. Schouten & H. Looren de Jong (Eds.), *The matter of the mind: Philosophical essays on psychology, neuroscience and reduction*. John Wiley & Sons.

Vlerick, M., & Boudry, M. (2017). Psychological closure does not entail cognitive closure. *dialectica*, 71(1), 101-115.

Wegener S.T., Jacobs M. (2011) Pain Asymbolia. In: Kreutzer J.S., DeLuca J., Caplan B. (eds) *Encyclopedia of Clinical Neuropsychology*. Springer, New York, NY