Faculteit letteren en wijsbegeerte

Anne Vangerven

(20023204)

Causaliteitsconcepten in de bewustzijnsfilosofie

Masterproef

neergelegd tot het behalen van de graad van

Master in de Wijsbegeerte

2012-2013

Promotor: Bert Leuridan

Commissarissen: Farah Focquaert
Raoul Gervais

- ✓ Word Count 1: 27.832
- ✓ Word Count 2: 28.992

Voorwoord

Sinds voor het begin van mijn opleiding wist ik dat mijn thesis 'iets' met het bewustzijn te maken zou hebben. Mijn beduimelde exemplaar van Dennetts *Consciousness Explained* staat sinds jaar en dag vol met woeste aantekeningen, maar toch zou het een ander paar mouwen worden om correct te verwoorden waar hij nu precies volgens mij zo verkeerd zat. Toen ik in de lessen wetenschapsfilosofie werd geconfronteerd met verklarings- en causaliteitsleer, maar vooral door het werk van Carl Craver te lezen, raakte ik gefascineerd door de nieuwe denkwerktuigen die hierdoor voorhanden kwamen. Vooral wat betreft Craver groeide een groot respect voor de manier waarop hij dit alles zo haarfijn en duidelijk kon afbakenen. Waarom dus niet het ene op het andere toegepast?

Ik ben echter geen logica- en nog veel minder een formulemens. Van die zaken herinner ik me alleen de frustratiehoofdpijn, en die van mijn geduldige lesgevers die waarschijnlijk even groot geweest moet zijn (wat dit betreft is een welgemeend hoeraatje voor prof. Joke Meheus zeker niet overbodig!).

Toch ben ik de uitdaging aangegaan om een thesisonderwerp te kiezen dat zich een eind weegs buiten mijn *'comfort zone'* bevond, en mocht ik daar de vruchten van plukken: het was dan ook een heel pak spannender dan het anders was geweest!

Een universitaire opleiding combineren met een job, ook al is het maar halftijds, is niet gemakkelijk. Ik wil dan ook mijn fantastische collega's bedanken voor de vele wissels en het begrip als ik weer eens met mijn hoofd in de wolken liep. Daarnaast kan ik in het diepste van mijn hart nooit genoeg dankbaarheid bijeen sprokkelen voor Piet en Sharon, zonder wie alles er helemaal anders had uitgezien, en natuurlijk mijn promotor, Bert Leuridan, voor zijn heldere uitleg van allerlei logische en metafysische concepten. Ten slotte mijn vriend, Frederik, voor zijn eindeloze geduld en humor: "Je had van in het begin al gelijk..."

Anne Vangerven

6 augustus 2013

Inhoud

1. Inleiding

Het probleem van mentale veroorzaking was oorspronkelijk een vraagstuk over het lichaam en de ziel. Het lijkt een universeel menselijk gegeven om aan onszelf zoiets toe te schrijven als een geest, een *chi*, een essentie die niet gelijk is aan wat we voor de rest zijn – een bouwsel van vlees en botjes. Dat bouwsel is maar omhulsel, die ziel, dat zijn we *echt*. Maar toch is dat lichaam toch ook iets dat we nodig hebben om ons door de wereld te bewegen en de voorwerpen om ons heen te manipuleren. Als we praten, komen onze gedachten er langs onze mond uit. Maar hoe gaat dat precies?
Sinds de dagen van toga's en amfora's – en waarschijnlijk daarvoor nog – hebben filosofen geprobeerd om die link tussen lichaam en geest te verklaren, om dat wat ze eerst van elkaar los hebben gedacht weer met elkaar te verenigen. Want de geest, zo ging het immers over de hele wereld, was eeuwig en onstoffelijk, en het lichaam bestond uit materie. Hoe kan het dan, dat zo'n gedachte er toch in slaagt om onze materiële voeten te bewegen?

Vandaag lachen we om dergelijke mysteries. De ziel, dat is een fabeltje. Alles wat bestaat valt onder de wetten van de fysica. Toch raken we het niet eens over wat er precies gaande is in ons hoofd. De geheimen van het brein zijn nog lang niet ontrafeld, en hoe hard we ook ons best doen om het tegendeel te geloven, toch blijft er dat gevoel dat er in ons hoofd een klein mannetje aan het roer zit van ons lichaam. Dezelfde problemen lijken de kop weer op te steken, maar in andere gedaantes. Als we zelf alleen nog robotjes zijn die zich plooien naar de natuurwetten, hoe kunnen we dan ooit nog iemand verantwoordelijk houden voor zijn handelingen? Zelfs fervente fysicalisten als Daniel Dennett, voor wie het bewustzijn volledig te reduceren valt tot de elektrochemische processen in ons brein, willen deze bittere pil niet slikken. Ze houden spartelend vol dat er toch iets bestaat als 'vrije wil'. Dat 'ik' uit "Ik denk, dus ik ben" is niet iets dat zelfs zij zomaar willen opgeven, of dat zomaar opgaat in een wolkje van logica. De reductionist neemt de taak op zich om alle tussenstapjes van atoom naar subjectiviteit te expliciteren en te verklaren.

Deze positie wordt echter niet door iedereen aangehangen. Hoewel er nog weinigen zijn die zichzelf dualist noemen in de zin als Descartes er één was (ik ken er in ieder geval geen), blijven sommige hedendaagse filosofen volhouden dat er eigenschappen zijn van het bewustzijn die onmogelijk te reduceren zijn tot het fysieke. Het gaat om fenomenologische eigenschappen, subjectiviteit, of ook wel qualia: die onbeschrijfelijke eigenschappen van onze ervaring die ervoor zorgen dat we er zo rotsvast van overtuigd zijn dat we thuis zijn in onze beleefwereld. We *voelen* die ervaringen in al hun intensiteit, we maken ze mee, het licht is *aan*.

De bewustzijntheorieën die hieruit voortvloeien, worden door hun bedenkers biologisch naturalisme (John Searle), anomaal monisme (Donald Davidson), niet-reductionistisch fysicalisme (Jaegwon Kim) of simpelweg eigenschapsdualisme (David Chalmers) genoemd. Elk op hun eigen manier proberen ze oplossingen te vinden voor een probleem dat zich volgens hen nog steeds aan ons opdringt: het mysterie van het bewustzijn.

Het probleem van mentale veroorzaking is echter niet alleen een ontologische vraag – wat het bewustzijn precies *is* – maar ook een vraag naar de werking ervan. We ontwaren immers niet alleen causale relaties tussen onze gedachten en ons lichaam, maar ook tussen gedachten onderling. Als we redeneren, laten we de ene gedachte de andere beïnvloeden, zo zegt Jaegwon Kim. Herinneren is een causaal proces waarbij we ervaringen die we fysiek hebben opgeslagen, weer ophalen in de vorm van een geloof. (Kim, 1998, p. 31) Als we willen volhouden dat er zoiets bestaat als een bewustzijn, zullen we volgens hem dus moeten aantonen hoe het in zo'n complexe causale interactie kan staan met de buitenwereld. De positie die volhoudt dat deze interactie ook echt plaatsvindt, noemen we interactionisme.

Niet iedereen is het daar echter mee eens. Filosofen die denken dat het bewustzijn een epifenomeen is geloven enkel dat mentale gebeurtenissen het gevolg zijn van fysieke gebeurtenissen, maar zelf geen gevolgen (we spreken over causale invloed, maar we zullen later zien dat we moeten oppassen met deze terminologie) kunnen hebben, *zelfs niet op andere mentale gebeurtenissen*. (Yoo, 2007) Het bewustzijn

dobbert in deze visie alleen een beetje mee op het fysieke, maar speelt verder geen enkele rol.

Een laatste positie, parallellisme, vindt sinds de tijd van Leibniz en Malebranche nog weinig aanhang. Zij stelden dat het fysieke en het mentale geen causale invloed hadden op elkaar, maar gewoonweg perfect met elkaar gelijk liepen. Hoe dat dan kon? Daar zat God voor iets tussen. Volgens Leibniz had die helemaal aan het begin de synchroniciteit vastgelegd, een beetje zoals de klapper aan het begin van een filmopname, waarna die voor altijd in perfecte harmonie bleef. In de visie van Malebranche (occasionalisme), had God echter veel meer werk te doen: telkens als we zin hadden om onze arm op te tillen, stond God immers voor ons klaar om onze arm omhoog te doen gaan. (Yoo, 2007) Omdat we vandaag nog weinig met God te maken hebben, kunnen we ons niet meer op hem beroepen om de coördinatie tussen lichaam en geest te verklaren. We gaan dit probleem nu met andere middelen te lijf.

Niet alleen over de aard van het bewustzijn en de relatie met het lichaam kunnen we een grote variatie van filosofische posities innemen. Sedert de dagen van David Hume beschikken we ook over steeds meer theorieën over causaliteit. Deze hebben niet alleen een metafysisch nut. Vandaag gebruiken we ze vaak in een wetenschappelijke context, wanneer we willen weten wat we nu precies mogen afleiden uit een hoeveelheid van experimentele data. Wat kunnen de mogelijke gevolgen zijn van een bepaald fenomeen? Hoe kunnen we een onwenselijk gevolg in de toekomst vermijden? Is één fenomeen het gevolg van het andere, zoals warmte en smeltend ijs, of zijn ze samen het gevolg van iets anders, zoals vieze gele vingers en een hoger risico op longkanker? Om een goed antwoord te kunnen geven op dit soort vragen waar wetenschappers zich dag in dag uit mee bezig houden, zullen we dus goed moeten weten wat we precies bedoelen als we het hebben over oorzaak en gevolg.

In de volgende hoofdstukken neem ik enkele bewustzijnsfilosofen onder de loep en zoek uit welk causaliteitsconcept zij er op nahouden. Niet iedere filosoof maakt dit immers even expliciet. Soms zullen ze overtuigd een positie kiezen, anderen

schipperen dan weer tussen verschillende opvattingen, of doen alsof wat we verstaan onder causaliteit helemaal niet zo omstreden is. Toch denk ik dat het intussen duidelijk is geworden dat de positie die een filosoof al dan niet kiest in het causaliteitsdebat, een belangrijke invloed zal hebben op de rest van zijn filosofie. Als we een theorie hebben over hoe het bewustzijn *werkt*, zijn we immers al een hele stap dichter bij de verklaring van wat het *is*.

De filosofen die ik bespreek bevinden zich in twee kampen: dualisme en fysicalisme. Ze komen dus elk tot zeer uiteenlopende conclusies met betrekking tot de aard van het bewustzijn. Maar ook binnen hetzelfde kamp kunnen de benaderingen al heel erg van elkaar verschillen. De selectie die ik heb gemaakt, kan hier natuurlijk geen recht aan doen. Wel zijn de denkers die ik heb gekozen zeker niet de minste, en kunnen ze als representatief worden aanzien voor een bepaalde benaderingswijze of stroming. Ik situeer elke denker even kort en vat de belangrijkste punten van zijn filosofie samen. Daarna ga ik in hun werk op zoek naar aanwijzingen voor hoe zij causaliteit definiëren, eventueel verwijzingen naar bestaande causaliteitstheorieën, en toepassingen hiervan. De hypothese die ik intussen bij de hand houd, is dat het causaliteitsconcept dat de besproken filosofen er al dan niet op nahouden een stempel zal drukken op hun argumentatie – hoe zij het probleem van mentale veroorzaking aanpakken – maar ook op hun uiteindelijke oplossing voor dit probleem.

Daniel Dennett en Jaegwon Kim werden de vertegenwoordigers van een fysicalistische benadering. David Chalmers staat in voor het dualisme. Het deel over Descartes en prinses Elizabeth dient als warmlopertje, waarin het probleem van mentale veroorzaking zoals dat zich binnen het dualisme stelt nog eens duidelijk wordt omschreven. Bovendien geef ik in dit deel ook al een illustratie van hoe een specifiek causaliteitsbegrip dit probleem *zou kunnen* oplossen. Bij sommige filosofen ga ik ook in op problemen als reductie en de vrije wil (dit laatste alleen bij Dennett), aangezien deze ook automatisch deel uitmaken van de bewustzijnsfilosofie, en opnieuw niet los te denken zijn van een bepaalde visie op causaliteit.

2. Causaliteits- en verklaringstheorieën

In de hierop volgende hoofdstukken zal ik enkele bewustzijnsfilosofen nader bekijken, uitzoeken wat hun is visie op causaliteit en verklaring, en proberen te achterhalen wat hiervan de gevolgen zijn voor hun visie op mentale veroorzaking en het bewustzijn zelf. Hiervoor zullen zowel ikzelf als deze filosofen zich beroepen op de diverse causaliteits- en verklaringstheorieën die sinds Hume zijn uitgedacht.[1] Om het verwijzen verderop in de tekst gemakkelijker te maken, volgt hierna een kort overzicht van de belangrijkste theorieën en begrippen. Ik zal me hiervoor baseren op *Causation & Explanation* (Psillos, 2002), 'Causation' (Weber en De Vreese, 2012) en 'Scientific Explanation' (Weber e.a., 2012). Mijn doel is hierbij zeker niet volledig te zijn. Als ik zaken weglaat, dan is dit enkel omdat ze in de hoofdstukken die volgen ook niet verder aan bod zullen komen.

2.1 Causaliteit als regelmatigheid

2.1.1 De regulariteitstheorie van David Hume

In *An Enquiry concerning Human Understanding* merkte David Hume op dat hij in zijn leven nog nooit het fenomeen dat we causaliteit noemen, had kunnen observeren. Het enige dat hij om zich heen zag, waren gebeurtenissen, eventueel gevolgd door andere gebeurtenissen. Maar het causale verband, '*the connexion*', dat had hij nog nooit waargenomen:

> *"When we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effects to the cause, and renders the one an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other. The impulse of one billiard-ball is attended with motion in the second. This is the whole that appears to the* outward *senses. The mind feels no sentiment or* inward *impression from this succession of objects: Consequently, there is not, in any single, particular*

[1] Natuurlijk wil ik niet doen alsof de metafysica van Aristoteles nooit heeft bestaan, maar ik denk diens relevatie vandaag toch wat is afgenomen.

> *instance of cause and effect, any thing which can suggest the idea of power or necessary connexion."* (Hume, [1748] 2007, p. 46, cursief in origineel)

Hieruit trok hij het besluit dat er ook geen noodzakelijk verband was tussen de twee gebeurtenissen, alleen een regelmatigheid. Daarom wordt zijn theorie over causaliteit een regulariteitstheorie genoemd. Causaliteit wordt bij Hume gereduceerd tot niet-causale feiten (namelijk dat de fenomenen ruimtelijk aangrenzend en opeenvolgend zijn):

> *c causes e iff*
> *c is spatially contiguous to e*
> *e succeeds c in time; and*
> *all events of type C (i.e., events that are like c) are regularly followed by (or are constantly conjoined with) events of type E (i.e. events like e)*
>
> (Psillos, 2002, p. 19)

Een gevolg hiervan is dat de notie van causaliteit een groot deel van haar kracht verliest. Als we spreken over een regelmatigheid, ontbreekt het idee van noodzakelijkheid dat we vaak wel met causaliteit associëren. Humes theorie resulteert dus in een zwak causaliteitsconcept.

2.1.2 Hempels deductief-nomologische verklaringsmodel

Zoals de titel zegt, is het model van Carl Hempel in de eerste plaats een verklarings- en geen causaliteitstheorie. Schatplichtig aan het project van de logische empiristen om causaliteit van haar mystiek te ontdoen, probeerden Hempel en zijn volgelingen om het concept van causaliteit te analyseren aan de hand van het concept van verklaring. (Psillos, 2002, p. 10)

Hempel stelt dat verklaringen eigenlijk argumentaties zijn, die we altijd kunnen vatten in de vorm van een logisch syllogisme (vandaar deductief), waarvan de eerste premisse een universele wet poneert (vandaar nomologisch).

Formeel gesteld, wordt dit:

> *"The ordered couple (L,C) constitutes a potential explanans for the singular sentence E if and only if*

> *(1) L is a purely universal sentence and C is a singular sentence,*
>
> *(2) E is deductively derivable from the conjunction L&C, and*
>
> *(3) E is not deductively derivable from C alone.*
>
> *(L,C) is a true explanans for E if and only if (L,C) is a potential explanans for E and both L and C are true."*
>
> (Weber e.a., 2012)

Volgens Hempel is het verklaren van een fenomeen gelijk aan aantonen dat het te verwachten viel, gezien de overkoepelende wet. De werkelijkheid begrijpen is dan niet meer dan te weten wat men kan verwachten. (Weber e.a., 2012)

Een probleem met dit model is dat het moeilijk bestand is tegen accidentele generalisaties. Zo kunnen we bijvoorbeeld uit "Alle appels in de mand zijn rijp," en "Deze appel ligt in de mand," makkelijk afleiden: "Deze appel is rijp." Toch is het geen natuurwet dat de appels die in deze mand liggen rijp moeten zijn, het is gewoon toeval. Om een behoorlijk DN-verklaringsmodel te ontwikkelen, zullen we dus ook een poging moeten doen om te beschrijven wat nu de werkelijke natuurwetten zijn, zo stelt Stathis Psillos (2002, p. 11).

2.1.3 De regulariteitstheorie van John Mackie

Volgens John Mackie zijn er twee problemen met klassieke regulariteitstheorieën. Het eerste probleem is dat niet alles dat we een oorzaak noemen, *voldoende grond* is voor haar effect. Een voorbeeld dat hij geeft is dat we zeggen dat een brand veroorzaakt werd door een kortsluiting, terwijl dat op zich niet voldoende is: we hebben ook zuurstof en brandbaar materiaal nodig. Het tweede probleem is dat opeenvolging in tijd geen goed manier is om de asymmetrie van oorzakelijkheid[2] te garanderen: soms noemen we iets een oorzaak van een gelijktijdige gebeurtenis. (Weber en De Vreese, 2012, p. 11)

Om deze problemen op te lossen, introduceert Mackie de term *INUS-conditie*:

> *"An INUS condition is an 'insufficient but non-redundant part of an unnecessary but sufficient condition'."* (Weber en De Vreese, 2012, p. 11)

[2] Dit wil zeggen dat effecten het gevolg zijn van hun oorzaken, en niet omgekeerd.

Een oorzaak hoeft volgens deze theorie niet op zich voldoende te zijn (zoals de kortsluiting) voor het gevolg, maar moet wel deel uitmaken van een set fenomenen die er samen voldoende voor zijn. In deze set mag het fenomeen ook niet overbodig zijn (denk aan het feit dat het dinsdag is in het verhaal van de kortsluiting en de brand).

Dit is de formele weergave van Mackie's theorie:

> *"Let c and e be two possible events. Then we call c a cause of e if we think that there is a time t, a time t' (t'<t) and a set of real events A for which:*
> *(1) e occurred at time t,*
> *(2) c occurred at time t',*
> *(3) the occurrence of c at t' is together with the events in A is a sufficient condition for the occurrence of e at t.*
> *(4) the events in A in themselves are not a sufficient condition e at t."*
>
> (Mackie, 1974, geciteerd door Weber en De Vreese, 2012)

2.2 Tegenfeitelijke theorieën

In de eerste *Enquiry* voegt David Hume, nadat hij zijn eerste causaliteitsdefinitie heeft gegeven (causaliteit als regelmatigheid), de volgende opmerking toe: *"Or in other words, where, if the first object had not been, the second never had existed."* (Hume, 1748, geciteerd in Psillos, 2002, p. 82)

Dit is een nieuwe definitie van causaliteit, zegt Psillos, niet in termen van causale regelmatigheden, maar in termen van *tegenfeitelijke afhankelijkheid*. (Psillos, 2002, p. 82)

Een tegenfeitelijke bewering heeft de vorm "Als *a* zich niet had voorgedaan, zou *b* zich niet hebben voorgedaan." We merken meteen dat als we causaliteit op deze manier begrijpen, de twee problemen die Mackie hierboven aanstipte, zich niet meer voordoen. We krijgen echter een nieuw probleem: overdeterminatie.

In een bekend voorbeeld gooien Billy en Suzy elk met een steentje naar een ruit. Toevallig is het het steentje van Billy dat het eerste met de ruit in contact komt, en de ruit breekt. Een fractie van een seconde later vliegt het steentje van Suzy naar binnen.

Als we hier een tegenfeitelijke bewering van maken, blijkt dat het steentje van Billy niet in aanmerking komt als oorzaak voor het breken van de ruit: als zijn steentje er immers niet was geweest, was de ruit nog steeds gebroken.

Een oplossing van Mackie voor dit probleem is dat de ruit niet *precies op die manier* was gebroken door het steentje van Suzy, en we dus met een ander gevolg te maken zouden hebben als haar steentje de ruit brak. Voor het *precies op deze manier* breken van het raam, was het steentje van Billy wél nodig. (geparafraseerd uit Psillos, 2002, die op p. 85 e.v. een ander voorbeeld gebruikt)

Tenslotte speelt bij tegenfeitelijke theorieën de notie van *mogelijke werelden* een rol. Tegenfeitelijke beweringen beschrijven immers geen 'volledig objectieve realiteit' (Mackie, 1974, geciteerd door Psillos, 2002, p. 83), maar een andere 'mogelijke wereld' waarin de zaken net iets anders zijn gelopen dan in de onze (er is bijvoorbeeld geen Billy, of hij heeft geen steentje gegooid). Voor ons tegenfeitelijk gedachte-experiment, zullen we eerst moeten uitmaken wat we 'als bagage' mee zullen nemen naar deze hypothetische wereld waarover we onze vraag zullen stellen. Voor de bewering "als we deze lucifer niet hadden aangestreken, was ze niet ontvlamd", kunnen we ons bijvoorbeeld beroepen op de bewering "lucifers die worden aangestreken, ontvlammen". Deze laatste bewering is immers voldoende zeker in onze wereld om ze in onze bagage te stoppen, aldus Mackie. Die bagage vormt dan samen het inductieve bewijs. (Mackie, 1974, geciteerd door Psillos, 2002, p. 83)

De mogelijke wereld waarover we de tegenfeitelijke vraag stellen, moet dus voldoende lijken op de onze om nog relevant te zijn. We kunnen ons immers een wereld voorstellen waarin alles onder water staat. In een dergelijke wereld gaat onze kennis over het gedrag van lucifers niet langer op. Deze wereld heeft echter nog weinig te maken met wat we willen onderzoeken.

2.2.1 De tegenfeitelijke theorie van David Lewis

De bekendste tegenfeitelijke theorie werd geformuleerd door David Lewis. Hij drukt causale afhankelijkheid als volgt uit:

> *"Let c and e be two distinct possible particular events. Then e depends causally on c iff the family O(e), ~O(e) depends counterfactually on the family O(c),*

> *~O(c). As we say it: whether e occurs or not depends on whether c occurs or not. The dependence consists in the truth of two counterfactuals: O(c) □→ O(e) and ~O(c) □→ ~O(e). There are two cases. If c and e do not actually occur, then the second counterfactual is automatically true because its antecedent and consequent are true: so e depends causally on c iff the first counterfactual holds. That is, iff e would have occurred if c had occurred. But if c and e are actual events, then it is the first counterfactual that is automatically true. Then e depends causally on c iff, if c had not been, e had never existed."*
>
> (Lewis, 1986, geciteerd door Weber en De Vreese, 2012)

Causale afhankelijkheid is wat we volgens David Lewis nodig hebben om causaliteit te definiëren. Omgekeerd impliceert causaliteit geen causale afhankelijkheid, omwille van de mogelijkheid van causale ketens:

> "*Let c, d, e, . . . be a finite sequence of actual particular events such that d depends causally on c, e on d, and so on throughout. Then this sequence is a causal chain. Finally, one event is a cause of another iff there exists a causal chain leading from the first to the second.*" (Lewis, 1973a)

2.3 De mechanistische theorieën van Salmon en Dowe

Een proces- (of mechanistische (Psillos, 2002, p. 107 e.v.)) theorie definieert causaliteit in termen van processen en interactie (Weber en De Vreese, 2012, p. 16).

Een fervente verdediger van dit idee van causaliteit als een mechanisme is Wesley Salmon, die stelt dat hij "Humes uitdaging ernstig wil nemen" (Salmon, 1997, geciteerd door Psillos, 2002, p. 110), en er dus op uit is een theorie te vinden die recht doet aan de verbinding tussen een oorzaak en haar gevolg. Om dit te bereiken, moeten we anders gaan denken over wat de bouwstenen (of relata) vormen van causale werking. In plaats van over gebeurtenissen, spreken we beter over processen, aldus Salmon. (Psillos, 2002, p. 111) Deze processen zijn "precies die verbindingen waar Hume naar op zoek was [...] (hoewel het geen *noodzakelijke* verbindingen hoeven te zijn)." (Salmon, 1997, geciteerd door Psillos, 2002, p. 111) Dit laatste is een belangrijk verschil met de theorieën die de nadruk juist wel op noodzakelijkheid of wetmatigheid leggen

(zie bvb. Hempel). Het heeft als voordeel dat we ook wanneer iets in het geheel niet te verwachten viel of gewoon toevallig gebeurde, hier nog steeds een causale relatie in kunnen zien optreden.

Een proces is volgens Salmon een 'wereldlijn' (*'world line'*) van een object: een collectie van punten in een tijdruimtediagram dat de geschiedenis van een object beschrijft. (Weber en De Vreese, 2012, p. 16) Een proces is continu, terwijl een gebeurtenis gelokaliseerd is in tijd en ruimte (op een dergelijk diagram zou het slechts een punt voorstellen, terwijl een proces een oneindige lijn vormt). (Psillos, 2002, p. 111) Interactie is wat optreedt wanneer twee causale processen elkaar kruisen. Niet alle interacties zijn echter causaal, omdat niet alle processen dat zijn. Causale processen onderscheiden zich doordat ze tijdens de interactie iets (*'a* mark') overdragen op het andere proces dat ze hebben gekruist, *dat ook na de interactie bij het gekruiste proces blijft horen*. Het prototype van een causale interactie is het geval waarin twee biljartballen elkaar raken. De overdracht (*'mark transmission'*) die in dit geval plaatsvindt is die van de aandrijving van de ene bal op de andere. Ook na de botsing blijft een deel van de richting en snelheid van de ene bal bij de andere bal horen. (Ze dragen er dus de blijvende gevolgen van, zou je kunnen zeggen.) Er bestaan ook wat Salmon 'pseudoprocessen' (*'pseudo-processes'*) noemt. Een voorbeeld hiervan is de schaduw van een rijdende auto. Deze vervormt wel als gevolg van de interactie met bijvoorbeeld een paal, maar neemt daarna zijn gewone vorm weer aan. Alleen causale processen zijn dus in staat tot overdracht van wat Salmon *'a mark'* noemt. (Psillos, 2002, p. 112)

Door de omschrijving van wat er precies wordt overgedragen (*'a* mark') opzettelijk vaag te laten, laat Salmon ruimte open voor een heel aantal soorten van causaliteit die voor andere theorieën problematisch zijn. Weber en De Vreese geven een voorbeeld waarin de toehoorders van een lezing de gevolgen ondervinden van het vertellen van de spreker: ze leren namelijk iets bij. (Weber en De Vreese, 2012, p. 17) Dit is niet alleen een voorbeeld van causaliteit op afstand, maar ook van mentale veroorzaking: de gedachten en ideeën van de spreker beïnvloeden via zijn handelingen (het spreken) de gedachten en ideeën van de toehoorders.

Weber en De Vreese benadrukken in dit geval dat het belangrijk is om over een duidelijk referentiekader te beschikken: zowel de ruimtelijke als de tijdsschaal zijn van belang als we een goede analyse willen maken van dergelijke voorbeelden. (Weber en De Vreese, 2012, p. 17 - 18)

Phil Dowes '*conserved quantity*'-theorie kan gezien worden als een uitbreiding op die van Salmon, waarin wel wordt geprobeerd om datgene wat wordt overgedragen, precies te definiëren:

> *"CQ1: A* causal process *is a world line of an object that possesses a conserved quantity*
> *CQ2: A* causal interaction *is an intersection of world lines that involves exchange of conserved quantity."*
>
> (Dowe, 2000, geciteerd door Weber en De Vreese, 2012, p. 18)

En verder:

> *"A* conserved quantity *is any quantity that is governed by a conservation law, and current scientific theory is our best guide to what these are. For instance, we have good reasons to believe that mass-energy, linear momentum, and charge are conserved quantities. (p. 91)*
> *An* exchange *occurs when at least one incoming, and at least one outgoing process undergoes a change in the value of the conserved quantity [.] (p. 92)*
> 'Possesses' *is to be understood in the sense of 'instantiates'. (p. 92)"*
>
> (Dowe, 2000, geciteerd door Weber en De Vreese, 2012, p. 19)

Het gevolg hiervan is dat er in de theorie van Phil Dowe geen ruimte is voor oorzakelijkheid buiten het fysieke domein (zoals het voorbeeld hierboven van de lezing). (Weber en De Vreese, 2012, p. 19)

2.4 De interventietheorie van Jim Woodward

In zijn hoofdwerk *Making Things Happen* (2003) verdedigt Jim Woodward een manipulationistische (*manipulationist*) of interventionistische (*interventionist*) causaliteitstheorie. Causale relaties (in tegenstelling tot louter correlaties) zijn relaties

die potentieel aanwendbaar zijn voor manipulatie en controle. De vraag of *P* de oorzaak is van *S*, wordt dan gelijkgesteld met de vraag of het mogelijk is om *P* op een bepaalde manier te manipuleren en daardoor *S* te veranderen. (Woodward in Hohwy en Kallestrup, 2008, p. 219-220) Die welbepaalde manipulatie waar Woodward het over heeft, noemt hij een *interventie*.

Belangrijk voor de theorie van Woodward is dat de causaliteitstheorie die hij beschrijft, betrekking heeft op *variabelen* (in tegenstelling tot bvb. gebeurtenissen of objecten). Hoewel dit niet courant is, heeft het als voordeel dat zijn theorie zeer bruikbaar is binnen een wetenschappelijke context, waar hoe dan ook al met dit concept wordt gewerkt. Bovendien is zijn idee van een interventie zeer goed te vergelijken met dat van een gecontroleerd experiment. Variabelen laten toe om met drempelwaarden te werken, en maken het makkelijk om causale beweringen uit te drukken in een formeel systeem. Een bijkomend voordeel is dat de waarde van een variabele ook gelijk kan zijn aan nul, wat als gevolg heeft dat we binnen deze theorie ook over afwezigheid als oorzaak kunnen spreken (als het ontbreken van een stimulus dus tot een gevolg leidt), wat binnen de meeste andere theorieën problematisch is. De volgende samenvatting van Woodwards causaliteitsconcept komt uit Carl Cravers *Explaining the Brain* (Cravers theorie is zelf grotendeels gebaseerd op die van Woodward):

We spreken over een causale relatie tussen *X* en *Y* wanneer we door interventie *I* op *X* (en enkel op *X*), de waarde *Y* kunnen veranderen.

Dit vooronderstelt dat:

(I1) *Y* niet via directe weg gewijzigd wordt door *I*

(I2) *I* geen causale tussenvariabele *S* tussen *X* en *Y* wijzigt, tenzij door de waarde van *X* te wijzigen

(I3) *I* niet gecorreleerd is met een andere variabele *M* die de oorzaak is van *Y*

(I4) *I* zich als een 'schakelaar' (*'switch')* gedraagt die de waarde van *X* wijzigt afgezien van de andere oorzaken van *X*

(Craver, 2009, p. 96)

Deze relaties zijn stabiel onder de relevante omstandigheden, maar hoeven niet universeel te zijn, wat een antwoord biedt op de problemen van fragiliteit en contingentie (Woodward spreekt zelf over *invariantie* (*invariance*)).[3] Belangrijk is ook dat een relatie enkel potentieel aanwendbaar moet zijn voor manipulatie, ook al zijn we (nog) niet in staat om dit ook werkelijk te testen. Het gaat erom dat we de relevantie van een variabele kunnen afleiden uit wat het effect zou zijn van haar manipulatie. (Craver, 2009, p. 99)

[3] Dit wil zeggen dat ook Woodward in zijn theorie kan spreken over causaliteit wanneer het geen fenomeen betreft dat wetmatig, of zelfs maar regelmatig verloopt, net zoals Salmon dat kan.

3. Dualisme

3.1 Descartes en prinses Elizabeth

Descartes is zeker niet de eerste die zichzelf geconfronteerd zag met het probleem van mentale veroorzaking: zowel in de *Phaedo* van Plato als in Aristoteles' *De Anima* kunnen we verwijzingen terugvinden naar gelijkaardige problemen, en pogingen om ze op te lossen. Evenmin was Descartes de eerste dualist, aangezien het idee van een onsterfelijke, want onstoffelijke ziel sinds mensenheugenis ingebakken was in tal van religies en filosofieën over de hele wereld. Toch is zijn naam welhaast synoniem geworden voor het dualisme dat hij zo bekend maakte, wat hem tot een goed startpunt maakt voor een systematische benadering van de problemen die optreden bij een dualistische visie op de verhouding tussen lichaam en geest. Het eerste probleem dat we hierbij ontmoeten, namelijk dat van locatie, vloeit immers rechtsreeks voort uit Descartes' metafysica.

In een belangrijke passage uit de zesde meditatie benadrukt Descartes dat, hoewel hij er zeker van is dat hij beschikt over een lichaam waarmee hij bovendien zeer nauw verbonden is, hijzelf zo radicaal verschillend is van zijn lichaam dat hij er zeker van onderscheiden is, en het zelfs niet nodig heeft om te bestaan. (Descartes, 1634, §9)
Dat 'ik' uit Descartes' beroemde 'ik denk dus ik ben', is volgens Descartes in de eerste plaats iets dat denkt, aangezien dat het enige is waar we echt zeker van kunnen zijn. Hij noemt het de *res cogitans* (letterlijk: de denkende substantie). Ons lichaam echter, behoort tot de materiële substantie, of *res extensa* (de uitgebreide substantie). Terwijl de primaire eigenschap van de *res extensa* is dat ze een fysieke plaats inneemt en deelbaar is (uitgebreidheid), is dit volgens Descartes uitgesloten voor de *res cogitans*. De denkende substantie heeft geen uitgebreidheid. Hierin zit volgens Descartes het radicale verschil tussen deze beide substanties, en daarom noemen we hem ook een substantiedualist.
Toch stelt Descartes onmiddellijk dat we wel zeer intiem met ons lichaam verbonden zijn: we zitten er niet slechts in "zoals een stuurman in een schip". Integendeel, we zijn er zodanig mee vermengd dat lichaam en geest een soort van eenheid vormen. Anders zouden we ook niet in staat zijn om pijn te voelen als ons lichaam gekwetst is, maar

zouden we het alleen kunnen afleiden zoals een stuurman schade vaststelt aan zijn schip. (Descartes, 1634, §13)

In deze meditatie duidt Descartes ook meteen al het brein aan als de plek waar de link tussen lichaam en geest wordt gemaakt:

> "[...] als ik pijn voel in de voet, leert de fysica me dat dit gevoel overgebracht wordt door middel van de zenuwen die verspreid zitten in de voet, en als koorden van daar naar het brein lopen. Wanneer ze samentrekken in de voet, trekken ze tezelfdertijd de delen van het brein samen van waaruit ze vertrekken, en lokken er een bepaalde beweging uit *waarvan de natuur heeft bepaald dat ze pijn doet voelen in de geest*, alsof ze in de voet was." (§21, mijn vertaling en cursief)[4]

Om de geest te bereiken moet een sensatie dus niet alleen via het brein gaan, ze doet dit volgens Descartes ook door middel van beweging. Dit is ook logisch, aangezien Descartes de grondlegger was van een mechanistische visie op levende wezens. Dieren waren volgens hem een soort van automaten, die min of meer op dezelfde manier in beweging werden gehouden als de diverse wonderbaarlijke machines die in zijn tijd populair werden. (Galilei's *Le Meccaniche* werd gepubliceerd kort na Descartes' geboorte.) Ook ons lichaam zag hij als een ingewikkelde machine: aan het begin van zijn postuum gepubliceerde *De Homine* stelt hij dat we de beweging van ons lichaam op dezelfde manier kunnen verklaren als die van klokken, fonteinen of windmolens. (Gaukroger, 2004) Hoewel hij bleef vasthouden aan het antieke idee van de verschillende lichaamssappen die met hun diverse effecten in evenwicht gehouden moesten worden, was Descartes' visie op het lichaam gebaseerd op een eenvoudige mechanica waarin tandwielen, hefbomen en gewichten de hoofdrol spelen.

Maar het verhaal is nog niet compleet:

[4] « […] quand je ressens de la douleur au pied, la physique m'apprend que ce sentiment se communique par le moyen des nerfs dispersés dans le pied, qui se trouvant étendus comme des cordes depuis là jusqu'au cerveau, lorsqu'ils sont tirés dans le pied, tirent aussi en même temps l'endroit du cerveau d'ou ils viennent et auquel ils aboutissent et y excitent un certain mouvement, que la nature a institué pour faire sentir de la douleur à l'esprit, comme si cette douleur était dans le pied. » (Descartes [1634] 2009, p. 206)

> "Dus, [...], wanneer de zenuwen in de voet sterk beroerd worden, [...] geeft hun beweging een teken aan de geest dat hem pijn doet voelen alsof die zich in de voet bevond, waardoor de geest wordt aangespoord om zijn absolute best te doen om de oorzaak [van het gevaar voor de voet] weg te nemen." (§22, mijn vertaling en cursief)[5]

Het lichaam kan dus niet alleen een effect sorteren in de geest, maar de geest kan ook een effect hebben op het lichaam. Sterker nog, zonder de geest om akte te nemen van diverse sensaties zoals honger, dorst of pijn en naar behoren te handelen, kan het lichaam zichzelf niet in stand houden. Hoe de geest dit precies allemaal klaarspeelt, legt Descartes in zijn *Meditaties* niet meer uit. Het voornaamste is, dat de geest dit doet. Deze visie op de verhouding tussen lichaam en geest waarbij de oorzakelijkheid in twee richtingen werkt, noemen we interactionisme.

Vandaag de dag blijkt interactionisme nog steeds de meest intuïtieve manier om het verband tussen lichaam en geest te denken. Binnen de volkspsychologie (*folk* psychology) is het ook zeker de meest populaire. Ze hangt nauw samen met onze notie van 'vrije wil' en bij uitbreiding van verantwoordelijkheid. Voor deze concepten is het immers van belang dat we in staat zijn om met onze geest te controleren wat ons lichaam doet. Van zodra we echter even dieper nadenken over dit interactionisme dat, als het even kan, toch onze voorkeur wegdraagt, blijkt al dat het een aantal prangende vragen oproept. Dit was in Descartes' tijd al zo. Hij kreeg ze gepresenteerd in een brief van prinses Elizabeth van Bohemen, die ze verstuurde in mei 1643, een jaar nadat Descartes zijn *Meditaties* reeds opnieuw had uitgebracht met toevoeging van een set bezwaren van andere filosofen, en zijn eigen antwoorden daarop:

> "Ik smeek u me te vertellen hoe de ziel van een mens de sappen van het lichaam zo kan bepalen dat het gewilde handelingen uitvoert (aangezien hij enkel een denkende substantie is). Want het lijkt dat elke bepaling van beweging gebeurt door een aandrijving van dat wat beweegt, volgens de

[5] "Ainsi, par exemple, lorsque les nerfs qui sont dans le pied sont remués fortement, et plus qu'à l'ordinaire, leur mouvement [...] fait une impression à l'esprit qui lui fait sentir quelque chose, à savoir de la douleur, comme étant dans le pied, par laquelle l'esprit est averti et excité à faire son possible pour en chasser la cause, comme très dangereuse et nuisible au pied. » (Descartes [1634] 2009, p. 207)

> manier waarop het geduwd wordt door dat waardoor het bewogen wordt, of volgens de eigenschappen, en de vorm van de oppervlakte van dit laatste. Contact is noodzakelijk voor de eerste twee condities, en uitgebreidheid voor de derde. U sluit [uitgebreidheid] totaal uit van uw notie van de ziel, en [contact] lijkt me incompatibel met een immaterieel iets. Daarom vraag ik u een meer specifieke definitie van de ziel dan in uw *Metafysica*, dat is te zeggen van zijn substantie, los van zijn actie, het denken. Want hoewel we deze [de substantie en het denken] als onscheidbaar zien [...] zoals de eigenschappen van God, toch kunnen we ons er een volmaakter idee van vormen, als we ze los van elkaar zouden beschouwen." (eerste brief van Elizabeth aan Descartes, mijn vertaling[6])

In deze korte passage vat prinses Elizabeth precies samen wat nu juist zo problematisch is aan Descartes' interactionisme. Bovendien maakt ze hier meteen ook duidelijk dat substantiedualisme op zichzelf onvoldoende is als recept voor het probleem. Het tweede ingrediënt dat we nodig hebben, is een welbepaalde visie op causaliteit.

Zoals ze zelf aangeeft, gelooft prinses Elizabeth dat er twee dingen nodig zijn om iets zoals een lichaam te kunnen doen bewegen. Het eerste is contact, en het tweede is uitgebreidheid. Aangezien de geest immaterieel is, heeft hij geen 'uitgebreidheid' die ergens tegen kan duwen, en ook geen locatie waar hij contact kan maken met het lichaam. Daaruit besluit Elizabeth dat de ziel nog iets méér moet zijn dan alleen een denkend ding.

De problemen die hier worden opgeworpen, zijn die van locatie en overdracht (later werd dit laatste het probleem van conservatie, maar de wet van behoud van energie was in Descartes' tijd nog niet geformuleerd), en zijn beiden het gevolg van een notie

[6] « [...] en vous priant de me dire comment l'âme de l'homme peut déterminer les esprits du corps, pour faire les actions volontaires (n'étant qu'une substance pensante). Car il semble que toute détermination de mouvement se fait par la pulsion de la chose mue, à manière dont elle est poussée par celle qui la meut, ou bien de la qualification et figure de la superficie de cette dernière. L'attouchement est requis aux deux premières conditions, et l'extension à la troisième. Vous excluez entièrement celle-ci de la notion que vous avez de l'âme, et celui-là me paraît incompatible avec une chose immatérielle. Pourquoi je vous demande une définition de l'âme plus particulière qu'en votre Métaphysique, c'est-à-dire de sa substance, séparée de son action, de la pensée. Car encore que nous les supposions inséparables [...], comme les attributs de Dieu, nous pouvons, en les considérant à part, en acquérir une idée plus parfaite. » (Descartes, Correspondance avec Élisabeth [Online]. Wikisource.)

van causaliteit waarbij oorzaak en gevolg zich niet alleen binnen dezelfde ruimtelijke locatie moeten bevinden, maar ook iets aan elkaar moeten doorgeven. Dat heeft deze visie dus gemeen met die van Salmon en Dowe. Hoewel de theorie van Salmon wel ruimte over laat voor causaliteit zonder fysiek contact, is dit laatste in die van Dowe wel een vereiste, en dit lijkt ook op te gaan voor de visie van prinses Elizabeth. Het typische voorbeeld van dit soort van causaliteit is dat van twee biljartballen die elkaar raken, en waarvan de ene iets van zijn impuls doorgeeft aan de andere. Dit is een zeer intuïtieve causaliteitsopvatting, en hoewel prinses Elizabeth zelf niet dieper ingaat op haar visie op oorzakelijkheid (ze heeft het hier specifiek over beweging, maar niet over causaliteit op zich), is het waarschijnlijk deze die ze vooronderstelt. Zoals we straks zullen zien, zou het best kunnen dat Descartes er een andere causaliteitsvisie op nahoudt, waardoor de problemen die zij opwerpt voor hem minder prangend zijn, of zich zelfs niet stellen. Toch probeert hij haar in zijn volgende brief een antwoord te geven:

> "Vooreerst geloof ik dat zich in ons [denken] bepaalde primitieve noties bevinden, die dienst doen als originelen, als patronen voor al onze andere kennis. [...] Ik geloof ook dat alle menselijke wetenschap niets anders is dan het goed te proberen onderscheiden van deze noties, en ze toe te schrijven aan datgene, waaraan ze toebehoren. Immers, als we iets moeilijks proberen te verklaren aan de hand van een notie die er niet aan toebehoort, kunnen we ook niet anders dan het verkeerd te begrijpen. Net zo, als we één van deze noties proberen te verklaren aan de hand van een andere, kunnen we niet anders dan haar verkeerd begrijpen: aangezien ze primitief is, kunnen we ze enkel verklaren door middel van zichzelf. Aangezien door de zintuigen de begrippen van uitbreiding, vorm en beweging voor ons veel bekender zijn dan de andere, is de belangrijkste oorzaak van onze vergissingen dat we ons meestal van deze noties willen bedienen om zaken te verklaren waaraan deze [noties] niet toebehoren [...] zoals wanneer we de manier waarop de ziel het lichaam beweegt proberen te begrijpen op dezelfde manier als die waarop een

> lichaam een ander lichaam beweegt." (eerste brief van Descartes aan Elizabeth, mijn vertaling[7])

Ook Descartes heeft het hier expliciet over beweging, en niet over 'oorzakelijkheid'. Hoe hij dit laatste precies denkt, brengt hij niet onder woorden. Het feit dat hij hier enkel over beweging spreekt, lijkt me dan ook onvoldoende om te veronderstellen dat zijn visie op oorzakelijkheid dezelfde was als die van Elizabeth. In tegendeel, ben ik zelfs geneigd om deze paragraaf te interpreteren als een waarschuwing om de mechaniek van het lichaam niet te extrapoleren naar het domein van de ziel. Het zou kunnen dat hij de interactie tussen lichaam en geest denkt op een manier die geen behoefte heeft aan contact of overdracht, maar daar gaat hij in de daarop volgende brief niet dieper op in. Hij sluit de zaak zelfs af door Elizabeth erop te wijzen dat het waarschijnlijk niet zo gezond is om zich al te veel met metafysica bezig te houden, en lijkt het debat daarmee op te geven:

> "Maar omdat Uwe Hoogheid opmerkte dat het gemakkelijker is, materie en uitbreiding toe te schrijven aan de ziel, dan de mogelijkheid om het lichaam te verplaatsen zonder materieel te zijn, verzoek ik u vrij om deze uitbreiding en materie aan de ziel toe te kennen, aangezien het niets anders is dan hem te begrijpen als verenigd met het lichaam. [...] Tenslotte, zoals ik geloof dat het noodzakelijk is om toch één keer in het leven de principes van de metafysica begrepen te hebben, omdat zij ons kennis verschaffen over God en de ziel, geloof ik ook dat het zeer schadelijk zou zijn voor de geest om er al te vaak over te mediteren [...] maar dat het beter is om de getrokken conclusies gewoon in

[7] « Premièrement, je considère qu'il y a en nous certaines notions primitives, qui sont comme des originaux, sur le patron desquels nous formons toutes nos autres connaissances. [...] Je considère aussi que toute la science des hommes ne consiste qu'à bien distinguer ces notions, et à n'attribuer chacune d'elles qu'aux choses auxquelles elles appartiennent. Car, lorsque nous voulons expliquer quelque difficulté par le moyen d'une notion qui ne lui appartient pas, nous ne pouvons manquer de nous méprendre ; comme aussi lorsque nous voulons expliquer une de ces notions par une autre ; car, étant primitives, chacune d'elles ne peut être entendue que par elle-même. Et d'autant que l'usage des sens nous a rendu les notions de l'extension, des figures et des mouvements, beaucoup plus familières que les autres, la principale cause de nos erreurs est en ce que nous voulons ordinairement nous servir de ces notions, pour expliquer les choses à qui elles n'appartiennent pas, comme [...] lorsqu'on veut concevoir la façon dont l'âme meut le corps, par celle dont un corps est mû par un autre corps. » (Descartes, Correspondance avec Élisabeth [Online]. Wikisource.)

> het achterhoofd te houden [...].” (2e brief van Descartes aan Elizabeth, mijn vertaling[8])

Hiermee lijkt voor Descartes de kous af te zijn. Toch zal hij in 1649 zijn *Les passions de l'âme* opdragen aan prinses Elizabeth, en daarin een poging ondernemen om toch een locatie toe te kennen aan de plek, waar de ziel “zijn functies in het bijzonder uitvoert”[9]: de pijnappelklier:

> “De reden waarom ik ervan overtuigd ben dat de ziel geen andere plaats kan hebben in het lichaam waar hij zijn functies onmiddellijk uitoefent is dat we alle andere delen van onze hersenen dubbel hebben, net zoals we twee ogen hebben, twee handen en twee oren, en tenslotte al onze zintuiglijke organen, en vooral omdat we maar één dezelfde gedachte hebben van één en hetzelfde ding tegelijkertijd, moet er een plek zijn waar de beelden uit deze twee ogen samenkomen om de ziel te bereiken [...] zodat ze maar één beeld presenteren in plaats van twee. En we kunnen ons gemakkelijk voorstellen dat deze beelden of andere indrukken samenkomen in deze klier door middel van de sappen die de holten van het brein vullen, maar er is geen enkele andere plek in het lichaam waar ze op die manier verenigd kunnen worden.” (artikel 32, mijn vertaling[10])

---

[8] « Mais, puisque Votre Altesse remarque qu'il est plus facile d'attribuer de la matière et de l'extension à l'âme, que de lui attribuer la capacité de mouvoir un corps et d'en être mue, sans avoir de matière, je la supplie de vouloir librement attribuer cette matière et cette extension à l'âme [...]Enfin, comme je crois qu'il est très nécessaire d'avoir bien compris, une fois en sa vie, les principes de la métaphysique, à cause que ce sont eux qui nous donnent la connaissance de Dieu et de notre âme, je crois aussi qu'il serait très nuisible d'occuper souvent son entendement à les méditer, [...]mais que le meilleur est de se contenter de retenir en sa mémoire et en sa créance les conclusions qu'on en a une fois tirées [...]. » (Descartes, Correspondance avec Élisabeth [Online]. Wikisource.)

[9] Descartes gebruikt zelf het woord ‘plus particulièrement’

[10] « La raiſon qui me perſuade que l’ame ne peut avoir en tout le corps aucun autre lieu que cette glande où elle exerce immédiatement ſes fonctions eſt que je conſidère que les autres parties de noſtre cerveau ſont toutes doubles, comme auſſi nous avons deux yeux, deux mains, deux oreilles, & enfin tous les organes de nos ſens extérieurs ſont doubles ; & que, d’autant que nous n’avons qu’une ſeule & ſimple penſée d’une meſme choſe en meſme temps, il faut néceſſairement qu’il y ait quelque lieu où les deux images qui viennent par les deux yeux, [...] ſe puiſſent aſſembler en une avant qu’elles parviennent à l’ame, afin qu’elles ne luy repréſentent pas deux objets au lieu d’un. Et on peut aiſément concevoir que ces images ou autres impreſſions ſe réuniſſent en cette glande par l’entremiſe des eſprits qui rempliſſent les cavitezdu cerveau, mais il n’y a aucun autre endroit dans le corps où elles puiſſent ainſi eſtre unies, ſinon en ſuite de ce qu’elles le ſont en cette glande. » (Descartes, [1649] [Online]. Wikisource.)

Hiermee heeft Descartes natuurlijk nog niets gezegd over hoe we deze uitwisseling tussen de ziel en het lichaam precies moeten denken, ook al heeft ze dan een preciezere locatie.

In een artikel uit 1997 argumenteren Daniel E. Flage en Clarence A. Bonnen dat Descartes' causaliteitsbegrip nog steeds diep geworteld was in de Aristotelische leer van de vier oorzaken. Ze baseren zich op de bezwaren van Arnaud op Descartes' *Meditaties*, en Descartes' antwoord hierop, waarin hij het onderscheid maakt tussen een vormoorzaak en een werkende oorzaak. Descartes gebruikt deze concepten om te verdedigen hoe God zijn eigen oorzaak kan zijn. Volgens Flage en Bonnen werkt het onderscheid in Descartes' metafysica echter nog verder door. Natuurwetten zouden bij Descartes hetzelfde ontologische en epistemologische statuut krijgen als eeuwige waarheden, en constitutief zijn voor de vorm van de wereld (p. 2). Daaruit concluderen zij dat volgens Descartes, afgezien van Gods actie waarbij hij de wereld creëerde, alle oorzaken vormoorzaken zijn (p. 2).

> *"A formal cause is explanatory. Unlike an efficient cause, which explains why or how something comes to be, a formal cause explains why something is what it is. [...] Such explanations are deductive: the essential definition (as major premise) plus a statement of conditions (minor premise) entails a description of the phenomenon to be explained."* (p. 7)

Verderop (p. 16 e.v.) argumenteren ze dat Descartes de gewoonte had om *deductief nomologische* verklaringen te geven waarbij hij gebruik maakte van vormoorzaken. Dit alles zou dan toelaten om beter te begrijpen hoe hij de verbinding tussen lichaam en geest precies dacht:

> *"The lawful connection between states of the will and states of the pineal gland also implies that formal causality helps us understand Descartes' account of the connection between mind and body: Mind and body are lawfully connected."* (p. 25)

Zonder Flage en Bonnen hierin zonder voorbehoud te willen volgen, biedt hun visie wel een mooie illustratie van hoe de problemen die we al dan niet ontmoeten bij mentale

veroorzaking, sterk afhangen van onze positie met betrekking tot causaliteit en verklaringen. Door Descartes een bepaalde visie op causaliteit (vormoorzaken) toe te dichten, plaatsen we meteen zijn metafysica in een ander licht. Als we hun theorie volgen, zou Descartes zijn verklaring immers als volgt gedacht kunnen hebben:

Alle toestanden X van de geest, gaan samen met een toestand X^1 van het lichaam (wet)
De geest is in toestand X (omstandigheid)

Het lichaam is in toestand X^1 (verklaring)

Volgens deze lezing kunnen we Descartes' schijnbare afwimpeling van Elizabeth ook echt als een antwoord lezen: de natuurwet die bepaalt hoe een lichaam een lichaam beweegt, is volgens hem anders dan deze die bepaalt hoe de ziel een lichaam beweegt. Natuurwetten hoeven we voor Descartes niet verder te verklaren (opnieuw, in de lezing van Flage en Bonnen), want ze zijn door God zo bepaald en zijn primitief: we kunnen ze enkel verklaren door middel van zichzelf. Hierbij moet wel worden opgemerkt dat als we deze lezing aanvaarden, Descartes eenzelfde beroep op God zou doen om de verbinding tussen lichaam en geest te verklaren als Leibniz en Malebranche na hem zouden doen. Dit zou hem niet langer tot een interactionist maken, maar net als zij, tot een parallellist.

3.2 David Chalmers

Het is moeilijk om vandaag de dag nog een filosoof te vinden die zichzelf een dualist noemt (hoewel Daniel Dennett ons ervan zou willen overtuigen dat we meer van Descartes' erfenis hebben overgehouden dan ons lief is, cf. infra). Zij die zichzelf wel nog comfortabel onder deze noemer onderbrengen, onderscheiden zich vaak van Descartes' substantiedualisme door zichzelf eigenschapsdualisten te noemen. In tegenstelling tot Descartes, poneren zij niet het bestaan van een immateriële geestelijke substantie die los staat van de wetten van de fysica. Wel beweren zij dat de mentale eigenschappen van hersentoestanden, *voor zover ze mentaal zijn*, niet reduceerbaar zijn tot diezelfde breintoestanden.

Wellicht de bekendste onder hen is David Chalmers, die we vooral kennen van zijn formuleringen van *'the hard problem'*[11], het zombie-argument, en *'the explanatory gap'*[12].

Hoewel Chalmers niet de eerste was om dit te verwoorden (hij verwijst zelf naar eerdere beroemde artikels zoals Thomas Nagel's '*What is it like to be a Bat?*' (1974)), staat hij er vooral om bekend dat hij ernaar streeft om ons het 'moeilijke probleem' (*hard problem)* van het bewustzijn ernstig te doen nemen. Volgens Chalmers is er een onderscheid tussen het 'makkelijke probleem' – in kaart brengen welke breintoestanden correleren met welke bewustzijnstoestanden, een project waar de wetenschap waarschijnlijk in zal slagen[13] – en het 'moeilijke probleem': verklaren *waarom* deze breintoestanden correleren met een fenomenologisch bewustzijn, of waarom er überhaupt zoiets als fenomenologische ervaring bestaat. De kloof tussen deze twee is *'the explanatory gap*'.

Het fameuze zombie-argument[14] is Chalmers' meest populaire manier om het bestaan van deze kloof te illustreren. Een filosofische zombie is iemand die in alle *fysieke* en *functionele* opzichten identiek is aan mezelf, maar die niet beschikt over

[11] Geïntroduceerd door Chalmers in ('Facing Up to the Problem of Consciousness', 1995)
[12] Oorspronkelijk een concept van J. Levine (1983)
[13] In de VS kondigde Obama vorige zomer het Brain Activity Map Project aan (Alivisatos e.a., 2012), in Europa is er het Blue Brain Project (The Blue Brain Project [online])
[14] Bedacht door Robert Kirk (1974)

fenomenologische ervaring. Aangezien we ons deze filosofische zombie perfect kunnen indenken, en hij in principe niet te onderscheiden is van mezelf, is de conclusie dat het bewustzijn nog iets meer (‘*over and above’*) moet zijn dan enkel het kunnen uitvoeren van bepaalde cognitieve operaties. Zo is het een gebruikelijke manier om het functionalisme van o.a. Dennett aan te vallen.

Hoewel dit allemaal een beetje eenvoudig kan lijken, is Chalmers zeker geen groentje. De indrukwekkende metafysische theorieën waar hij zich in zijn boeken aan waagt nopen ons er op zijn minst toe om hem als filosoof ernstig te nemen, en respect te hebben voor de onversaagdheid waarmee hij geen enkele obscure mogelijkheid onbesproken laat. De uitgebreidheid van zijn werk, en het feit dat hij zelf uiteindelijk geen definitieve keuze maakt (hij acht zowel epifenomenalisme als interactionisme mogelijk, maar ook – wat veelal op gehoon wordt onthaald – een vorm van panpsychisme (Chalmers, 2010, p. xviii)), hebben me gedwongen om in wat volgt flink te selecteren. De elementen die ik aan bod laat komen, moeten illustreren hoe echo’s van Descartes’ bevattelijkheids-argument (*conceivability argument)* vandaag doorklinken in de argumenten die Chalmers geeft voor zijn eigen dualisme, en hoe eigenschapsdualisme naast het probleem van causale geslotenheid (dat ik verder bij Jaegwon Kim bespreek) te maken krijgt met een nieuw probleem: de causale werkzaamheid (*efficacy*) van eigenschappen. Tenslotte schets ik hoe Chalmers zijn kennis over de metafysica van informatie gebruikt om te argumenteren voor de mogelijkheid van ‘*pan-proto-psychisme’*.

3.2.1. De superveniëntie van het mentale

In de zesde meditatie beroept Descartes zich op twee zaken om ons ervan te overtuigen dat lichaam en geest gescheiden substanties zijn. Het eerste is onze mogelijkheid om ons dit voor te stellen (vandaar: *conceivability*), het tweede de almachtigheid van God:

> “[...] omdat ik weet dat alles wat ik me klaar en duidelijk kan voorstellen precies zo door God gemaakt kan worden als ik het me voorstel, is het voldoende dat ik in staat ben om me één ding los van het andere voor te stellen, om er zeker van te zijn dat het ene verschillend is van het andere, aangezien ze door de almacht

> van God op zijn minst zo gemaakt zouden kunnen zijn, dat ze apart van elkaar bestaan […]" (§9, mijn vertaling[15])

Chalmers' argumenten vertonen enkele gelijkenissen met die van Descartes. De belangrijkste daarvan zit hem, zoals we al zagen bij het zombie-argument, in onze mogelijkheid om ons iets los van het andere te kunnen voorstellen. Waar Descartes een beroep doet op God om de sprong van deze 'denkbaarheid' naar de realiteit te maken (God zou ons geen helder en duidelijk idee geven van iets dat onwaar was), moet Chalmers zijn redenering echter elders op steunen. De centrale steunpoten van zijn argumentatie zijn de notie van superveniëntie en die van (reductionistische) verklaring.

De definitie die Chalmers van dit eerste geeft in *The Conscious Mind* is de volgende:

> *"B-properties supervene on A-properties if no two possible situations are identical with respect to their A-properties while differing in their B-properties."* (Chalmers, 1997, p. 33)

De B-eigenschappen in kwestie betreffen hier hogere-orde-eigenschappen, en de A-eigenschappen lagere-orde-eigenschappen. Als voorbeeld van een hogere-orde-eigenschap noemt Chalmers 'levend zijn': 'levend zijn' (een biologische eigenschap) supervenieert op de lagere-orde (fysische) eigenschappen van wezens. Het is met andere woorden onmogelijk dat twee wezens identiek zijn wat betreft hun fysische eigenschappen, terwijl het ene wezen leeft en het andere niet. Dit houdt ook in dat de hogere-orde-eigenschap niet kan veranderen (bvb. in 'dood zijn'), zonder dat er ook een verandering optreedt in de lagere-orde-eigenschappen waarop ze supervenieert. Het omgekeerde is wel mogelijk, omdat hogere-orde-eigenschappen meervoudig realiseerbaar kunnen zijn; soms kunnen verschillende constellaties van lagere-orde-eigenschappen, dezelfde hogere-orde-eigenschap tot stand brengen. Ook hiervan is 'levend zijn' een voorbeeld: zowel de lagere-orde-eigenschappen van een konijn als die

[15] "[...] parce que je sais que toutes les choses que je conçois clairement et distinctement peuvent être produites par Dieu telles que je les conçois, il suffit que je puisse concevoir clairement et distinctement une chose sans une autre, pour être certain que l'une est distincte ou différente de l'autre, parce qu'elles peuvent être posées séparément au moins par la toute-puissance de Dieu [...]" (Descartes [1634] 2009, p. 190)

van een kangoeroe realiseren de hogere-orde-eigenschap 'levend zijn'. Bewustzijn is precies zo'n eigenschap waarvan wordt geponeerd dat ze meervoudig realiseerbaar is (cf. infra: Jaegwon Kim). Toch gaat Chalmers op dit laatste in dit deel niet verder in.
Hij gaat verder met het verschil aan te stippen tussen locale en globale superveniëntie:

> *"B-properties supervene locally on A-properties if the A-properties of an individual determine the B-properties of that individual"* (p. 33)

Als voorbeeld hiervan noemt Chalmers 'vorm': dit supervenieert locaal op de fysieke eigenschappen van een object. Eigenschappen als 'waarde' of 'fitheid', worden echter evenzeer bepaald door de feiten met betrekking tot de omgeving van een object of wezen als door de eigenschappen van dit wezen zelf:

> *"B-properties supervene* globally *on A-properties, by contrast, if the A-facts about the entire* world *determine the B-facts: that is, if there are no two possible worlds identical with respect to their A-properties, but differing with respect to their B-properties."* (p. 34)

Belangrijker voor zijn argumentatie echter, is het verschil dat hij maakt tussen 'logische' en 'natuurlijke' superveniëntie. Dit lijkt een onderscheid te zijn dat hij zelf introduceert. Hoewel hij meteen vermeldt dat 'logische -' eigenlijk hetzelfde is als 'conceptuele superveniëntie', en we 'natuurlijke -' mogen opvatten als 'nomische -' of 'empirische superveniëntie', zijn ook dit geen courante termen.[16]

> *"B-properties supervene* logically *on A-properties if no two* logically possible *situations are identical with respect to their A-properties but distinct with respect to their B-properties. […] In determining whether it is logically possible that some statement is true, the constraints are largely* conceptual*. The notion of a male vixen is contradictory, so a male vixen is logically impossible; the*

[16] De definities die Chalmers ervan geeft lijken een beetje gelijk te lopen met Kim's '*strong* -' en '*weak supervenience*'. (Kim, 1993, p. 79 e.v.) Bovendien begint hij zijn illustratie van '*natural supervenience*' op p. 36 met de woorden: '*The weaker variety of supervenience…*'.
Chalmers maakt het nog iets verwarrender als hij onderstreept dat we zijn '*logical supervenience*' niet in de zin van logische deduceerbaarheid mogen interpreteren, maar een pagina verderop toch stelt dat *'In general, when B-properties supervene logically on A-properties, we can say that the A-facts entail the B-facts'* (p. 35-36)
Over het gebruik van al deze termen wordt binnen metafysische kring nog heftig gedebatteerd. De consensus die Chalmers hier voorwendt is dus een pak problematischer dan hij doet uitschijnen.

> *notion of a flying telephone is conceptually coherent, if a little out of the ordinary, so a flying telephone is logically possible. […] In a sense, when logical supervenience holds,* all there is *to the B-facts being as they are is that the A-facts are as they are."* (p. 35-36)

Het is dit laatste dat voor Chalmers van het grootste belang zal zijn. Als B-feiten niet meer zijn dat A-feiten in hun betreffende constellatie, dan kunnen we de wereld volledig beschrijven door alleen te verwijzen naar de A-feiten (bijvoorbeeld feiten over atomen).

Er is echter nog een andere (zwakkere) zin waarin B-feiten kunnen superveniëren op A-feiten, en dat is wat Chalmers 'natuurlijke superveniëntie' noemt:

> *"The weaker variety of supervenience arises when two sets of properties are systematically and perfectly* correlated *in the natural world. For example, the pressure extended by one mole of a gas systematically depends on its temperature and volume according to the law* pV = KT*, where* K *is a constant. […] It follows that the pressure of a mole of gas supervenes on its temperature and volume in a certain sense. […] But this supervenience is weaker than logical supervenience. It is* logically *possible that a mole of gas with a given temperature and volume might have a different pressure; imagine a world in which the gas constant* K *is larger or smaller, for example."* (p. 36)

In gevallen van 'natuurlijke superveniëntie', hebben we dus nog iets meer nodig dan enkel de A-feiten om de B-feiten te beschrijven: we moeten nog iets meer zeggen over de wereld waarin deze A-feiten zich afspelen (in dit geval iets over de gasconstante). Dit is natuurlijk waar Chalmers naartoe wil:

> *"It seems very likely that consciousness is naturally supervenient on physical properties, locally or globally, insofar as in the natural world, any two physically identical creatures will have qualitatively identical experiences. It is not at all clear that consciousness is logically supervenient on physical properties, however. It seems logically possible, at least to many, that a creature physically identical to a conscious creature might have no conscious experience at all, [...].*

> *If this is so, then conscious experience supervenes naturally but not logically on the physical."* (p. 38)

Om te concluderen dat een wezen een fenomenologisch bewustzijn heeft, zullen we dus nog iets meer moeten doen dan alleen maar te verwijzen naar de fysische eigenschappen ervan. Dat bewustzijn correleert hier wel mee (in onze wereld), maar dat het dit doet, volgt hier niet uit zonder dat we meer informatie hebben over de wereld (zoals de gasconstante uit het vorige voorbeeld). Wat het dan precies is dat ontbreekt, bevindt zich natuurlijk in die *'explanatory gap'*. Maar dat het 'zo lijkt', zelfs dat het 'voor velen' zo lijkt, is nog niet voldoende om te concluderen dat er ook werkelijk zo'n kloof bestaat. Dat laatste hangt af van wat je precies verstaat onder een verklaring.

3.2.2. Chalmers' verklaringsopvatting

Hoewel Chalmers ook aandacht schenkt aan andere soorten van verklaringen (zoals 'historische', die 'verwijzen naar de ontstaansgeschiedenis van een fenomeen' (1997, p. 43)) is de belangrijkste soort van verklaring voor dit deel van Chalmers' argumentatie de reductionistische (*'reductive'*) soort. Hij wil immers aantonen dat (sommige) mentale eigenschappen niet reduceerbaar zijn tot hun neurologische correlaten. Als voorwaarde voor een reductionistische verklaring, stelt hij de mogelijkheid tot 'een bepaald soort' analyse van een fenomeen:

> *"Reductive explanation requires some kind of analysis of the phenomenon in question, where the low-level facts imply the realization of the analysis."* (1997, p. 48)

En:

> *"The possibility of this kind of analysis undergirds the possibility of reductive explanation in general. Without such an analysis, there would be no explanatory bridge from the low-level physical facts to the phenomenon in question."* (1997, p. 44)

Uit het voorgaande onderscheid dat Chalmers maakt tussen natuurlijke en logische superveniëntie, kunnen we besluiten dat in de beide gevallen een ander soort van analyse nodig zal zijn om die brug te maken. Dit wordt ook door Chalmers bevestigd:

> *"The epistemology of reductive explanation meets the metaphysics of supervenience in a straightforward way. A natural phenomenon is reductively explainable in terms of some low-level properties precisely when it is logically supervenient on those properties."* (1997, p. 47)

Chalmers stelt dat fenomenologisch bewustzijn niet logisch supervenieert op de '*physical facts*' van onze wereld.

In deze context is het een beetje duister of hij fysische *wetten*, zoals de valwet of de wet van inertie, tot deze *'physical facts'* rekent. In het geval van de gasconstante uit ons voorbeeld, werd ons immers duidelijk gemaakt dat juist de nood aan zo'n extra principe, het verschil maakte tussen logische en 'louter' natuurlijke superveniëntie. Bovendien zal hij stellen (p. 71 e.v.) dat *bijna alles* logisch supervenieert op '*the physical'*. Dat het fenomenologische bewustzijn dit niet doet, zal juist zijn argument zijn tegen materialisme:

> *"In our language, materialism is true if all the positive facts about the world are globally logically supervenient on the physical facts."* (1997, p. 41)

De mogelijkheid tot verwarring bij de lezer zit hier in het betekenisonderscheid tussen 'fysiek' ('betrekking hebbend op de stoffelijke natuur') en 'fysisch' ('betrekking hebbend op de natuur' of 'natuurkundig'), dat wel in het Nederlands, maar niet in het Engels bestaat, waardoor we telkens uit de context moeten afleiden welke van de twee Chalmers bedoelt. Hij stelt dan ook niet dat alle positieve feiten over de wereld logisch superveniëren op de 'fysieke' (materiële) feiten, maar wel op de 'fysische' feiten, d.w.z. feiten over materie met toevoeging van fysische wetten. Door fysische wetten toe te voegen aan de lagere-orde-feiten (wat Chalmers de *'supervenience base'* noemt (p. 86)), kunnen we juist over gaan van louter natuurlijke superveniëntie naar logische superveniëntie[17]. We kunnen een fenomeen met andere woorden

[17] Hij bevestigt deze lezing ook verder op p. 86: *"I have bypassed these problems elsewhere by including physical laws in the supervenience base […]."*

reductionistisch verklaren door het te analyseren in termen van zowel de fysieke als de fysische onderliggende feiten, en aan te tonen dat zij het te verklaren fenomeen realiseren. Om de brug te kunnen maken van de onderliggende feiten naar het te verklaren fenomeen, moeten we van dit laatste eerst een functionele analyse kunnen maken.

> *"Without such an analysis, there would be no explanatory bridge from the lower-level physical facts to the phenomenon in question. With such an analysis in hand, all we need to do is to show how certain lower-level physical mechanisms allow the analysis to be satisfied, and an explanation will result."* (1997, p. 44)

Van de meest interessante fenomenen die om verklaring vragen, zoals 'reproductie' of 'leren', kunnen we volgens Chalmers een analyse maken in termen van de 'causale functie' die het fenomeen vervult, of in staat is, te vervullen. (p. 44) Zelfs van een louter fysische notie zoals hitte, kan volgens hem zo een functionele analyse gemaakt worden, op basis van de causale rol (*causal role)* die het fenomeen speelt. (p. 44-45)[18] Op die manier kunnen we dus van bijna alles een functionele analyse maken die leidt tot logische superveniëntie, en ons dus in staat stelt om het fenomeen reductionistisch te verklaren. Van alles, behalve, zoals je al kan raden, fenomenologisch bewustzijn:

> "*Whatever functional account of human cognition we give, there is a* further question*: Why is this kind of functioning accompanied by consciousness? […] Phenomenal states, unlike psychological states, are not defined by the causal roles that they play.*" (1997, p. 47, cursief in origineel)

Bij elke poging om een cognitief proces functioneel te analyseren, kunnen we ons de vraag blijven stellen waarom dit functionele proces precies gepaard gaat met een

[18] In deze context verwijst Chalmers naar een bezwaar van Kripke (1980), die opmerkte dat aangezien hitte wordt gerealiseerd door de beweging van moleculen, de 'beweging van moleculen' het fenomeen van hitte kan vervangen in een tegenfeitelijke wereld, of deze beweging nu de relevante causale rol speelt of niet. Chalmers brengt hier tegen in dat het de causale rol (bvb. het smelten van ijzer of verbranden van ander materiaal) is die we willen verklaren als we een verklaring willen voor het fenomeen 'hitte', en niet 'de beweging van moleculen'. Dit laatste is het resultaat van de verklaring (a posteriori), maar niet de functionele analyse die ons in staat stelt de verklaring te geven. (p. 45) Het feit dat die causale rol in verschillende zaken gerealiseerd kan worden, zal Chalmers later gebruiken om te argumenteren voor '*strong AI*' en 'pan-proto-psychisme'.

fenomenologisch bewustzijn, waar uit volgt dat we het fenomeen niet exhaustief hebben geanalyseerd: er blijft een deel van het fenomeen dat aan onze analyse ontsnapt. Hieruit mogen we niet besluiten dat fenomenologische toestanden geen causale rol spelen, alleen dat we die niet (volledig) kennen. We missen dus een cruciaal element voor onze reductionistische verklaring.

Hier moeten we ons misschien even afvragen of dit wel voldoende is om de stap te zetten naar dualisme. Ten slotte heeft Chalmers enkel aangetoond dat er een kloof zit in onze mogelijkheden om de wereld te verklaren, maar niet in de wereld *zelf*. Toch is er een manier waarop hij zijn sprong van een epistemologisch naar een ontologisch feit kan verantwoorden, en een aanwijzing hiervoor zit in zijn beschrijving van de voorwaarde voor materialisme: “Materialisme is waar, als alle positieve feiten globaal logisch superveniëren op de fysische feiten.” Dit zou inhouden dat de mogelijkheid om in een bepaalde kenbaarheidrelatie te staan, een ontologische eigenschap is van een reële zaak (een eigenschap die er bovendien voor zou zorgen dat een bepaald epistemologisch paradigma, namelijk materialisme, waar is). Een andere mogelijkheid is dat Chalmers er een ontisch verklaringsconcept op nahoudt, vergelijkbaar met dat van bijvoorbeeld Carl Craver, die beweert dat een verklaring zich niet in ons denken, maar in de werkelijkheid bevindt. (Craver, 2009, p. 27) Wij kunnen dan enkel een meer of minder correcte *beschrijving* geven van de (reductionistische) verklaring. In *The Conscious Mind* bevestigt Chalmers nergens letterlijk dat hij deze visie aanhangt, maar als hij dat zou doen, zou dit er wel voor zorgen dat hij geen sprong van epistemologie naar ontologie meer hoeft te maken.
Iets verderop stelt hij bovendien:

> *“[…] conscious experience involves properties of an individual that are not entailed by the physical properties of that individual, although they may depend lawfully on those properties. […] All we know is that there are properties of individuals in this world – phenomenal properties* – that are ontologically independent *of physical properties.”* (1997, p. 125, mijn cursief)

Er zijn dus twee soorten fenomenen in de wereld: fenomenen die we exhaustief kunnen beschrijven met de (huidige) wetten van de fysica, en fenomenen die we

hiermee niet kunnen beschrijven. Het al dan niet beschrijfbaar zijn op deze manier, ziet Chalmers als een ontische eigenschap van het fenomeen zelf. Hieruit trekt hij een dualistisch besluit.

De stap van een epistemologische naar een ontologische kwalificatie (of het weglaten van deze stap) kan vreemd overkomen. Toch moeten we hierbij opmerken dat een materialist als Daniel Dennett eigenlijk hetzelfde doet wanneer hij uit de onmogelijkheid om qualia op een objectieve manier te bestuderen besluit dat er geen qualia *zijn*: de conclusie die hij trekt is gewoon een andere. Voor David Chalmers is het bestaan van fenomenologische ervaring een bruut feit (Chalmers noemt het een '*brute explanandum'*) dat zich aan ons opdringt, en waar we niet om heen kunnen (p. 86). Dit is natuurlijk bepalend voor zijn conclusie.

Ten slotte zouden we kunnen opperen dat er misschien een wet of principe ontbreekt aan de fysica zoals we die vandaag kennen. Door toevoeging van dit principe zouden we dan misschien wel kunnen aantonen dat fenomenologisch bewustzijn logisch supervenieert op de fysische feiten. Hier heeft Chalmers zelf weinig hoop in, hoewel hij de mogelijkheid niet volledig uitsluit.

> *"The trouble is that the basic elements of physical theories seem always to come down to two things: the structure and dynamics of physical properties. […] No set of facts about physical structure and dynamics can add up to a fact about phenomenology."* (1997, p. 118)

Hierin horen we een verre echo van Descartes, als hij Elizabeth aanspoort om niet te proberen om de mechanische principes van de materiële wereld toe te passen op de verbinding tussen lichaam en geest. Het principe dat in deze kloof past, zal een geheel ander soort principe zijn. Voor Descartes was het misschien een natuurwet die gegarandeerd werd door God, voor Chalmers zal het geen principe van de fysica, maar een *verklaringsprincipe* zijn:

> *"The possibility of explaining consciousness nonreductively remains open. This would be a very different sort of explanation, requiring some radical changes in the way we think about the structure of the world."* (1997, p. 122)

In het volgende deel geeft Chalmers ons een idee van hoe we dat principe moeten denken:

> *"Here the fundamental laws will be* psychophysical *laws, specifying how phenomenal (or protophenomenal) properties depend on physical properties. These laws will not interfere with physical laws; physical laws already form a closed system. Instead, they will be* supervenience laws*, telling us how experience arises from physical processes."* (1997, p. 126, cursief in origineel)

De oplossing zal dus volgens Chalmers te zoeken zijn in de verklaringstheorie.

3.2.3. De werkzaamheid van mentale eigenschappen

Eigenschapsdualisten zoals Chalmers stellen dat het mentale bepaalde eigenschappen heeft die ontologisch verschillen van de eigenschappen die we reductionistisch kunnen verklaren door middel van hun causale functie. Maar ook materialisten geven meestal toe dat mentale processen, hoewel volledig materieel, ook een fenomenologisch *aspect* hebben. Het wordt echter moeilijk voor hen om aan dat fenomenologische aspect ook een causale rol toe te kennen, aangezien dit tot overdeterminatie zou leiden (cf. infra, Jaegwon Kim).

Het onderscheid tussen deze twee soorten aspecten, wordt mooi geïllustreerd met een voorbeeld van Fred Dretske (1989, geciteerd door The Stanford Encyclopedia of Philosophy [online]): als we een voorwerp in zachte klei laten vallen, zal dat daar een afdruk achterlaten: zowel de vorm als het gewicht van het voorwerp zullen bepalend zijn voor de afdruk. We zeggen dat de vorm en het gewicht causaal werkzame eigenschappen zijn in dit verhaal. Toch heeft het voorwerp ook ontegensprekelijk een kleur, alleen heeft die op de afdruk geen invloed. De kleur van het voorwerp is in dit geval causaal irrelevant.

De visie waarbij men stelt dat fenomenologische eigenschappen geen causale rol spelen, maar toch deel uitmaken van het mentale, noemt men epifenomenalisme.

Het onmiddellijke probleem dat hierbij opduikt, wordt door Daniel Dennett via één van zijn typische analogieën scherp gesteld in *Consciousness Explained*:

> *"Consider, for instance, the hypothesis that there are fourteen epiphenomenal gremlins in each cylinder of an internal combustion engine. These gremlins have no mass, no energy, no physical properties; they do not make the engine run smoother or rougher, faster or slower. There is and could be no empirical evidence of their presence, and no empirical way in principle of distinguishing this hypothesis from its rivals: there are twelve or thirteen or fifteen ... gremlins."* (Dennett, 1993, p. 403)

Deze absurdheid is voor Dennett het logische gevolg van het idee van een fenomeen dat geen enkele causale rol speelt in de fysische wereld: er is geen enkele manier om het bestaan ervan te bevestigen of te ontkennen. Dennett maakt zelf (p. 402) het onderscheid tussen de bovenstaande vorm van epifenomenalisme, die volgens hem door de meeste filosofen bedoeld wordt, en epifenomenalisme waarbij een 'epifenomeen' gelijk is aan een 'bijproduct', iets dat samengaat met, maar niet essentieel is voor, een te verklaren fenomeen.

De sterke interpretatie die Dennett hier te kijk zet is er zeker verantwoordelijk voor dat epifenomenalisme meestal niet als een aantrekkelijke optie wordt gezien, en auteurs zich vaak genoopt voelen om zich hiertegen te verdedigen. De illustratie van Dretske toont echter dat zeker niet alle filosofen deze sterke interpretatie aanhangen. De kleur van het voorwerp speelt dan misschien geen causale rol bij het maken van een afdruk, maar dat wil nog niet zeggen dat ze in het geheel geen causale rol *kan* spelen!

De reden waarom sommige filosofen zich toch met de sterke vorm van epifenomenalisme geconfronteerd zien, is dat, zoals Chalmers al stelde, zowat alles in onze wereld te verklaren is aan de hand van fysische fenomenen. We hebben het fenomenale aspect niet *nodig* om onze wereld draaiende te houden of te verklaren: het fysische domein is causaal gesloten. Hieruit zouden we zonder meer kunnen concluderen dat het fenomenologische causaal irrelevant is. (Chalmers, 1997, p. 150)[19] Volgens Chalmers hoeft dit echter niet zo te zijn (in ieder geval hoeft het niet onmiddellijk te leiden tot epifenomenalisme in de sterke zin), maar hangt dit opnieuw

[19] Chalmers verwijst zelf naar Kirk (1979), Horgan (1987) en Seager (1991) voor eerdere formuleringen van deze redenering.

samen met welke visie op causaliteit we verkiezen. Onder de noemer *'strategies for avoiding epiphenomenalism'* reikt hij alvast een aantal mogelijke strategieën aan:

a) Causaliteit als regelmatigheid (*regularity-based causation*)

Als we een sterk Humeaanse causaliteitsopvatting zouden aanvaarden, waarbij 'A veroorzaakt B' enkel wil zeggen dat er een sterke regelmatigheid is tussen A-types van gebeurtenissen en B-types van gebeurtenissen, komt er volgens Chalmers ruimte vrij voor een 'causale' rol voor het fenomenologische. Ook zonder die Humeaanse interpretatie kunnen we causaliteit gelijk stellen met een wetmatige verbinding (*lawful connection* (p. 151)). Dit zou er dan voor zorgen dat de tegenfeitelijke bewering "Als ik geen fenomenologische pijngewaarwording had gevoeld, had ik mijn hand niet teruggetrokken", waar is[20], ook als die fenomenologische gewaarwording een louter 'bijverschijnsel' van een bepaalde hersentoestand was. Zonder dit bijverschijnsel, zou de hersentoestand immers anders zijn geweest (andere eigenschappen hebben), wat dan weer de tegenfeitelijke bewering zou beïnvloeden. Deze uitweg lijkt een beetje op het voorbeeld dat we bij Descartes hebben besproken: als causaliteit enkel een wetmatigheid is, ontsnappen we aan de problemen van contact en overdracht, maar (volgens Chalmers) ook aan dat van conservatie en causale geslotenheid. Zelf verwerpt hij echter deze optie, aangezien de opvatting van causaliteit als regelmatigheid, zelfs al is het een *wetmatige* regelmatigheid, voor hem te zwak is. Hier volg ik hem ook in: louter (wetmatige) correlatie gelijk stellen met veroorzaking is tegenintuïtief, en zou causaliteit tot een impotente notie maken.

b) Causale overdeterminatie (*causal overdetermination*)

Een tweede mogelijkheid ligt er volgens Chalmers in om de mogelijkheid van causale overdeterminatie ernstig te nemen. We weten nog te weinig van causaliteit om deze mogelijkheid bij voorbaat uit te sluiten, zo stelt hij, dus moeten we ze open laten. (p. 152) Gezien de geringe aandacht die hij hier zelf aan besteedt, moeten we echter besluiten dat hij deze optie er terecht terzijde bij vermeldt voor de volledigheid,

[20] Chalmers heeft het hier enkel over actuele regulariteit: de tegenfeitelijke bewering in kwestie geldt enkel in een wereld waarin het bewustzijn gelijkaardig is aan dat van ons.

aangezien het aanvaarden van overdeterminatie enkel en alleen om de mogelijkheid van mentale veroorzaking te behouden een weinig bevredigend *ad hoc* manoeuvre zou zijn.

c) Het niet-superveniëren van causaliteit (*the non-supervenience of causation)*

Chalmers stelt dat er twee soorten van feiten zijn die niet logisch superveniëren op het fysische: feiten over het bewustzijn, en feiten over causaliteit (dit laatste beargumenteert hij verder niet). Hieruit besluit hij dat het "natuurlijk is" om te speculeren dat deze twee manieren van niet-superveniëren nauw verwant zijn met elkaar (p. 152), of "in een hechte metafysische relatie staan" (p. 86), wat op zichzelf maar een zwak gefundeerde veronderstelling lijkt. Verder verwijst hij naar het werk van Gregg Rosenberg, meer specifiek naar een niet gepubliceerd manuscript uit 1996 ('*Consciousness and Causation: Clues toward a double aspect-theory*'), het jaar waarin hij zijn eigen *The Conscious Mind* publiceerde. Hierin stelt Rosenberg dat veel van de problemen omtrent het bewustzijn parallellen vinden in problemen omtrent causaliteit. Daaruit besluit hij dat ervaring causaliteit, of bepaalde aspecten van causaliteit *realiseert* (*realizes*), waardoor het juist het bestaan van ervaring is dat causaliteit mogelijk maakt.[21] (Rosenberg, 1996, geparafraseerd door Chalmers op p. 152)

d) De intrinsieke aard van het fysische (*the intrinsic nature of the physical*)

De positie waar Chalmers zelf het meeste voor te vinden is, gaat uit van het feit dat de fysica haar basiselementen enkel relationeel definieert, op basis van hun causale relaties tot andere elementen:

[21] De ideeën die Chalmers hier aanhaalt zijn wat vaag, en hij gaat er zelf ook niet verder op in. Belangrijk met betrekking tot epifenomenalisme is dat deze zienswijze de causale werkzaamheid van het mentale zou 'redden' door het mentale tot de drager te maken waarin causaliteit wordt verwezenlijkt: een panpsychistische visie. Intussen heeft Rosenberg zelf *A Place for Consciousness* (2004) uitgebracht, waarin hij deze ideeën verder uitwerkt. Interessant genoeg zal Chalmers in zijn latere boek uit 2010 niet langer naar hem verwijzen, hoewel hij ook daarin onomwonden stelt dat zijn eigen sympathieën evenredig verdeeld zijn onder de drie gezichtspunten van interactionisme, epifenomenalisme en pan-(proto-) psychisme. (p. xviii) Misschien was het panpsychisme van Rosenberg zelfs voor Chalmers te radicaal? Zelf heeft hij het immers over 'pan-proto-psychisme'.

> *"Their mass and charge is specified, to be sure, but all that a specification of mass ultimately comes to is a propensity to be accelerated in certain ways by forces, and so on."* (1997, p.153)

De enige eigenschappen van de basiselementen waaruit de wereld bestaat die we kennen, zijn dus relationele eigenschappen. Maar daarmee weten we nog steeds niet alles over wat het precies is, dat in al deze relaties treedt, en wat er de *intrinsieke* eigenschappen van zijn. Een atoom is tenslotte zelf een abstractie van een hele verzameling relaties die zich vertonen met (voornamelijk) andere atomen en meetinstrumenten.

> *"One might be attracted to the view of the world as pure causal flux, with no further properties for the causation to relate, but this would lead to a strangely insubstantial view of the physical world.*[22] *It would contain only causal and nomic relations between empty placeholders with no properties of their own."* (1997, p. 154)

Chalmers gaat verder door te suggereren dat deze basiselementen misschien wel (deels) fenomenologisch van aard zijn. Hij stipt zelf het 'neutrale monisme' (*neutral monism*) van Bertrand Russel aan als de oorsprong van deze visie. Die stelde immers voor om als basis voor de werkelijkheid een neutraal element te nemen, waaruit dan zowel het fysieke als het mentale zouden zijn opgebouwd. Als de basiselementen echter volledig mentaal zouden zijn, leunen we sterker aan bij een idealistische visie (echter één die verschilt van die van Berkeley, verduidelijkt Chalmers op p. 155, aangezien de wereld niet supervenieert op de geest van het subject, maar op een causaal netwerk van mentale basiselementen). Deze visie, zo vertrouwt hij de lezer toe in een voetnoot, "*has been relentlessly pushed on me by Gregg Rosenberg.*"[23]
Zelf stelt Chalmers voor om het over 'protofenomenale' (*protophenomenal*) eigenschappen te hebben. Het instantiëren van een enkele protofenomenologische

[22] Chalmers verwijst hier naar Sydney Shoemaker's 'Causality and Properties' (1980) als een voorbeeld van een dergelijke visie, maar noemt Shoemaker's argumenten om die te verdedigen "grotendeels verificationistisch".
[23] Rosenberg verwijst op zijn beurt naar Russell en Whitehead: *"Scholars should see it as an attempt to make a substantial advance in the development of Bertrand Russell's* Structural Realism *by borrowing some inspiration from Alfred North Whitehead's process philosophy."* (Rosenberg, 2004, p. ix, cursief in origineel)

eigenschap (bvb. in een atoom) zou dan nog niet tot bewuste ervaring leiden, maar misschien zou die wel ontstaan van zodra deze protofenomenologische eigenschappen in een bepaalde onderlinge organisatie treden. (p. 154, cf. infra)

Het probleem van epifenomenalisme blijft hierbij echter om het hoekje loeren, en Chalmers erkent dit ook: het is immers niet hun intrinsiek protofenomenologisch-zijn dat ervoor zorgt dat de basiselementen in allerlei causale relaties treden. Het is zelfs zo dat als we deze basiselementjes zouden vervangen door elementjes die *niet* protofenomenologisch van aard waren, deze causale relaties nog steeds zouden optreden. Natuurlijk zou dit een "subtieler soort van causale relevantie" zijn dan gewoonlijk, geeft hij toe. (p. 154)

3.2.4. Organisationele invariantie, informatie en pan-proto-psychisme

Om de verklaringskloof tussen het fysieke en het mentale te overbruggen, zullen we psychofysische wetten nodig hebben die ons vertellen hoe het precies komt dat het mentale supervenieert op fysische processen. Voordat Chalmers op zoek gaat naar wat die fundamentele wetten zouden kunnen zijn, introduceert hij een aantal principes – regelmatigheden die hij opmerkt in de relatie tussen mentale en fysische processen, en waar de uiteindelijk opgestelde wetten rekening mee zullen moeten houden.

Het eerste van deze principes die hij poneert is dat van coherentie (*coherence*) (Chalmers, 1997, p. 218 e.v.): er is een opmerkelijke coherentie tussen bewustzijn en cognitie – wat Chalmers het fenomenologische en het psychologische aspect van de geest noemt.[24] Zonder cognitie is er geen bewustzijn (zoals Dennett het zou formuleren: er kan geen bewustzijn zijn zonder dat we er ons bewust van zijn). Bovendien is er niet alleen samenhang tussen de twee, maar vertonen de instanties waarin ze zich voordoen ook een grote structurele gelijkenis. Dit noemt Chalmers het principe van structurele coherentie (*structural coherence*). Deze coherentie heeft

[24] *"[The phenomenal concept of mind] is the concept of mind as conscious experience, and of a conscious state as consciously experienced mental state. [The psychological concept of mind] is the concept of mind as the causal or explanatory basis for behavior. [...] According to the psychological concept, it matters little whether a mental state has a conscious quality or not. What matters is the role it plays in a cognitive economy."* (Chalmers, 1997, p. 11)

betrekking op de structuur van de respectievelijke ervaringsvelden (waarin de mogelijke ervaringen zich afspelen):

> "*We might say there is a* difference structure *in our conscious experience [...] that is mirrored by a difference structure in awareness*[25]*: to the manifold of color experiences and relations among them, there corresponds a manifold of color representations and relations among them.*" (Chalmers, 1997, p. 224, cursief in origineel)

Zelfs wanneer een gewaarwording op zich moeilijk te vatten is (zoals de gewaarwording van een specifieke kleur rood), kunnen we wel zonder problemen praten over hoe gewaarwordingen zich tot elkaar verhouden (bvb. met betrekking tot intensiteit, verschil, gelijkenis, etc.). Het zijn dan ook deze relaties die gespiegeld worden in de cognitieve representatie van onze ervaring. (Chalmers, 1997, p. 224-5)

Het derde principe is dat van organisatorische invariantie (*organizational invariance*). De relevante eigenschappen in de fysieke superveniëntiebasis voor het bewustzijn zijn volgens Chalmers *organisationele* eigenschappen. Bewustzijn ontstaat door de functionele organisatie van ons brein. *"Functional organization is best understood as the abstract pattern of causal interaction between various parts of a system, and perhaps between these parts and external inputs and outputs,"* aldus Chalmers. (Chalmers, 1997, p. 147) Het is belangrijk hierbij rekening te houden met het niveau waarop we deze functionele organisatie waarnemen. Maar op een voldoende fijnmazig niveau, moet het mogelijk zijn om neuronen te vervangen door siliconenchips, waarbij het systeem hetzelfde gedrag zal produceren zolang de toestanden van deze siliconenchips hetzelfde patroon van causale interactie vertonen. Twee systemen die in deze strikte (fijnmazige) zin gelijk zijn qua functionele organisatie, noemt Chalmers functioneel isomorf. Het principe van organisationele invariantie stelt nu dat voor ieder systeem dat bewuste ervaringen heeft, ieder ander systeem dat hieraan functioneel isomorf is, kwalitatief identieke ervaringen zal hebben: *"As long as the*

[25] Chalmers gebruikt hier '*awareness'* om te verwijzen naar het psychologische ervaringsaspect.

functional organization is right, conscious experience will be determined." (Chalmers, 1997, p. 249)

Hierbij moeten een aantal belangrijke onderscheiden worden gemaakt. Zoals we zullen zien bij Daniel Dennett, die zichzelf een functionalist noemt, is het bewustzijn voor hem niet meer dan de functionele organisatie van de microstructuren in ons brein. Ook bij Jaegwon Kim komt de term functionalisme terug in de context van functionele analyse voor reductie. Kim zal immers argumenteren voor een niet-reductionistisch fysicalisme omwille van zijn onmogelijkheid om qualia functioneel te analyseren, net zoals Chalmers dit gebruikt om voor een eigenschapsdualisme te kiezen (cfr. supra). Voor Chalmers staat het bewustzijn echter niet *gelijk* aan deze functionele organisatie, maar *ontstaat ze eruit* (*arises*). De relatie tussen het bewustzijn en de functionele organisatie, is geen gelijkheids- maar een superveniëntierelatie.[26] Chalmers noemt deze visie dan ook niet-reductionistisch functionalisme. Een uitermate interessant gevolg van dit soort visie is dat het aanleiding geeft tot het bekende '*Chinese nation argument*' van Ned Block: Stel dat ieder lid van de populatie van China een individueel neuron zou simuleren. Iedere persoon zou met walkietalkies of een ander communicatiesysteem in contact staan met de andere leden, net zoals de neuronen in ons brein met elkaar verbonden zijn. Als functionalisme een correcte visie is, zou een dergelijk systeem, vermits het functioneel volledig gelijk is aan een mensenbrein, ook een bewustzijn hebben. Block ontkent dit. (Block, 1978) Chalmers, maar ook Dennett bevestigen dit echter, zoals ze ook beiden zullen argumenteren voor *'Strong AI'* (cfr. infra).

De basis van Chalmers argumenten hiervoor maakt gebruik van een uitgebreide informatietheorie die hij deels overneemt van C. E. Shannon (1948). In plaats van als een semantische notie, zag Shannon informatie als een formeel systeem waarin het een kwestie was van het selecteren van een bepaalde toestand uit een reeks van

[26] Terwijl Kim veel aandacht besteedt aan de precieze aard van de superveniëntierelatie tussen het mentale en haar superveniëntiebasis, en uitvoerig ingaat op de door hem gekozen term '*realises'*, kiest Chalmers zoals we zien voor de term '*arises'*, maar gebruikt hij verderop ook *'realizes'*. Hij gebruikt deze termen echter veel informeler en zonder ze verder te proberen definiëren.

mogelijkheden. De bit, de kleinste eenheid van informatie, representeert zo een keuze twee mogelijke toestanden: 1 en 0. (Shannon, 1948, geparafraseerd in Chalmers, 1997) Chalmers gebruikt deze notie van informatie als springplank voor het formaliseren van wat hij een informatieruimte (*information space*) noemt:

> *"An information space is an abstract space consisting of a number of states, which I will call information states, and a basic structure of difference relations between those states."* (Chalmers, 1997, p. 278)

De eenvoudigste (niet-triviale) informatieruimte bestaat uit twee toestanden (1 en 0), en de informatieruimte wordt volledig beschreven door het verschil tussen deze twee toestanden, wat Chalmers de verschilstructuur (*difference* structure) van de informatieruimte noemt. De verschilstructuur tussen de toestanden zelf is echter maar één manier waarop we een informatieruimte complexer kunnen maken. De informatietoestanden zelf kunnen immers ook een interne structuur bezitten. Zo kunnen we bijvoorbeeld een ruimte construeren van vier mogelijke informatietoestanden, die elk tien bits bevatten. Die tien bits beschrijven op hun beurt weer elk een verschilstructuur van twee mogelijkheden. Dit noemt Chalmers de combinatorische structuur van de ruimte. De verschilstructuur tussen de informatietoestanden noemt hij de relationele structuur. Natuurlijk kunnen informatieruimtes ook een continuüm beschrijven. De essentie van deze theorie is dat ze in principe eender welke hoeveelheid van informatie kan uitdrukken in een verschilstructuur.

Informatieruimtes en –toestanden kunnen op verschillende manieren in onze wereld worden gerealiseerd, zegt Chamers: in de fysieke wereld, en in de fenomenale wereld. Als voorbeeld van een fysiek gerealiseerde informatietoestand noemt Chalmers die van een lichtschakelaar. Die heeft twee toestanden, omhoog en naar beneden, en realiseert dus een informatieruimte zoals die van de bit. Natuurlijk kan de schakelaar nog veel meer toestanden aannemen: hij kan bijvoorbeeld halverwege staan, of afgebroken zijn. De informatieruimte van fysieke objecten, wordt echter altijd gedefinieerd door een *causaal pad* (*causal pathway*, in dit geval dat van de schakelaar naar de lamp) en een ruimte van *mogelijke effecten* aan het eind van dat causale pad.

(Chalmers, 1997, p. 281) Aangezien een klassieke lamp zonder dimmerschakelaar slechts twee mogelijke toestanden kan aannemen, aan of uit, heeft de schakelaar ook maar twee toestanden die *relevant* zijn. Het enige relevante verschil in een fysiek gerealiseerde informatieruimte is dus dat waarvan haar output een verschil maakt. Chalmers leent een slogan van Gregory Bateson: *"Information is difference that makes a difference."*

Maar wat nu met die andere manier waarop informatie gerealiseerd kan worden? Zoals Chalmers zelf aangaf toen hij het over functionele analyse had, kunnen we onze fenomenologische ervaringen niet analyseren op basis van hun causale rol. Toch kunnen we ook in onze fenomenologie de verschilstructuur van een informatieruimte waarnemen, zegt hij. Als we een abstractie maken van de patronen van verschil en gelijkenis tussen bijvoorbeeld visuele ervaringen, bekomen we een informatieruimte die net zo complex kan zijn als wanneer die fysiek gerealiseerd was.

> *"To find information spaces realized phenomenally, we do not rely on the causal 'difference that makes a difference' principle that we used to ,find information spaces realized physically. Rather, we rely on* ***the intrinsic qualities of experiences and the structure among them*** *– the similarity and difference relations that they bear to each other, and their intrinsic combinatorial structure."* (Chalmers, 1997, p. 284, mijn vetjes)

Dit laatste is wat Chalmers nog nodig had om zijn tweevoudig-aspectprincipe (*double aspect pinciple*) naar voor te schuiven. De informatie die fenomenaal gerealiseerd wordt, wordt immers ook fysiek gerealiseerd. Steunend op het principe van structurele coherentie, kunnen we immers vaststellen dat het om dezelfde informatie gaat: ze beschrijft dezelfde verschilstructuur.

> *"We might put this as a basic principle that information (in the actual world) has two aspects, a physical and a phenomenal aspect. Wherever there is a phenomenal state, it realizes and information state, an information state that is also realized in the cognitive system of the brain. Conversely,* ***for at least some physically realized information spaces****, whenever an information state in that space is realized physically, it is also realized phenomenally."* (Chalmers, 1997, p. 286, mijn vetjes)

Voorlopig laat Chalmers echter open voor welke fysiek gerealiseerde informatie dit laatste zal gelden. Het antwoord dat men hierop zal geven, heeft natuurlijk grote gevolgen voor de ontologie die uit deze theorie zal voortvloeien. Chalmers bespreekt achtereenvolgens verschillende mogelijke posities, maar stipt eerst nog even aan hoe deze informatietheorie een mogelijke verklaring zou kunnen bieden voor de moeite die we hebben om onze fenomenale oordelen onder woorden te brengen:

> *"[...] these judgments arise because our processing system is thrust into locations in information space, with direct access to those locations but nothing else. This direct knowledge will strike the system as a brute 'quality': it knows that the states are different, but cannot articulate this beyond saying, in effect, 'one of those'. This immediate access to brute differences leads to judgments about the mysterious primitive nature of these qualities, about the impossibility of explicating them in more basic terms, and to many of the other judgments that we often make about conscious experience."* (Chalmers, 1997, p. 291)

Hoe ernstig moeten we dat *double aspect pinciple* nu nemen? Als *alle* informatie een fenomenologisch aspect heeft, wil dat dan zeggen dat de wereld even vol ervaring is als hij vol is van informatie?[27] Een eerste beperking die we hieraan kunnen stellen, zegt Chalmers, is dat het een bepaald *soort* van informatie moet zijn om relevant te zijn voor ervaring. Toch is het voor hem niet bij voorbaat uitgesloten dat met elke vorm van informatie ook een vorm van ervaring samen gaat. Als we toelaten dat bewustzijn een gradueel fenomeen is dat toeneemt met een toenemende graad van complexiteit (tussen, zeg maar, een insect en een mens), waarom kunnen we dan niet verder van de ladder afdalen zonder een punt tegen te komen waar bewustzijn plots "uitdooft"?

> *"As we move along the scale from fish and slugs through simple neural networks all the way to thermostats, where should consciousness wink out? [...] Before phenomenology winks out altogether, we presumably will get some sort of maximally simple phenomenology. [...] The thermostat seems to realize the sort of information processing in a fish or a slug stripped down to its simplest*

[27] Gerg Rosenberg zou hier bevestigend op antwoorden. Helaas is zijn theorie wegens plaatsgebrek uit de boot gevallen. (Rosenberg, 2004)

> *form, so perhaps it might also have the corresponding sort of phenomenology in its most stripped-down form."* (Chalmers, 1997, p. 295)

We kunnen de fenomenale eigenschappen op deze laagste niveaus van complexiteit ook aanduiden met wat Chalmers proto-fenomenale (*proto-phenomenal*) eigenschappen noemt. Om die reden noemt hij deze visie ook *pan-proto-psychisme*. Dit is natuurlijk maar een schets van hoe de superveniëntie van het mentale op het fysieke er zou kunnen uitzien. Zoals we verder bij Jaegwon Kim zullen zien, zegt superveniëntie op zich niets over hoe deze relatie concreet tot stand komt. Chalmers blijft voorzichtig in zijn claims, maar vindt geen sluitende redenen om dit soort van mogelijkheden bij voorbaat uit te sluiten: *"There seem to be no knockdown arguments against the view, and there are various positive reasons why one might embrace it."*
Het resultaat van deze houding is dat Chalmers geen concrete theorie naar voor kan schuiven: de vragen die we aan het begin hadden, zijn nog steeds niet opgelost. Wel is het zo dat zijn weigering om mogelijkheden onbesproken te laten een bijzonder rijke visie tot gevolg heeft, die zeer tot de verbeelding spreekt.

4. Fysicalisme

4.1 Daniel Dennett

In de eerste pagina's van zijn boek *Consciousness Explained* stelt Dennett niet alleen voorop dat hij een "empirische, wetenschappelijk respectabele theorie" (p. 4) wil ontwikkelen, maar ook dat hij "de verschillende fenomenen waaruit wat we het bewustzijn noemen is opgebouwd, zal verklaren, en aantonen dat ze allemaal fysische effecten zijn van de activiteiten van het brein, hoe deze activiteiten zijn geëvolueerd, en hoe ze aanleiding geven tot *illusies over hun eigen krachten en eigenschappen*." (p. 16, mijn vertaling en cursief)

Dit zijn grote beloftes, en zonder meteen al te beoordelen of hij deze zal waarmaken of niet, moet worden opgemerkt dat Dennett geen twijfel laat bestaan over zijn positie: hij is een fysicalist, die enkel waarde hecht aan wat objectief ("empirisch") meetbaar is.

Op verschillende plaatsen in zijn boek lijkt Dennett weliswaar de mogelijkheid open te houden dat er een ander soort fenomenen zou kunnen *bestaan* (cf. supra zijn woordkeuze als hij het over Gremlins heeft), maar dat deze alleen geen plaats hebben in een "wetenschappelijk respectabele theorie". Dat we het bestaan van deze fenomenen kunnen bevestigen, noch ontkennen, is het uitgangspunt van zijn 'heterofenomenologische' methode. Bovendien maakt hij hier en daar een claim die nog sterker is, en dus verre van neutraal:

> *"It has been said of behaviorists that they feign anesthesia – they pretend they don't have the experiences we know darn well they share with us. If I wish to deny the existence of some controversial feature of consciousness,* ***the burden falls on me to show that it is somehow illusory****."* (p. 40, mijn vetjes)

Als we mogen aannemen dat we onder deze 'controversiële' aspecten op zijn minst concepten als qualia mogen rekenen, doet Dennett opnieuw een grote toezegging wanneer hij ridderlijk belooft dat hij zal moeten bewijzen dat ze *niet* bestaan, in plaats van enkel te stellen dat hij over hun bestaan geen uitspraken kan doen. Dit is stoere taal.

Daarnaast maakt hij het onderscheid tussen wat hij zelf 'gulzig' reductionisme en 'goed' reductionisme noemt. Het soort reductionisme dat we moeten nastreven, verklaart fenomenen door uit te leggen hoe ze voortkomen uit hun onderdelen en de interacties hiertussen. 'Gulzig' reductionisme zou dan zijn wanneer denkers in hun pogingen om fenomenen zo snel mogelijk te funderen in de meest basale elementen, niveaus overslaan, en niet duidelijk maken hoe we precies kunnen overgaan van het ene niveau naar het andere. (1995, p. 82) Voor deze laatste vorm van reductionisme introduceert hij de metafoor 'hemelhaken', en voor de goede vorm de term 'kranen'[28]. (1995, p. 73 e.v.) Welke vorm van reductionisme hij zal toepassen en verdedigen, is een belangrijke keuze voor elke fysicalist, aangezien hij zal moeten kunnen aantonen dat mentale fenomenen reduceerbaar zijn tot fysische processen. Dit is immers juist wat de (eigen-schaps-) dualist betwist.

Eén van de belangrijkste doelen in *Consciousness Explained* is zijn missie om de invloed van Descartes uit te roeien, die volgens hem zelfs bij "de meest gesofisticeerde materialisten" nog steeds aanwezig is (1993, p. 106) wanneer zij spreken alsof er een centraal punt zou zijn in ruimte en tijd waar het bewustzijn "optreedt". (Dennett noemt dit centrale punt het 'Cartesiaans Theater' (1993, p. 107), en het zal één van de belangrijkste concepten in zijn argumentatie zijn.) Om dit te doen op de manier die hij zelf vereist, zal hij echter bij elke relevante reductiestap duidelijk moeten maken waarop deze is gebaseerd (cf. supra bij Chalmers: *'explanatory bridge'*).

Dennett heeft echter nog een andere missie, die hij aanvangt in zijn in 1985 gepubliceerde boek 'Elbow Room', en verder zet in het latere *Freedom Evolves*: het redden van 'de vrije wil'. Dit is bij uitstek een probleem van mentale veroorzaking: om vast te kunnen houden aan het idee van vrije wil, zal Dennett immers moeten aantonen dat onze fysieke handelingen het gevolg zijn van onze gedachten, op een niveau waarop ze betekenis hebben. Ondanks zijn fervent fysicalisme, krijgt ook hij dus te maken met het probleem van mentale veroorzaking, en zal ook hij nood hebben aan een robuuste causaliteitstheorie om deze metafysische knoop te ontwarren. In

[28] 'Kranen' en 'hemelhaken' slaan op de tussenstappen die worden genomen om van het ene reductieniveau naar het andere over te gaan, en hoe die worden verklaard. Een overgang die gebruik maakt van 'kranen', maakt met andere woorden gebruik van een heldere verklaring die gebaseerd is op fysisch verklaarbare fenomenen, een overgang die gebruik maakt van 'hemelhaken' wordt niet of nauwelijks gelegitimeerd.

Freedom Evolves gebruikt hij de tegenfeitelijke causaliteitstheorie van David Lewis om de notie van determinisme af te zwakken, en stelt hij dat onze capaciteit om verschillende mogelijke uitkomsten met elkaar te vergelijken, voldoende is voor morele verantwoordelijkheid. Deze capaciteit, ook al is het beoefenen ervan een mechanisch proces, is *"free will worth wanting"* (de ondertitel van 'Elbow Room'). Elke andere vorm ervan zou berusten op een geloof in het Cartesiaans Theater en dus inconsistent zijn. Toch lijkt dit geen bevredigend antwoord. Vrije-wil-aanhangers willen (!) immers dat onze handelingen het gevolg zijn van onze gedachten, op een niveau waarop ze betekenis hebben (en dus meer zijn dan datasoep). Met de notie van memen, die zowel in *Consciousness Explained* als in *Freedom Evolves* een belangrijke rol speelt, lijkt Dennett hen dit te gunnen. De vraag die echter open blijft, is hoe memen in relatie staan tot de neurologische structuren die ze mogelijk maken. Zijn neuronenprocessen enkel de drager van de informatie, of zijn ze de informatie zélf? Als een meme een bepaalde handeling kan veroorzaken, hoe doet ze dat dan precies? En waaraan ontleent een meme haar semantische inhoud (intentionaliteit)? Op deze vragen biedt Dennett geen antwoord.

4.1.1 Functionalisme en reductie: homunculi

Het belangrijkste manoeuvre dat Dennett uitvoert aan het begin van zijn *Consciousness Explained*, is wel het introduceren van wat hij de 'heterofenomenologische methode' noemt (p. 72 e.v.). Net na al de moeite die hij heeft gedaan om de lezer ervan te overtuigen dat hij onze (fenomenologische) ervaring ernstig zal nemen (cf. de inleiding van dit deel), pareert hij de hele kwestie door te stellen dat we informatie die enkel subjectief kenbaar is, niet als wetenschappelijke data mogen gebruiken: wetenschap "dringt immers aan op het derde-persoons-perspectief" (p. 72). Dit zal hij dan ook toepassen: de 'heterofenomenologische methode' houdt in dat we als onderzoeker enkel de verbale *rapporteringen* van onze onderzoekssubjecten als data zullen gebruiken. We moeten ons hiertoe verhouden als een antropoloog, stelt hij, die een vreemd geloof van een groep mensen onderzoekt. We doen geen uitspraken over de waarheid van datgene waarin zij geloven, maar kunnen niettemin informatie vergaren over dit (al dan niet

fictieve) fenomeen, aan de hand van hun verhalen. Het komt er op neer dat we vaststellen dat deze mensen in ieder geval praten alsof ze rotsvast geloven dat ze beschikken over een rijke fenomenologische wereld, en willen bestuderen hoe het komt dat ze dat doen. Dit wil dus zeggen dat Dennett aan zijn eigen onderzoeksobject geen ontologisch statuut wil toekennen. Een bizarre beweging, zo vindt naast ikzelf ook David Chalmers (1997, p. 30): deze indirecte manier van bestuderen heeft als resultaat dat Dennett geen model overhoudt van het bewustzijn, maar van "de capaciteit van een subject om over een mentale toestand verbaal te kunnen rapporteren", aldus Chalmers.

Toch wendt hij ons gulheid voor door te stellen dat dit de enige manier is om alsnog recht te doen aan deze fenomenologische informatie:

> *"[...] here is the* neutral *path leading from objective physical science and its insistence on the third-person point of view, to a method of phenomenological description that can (in principle) do justice to the most private and ineffable subjective experiences, while never abandoning the methodological scruples of science."* (p. 72, cursief in origineel)

Een gevolg van deze methode is dat Dennett een functionalist is: we kunnen mentale toestanden enkel beschouwen in termen van de functie die zij vervullen (bvb. zorgen voor rapporteringen van een subject). Dit houdt ook in dat als we erin zouden slagen om een robot te bouwen die precies dezelfde functies zou kunnen vervullen als een mensenbrein (Dennett verwijst naar de 'Turing test'[29] (p. 310-311)), die robot per definitie ook over een bewustzijn zou beschikken.[30] Het is interessant om op te merken dat David Chalmers iets gelijkaardigs poneert met zijn pan-proto-psychisme: volgens

[29] De 'Turing test' verwijst naar een voorstel van Alan Turning (1950, geciteerd door The Stanford Encyclopedia of Philosophy, 2011), waarbij de maat van intelligentie van een artificieel systeem (soms ook: bewustzijn) wordt gededuceerd uit de capaciteit van dit systeem om zich ten aanzien van een menselijke ondervrager voor te doen als een mens (ook het 'imitatiespel' genoemd).

[30] Een bekend tegenvoorbeeld voor dit soort redenering is John Searles gedachte-experiment van de 'Chinese kamer' (Searle, 1980): we kunnen ons volgens hem voorstellen dat iemand die gevangen zit in een kamer zonder vensters maar met een brievenbus, aan de hand van een uitgebreide decoder (een lijstje waarop de passende antwoorden worden gelinkt aan de mogelijke vragen) steeds in het Chinees kan antwoorden op onze in het Chinees geschreven boodschappen, zonder zelf een woord Chinees te begrijpen. Dit toont voor Searle aan dat er voor een bewustzijn nog iets méér nodig is dan enkel het kunnen uitvoeren van de functies van een mensenbrein. Volgens Dennett is in het Chinees kunnen antwoorden op boodschappen in het Chinees, niet meer of minder dan wat 'Chinees begrijpen' betekent. (Dennett, 1993, p. 435-440)

hem wordt bewustzijn gerealiseerd door een bepaald soort functionele organisatie, wanneer die een zekere graad van complexiteit bereikt. Het verschil zit hem in hoe deze functionele relatie door beiden wordt gedacht. Voor Dennett gaat het om de relatie tussen dit systeem en de buitenwereld, of een externe observator. Zo lang het systeem in staat is om in de wereld te functioneren en zich te gedragen zoals een mens, zal het over een bewustzijn beschikken, *op welke manier het hier ook in slaagt*. Zoals we zagen, is Chalmers ervan overtuigd dat een robot, of een 'zombie', ook zou kunnen slagen voor deze test zonder over een bewustzijn te beschikken: de uitwendige tekenen zijn er, maar vanbinnen is er niets gaande. Dit komt omdat voor Chalmers bewustzijn pas kan worden gerealiseerd door de *interne* functionele organisatie van het systeem (de bedrading zeg maar), vanaf dat die een zekere graad van complexiteit bereikt. Een filosofische zombie zou dus volgens hem kunnen slagen voor de Turing test (en dus bewust zijn volgens Dennett) zonder dat zijn interne bedrading dermate complex is dat hij daarbij ook nog eens een bewuste ervaring heeft. Een gevolg van deze posities is wel dat beiden een geloof in '*strong artificial intelligence*'[31] aanhangen; alleen is er voor Chalmers heel wat meer nodig om dit te bereiken.

Dennett noemt zijn visie ook 'eerste-persoons-operationalisme'. Operationalisme wil zeggen dat we een concept (bvb. 'lengte' of 'massa') slechts kunnen definiëren aan de hand van de manier die we gebruiken om het te meten. In die zin is operationalisme dus verwant aan functionalisme. 'Eerste-persoons-operationalisme' betekent dan volgens Dennett dat er geen bewuste ervaring kan zijn zonder dat het subject (vandaar 'eerste-persoons-') in staat is om erover te rapporteren. Het verbale rapport is wat er dan voor zorgt om de gegevens om te zetten naar de derde persoon die Dennett nodig acht om wetenschappelijk te zijn.

Belangrijk is in beide gevallen dat Dennett het enkel wil hebben over de *objectieve* effecten van een fenomeen. John Searle vindt dat Dennett daarmee een categoriseringsfout maakt: we hebben inderdaad data nodig die epistemologisch objectief (intersubjectief kenbaar) zijn, maar dat hoeft niet te betekenen dat ze ook

[31] Deze visie houdt in dat een computerprogramma dat een bewustzijn kan simuleren, ook over een bewustzijn beschikt (Searle, 2004, p. 46). Het betekent met andere woorden dat een bewuste robot theoretisch mogelijk is.

ontologisch objectief hoeven te zijn. (Searle e.a., 1997) Dit zou echter inhouden dat we aan het subjectieve een apart ontologisch statuut zouden moeten toekennen (wat de eigenschapsdualist ook doet) maar dit wordt door Dennett verworpen.

Dennett wil een mechanistische verklaring geven voor het bewustzijn. Dit stelt hij tenminste in zijn inleiding, zonder verder uit te leggen wat hij hier precies onder verstaat. Wel is duidelijk dat hij (meer nog dan David Chalmers) van mening is dat de nodige reductie het beste bereikt kan worden via een functionele analyse[32] van het te verklaren fenomeen. De kern van deze analyse is wat hij 'het meervoudige kladmodel' noemt. Dit model heeft tot doel om het instinctieve idee van het bewustzijn als een steeds doorlopende gedachtestroom zoals in James Joyce's *Ulysses* (het 'Cartesiaans theater', cf. supra), te vervangen door een parallel systeem waarin een ongestructureerde wirwar van outputgegevens die door verschillende subsystemen worden uitgespuugd onder elkaar uitvechten welke zullen worden neergeschreven in het geheugen: een 'Joyceaanse virtuele machine', noemt hij het. Ter illustratie hiervan besteedt hij aandacht aan het fenomeen van visuele perceptie. Verschillende aspecten van visuele informatie bereiken op verschillende momenten verwerkingsmodules van het brein, en worden gebruikt voor snel opeenvolgende herinterpretatie van wat alleen lijkt op een visuele stroom, maar in werkelijkheid niet de continuïteit bezit die wij denken te ervaren. De fragmentarische aard van onze visuele perceptie is natuurlijk gekend, en Dennett breidt dit gegeven uit naar onze hele bewuste ervaring. In deze context gaat hij ook dieper in op het gedrag van diverse zelforganiserende systemen.
De verschillende submechanismen die hun eigen taken uitvoeren, noemt Dennett humunculi, hiermee verwijzend naar het oude idee van de centrale homunculus, een mini-mensje dat aan de besturingsknoppen zou zitten ergens diep vanbinnen in ons brein. Zij vormen de kleinere eenheden van functionele organisatie die de reductie van het te verklaren fenomeen mogelijk moeten maken. Door de diverse taken uit te besteden aan subsystemen die zelf perceptieloos zijn, ontwijkt Dennett de oneindige regressie die het gevolg zou zijn van het postuleren van een centrale homunculus. Wat

[32] Hoewel beide auteurs hun concept van functionele analyse lijken te ontlenen aan Robert Cummins (1975), wordt door geen van beide naar hem verwezen.

precies de interne structuur is van zo'n humunculus, is niet helemaal duidelijk. We weten enkel dat ze zich op een hoger niveau van organisatie en abstractie bevinden dan neuronengroepen:

> *"[...] at this level of complexity and sophistication, even if we succeeded in explaining the process at the level of synapses and bundles of neurons, we would be mystified about other aspects of what must be happening. If we are to make sense of this at all, we must first ascend to* ***a more general and abstract level****. Once we understand the process in outline at the higher level, we can think about descending once again to the more mechanical level of the brain."* (Dennett, 1993, p. 193, mijn vetjes)

Dit is één van de momenten waarop Dennetts ontologische standpunt verwarrend wordt: het lijkt er initieel op dat Dennett met zijn meervoudige kladmodel enkel een complexe informatieverwerkende machine heeft beschreven, zonder dat hij de sprong kan maken van homunculi naar bewustzijn, terwijl het juist deze stap is die we willen verklaren. In dit geval zou hij zijn functionele reductie bereiken door voor complexe functionele fenomenen zoals leren of het geheugen, combinaties van minder complexe mechanismen (homunculi) in de plaats te stellen, waardoor hij kan afdalen naar lagere niveaus van complexiteit. Elders lijkt hij echter te suggereren dat het bewustzijn de informatie *zelf* is (zie bvb. 'Joycean machine'(p. 225-226)), en er verder niets te verklaren valt.

Het is interessant om op te merken dat waar Chalmers de causale werkzaamheid van het fenomenale als een openstaande vraag beschouwt (hij ziet immers epifenomenalisme als één van de valabele visies), het bij Dennett niet helemaal duidelijk is of hij deze werkzaamheid nu ontkent (wat zijn ontbrekende sprong naar de ervaring zou verklaren), of gelijkstelt aan die van onze cognitieve operaties. Het eerste lijkt me, gezien zijn smalende houding ten aanzien van epifenomenalisme, weinig waarschijnlijk. In dit laatste geval zou het bewustzijn echter gelijk staan met de objectief meetbare hersenprocessen in ons brein en verder niets, wat zijn latere standpunt over de vrije wil lijkt te zullen ondermijnen. Bovendien lijkt hij dit zelf tegen te spreken in *Freedom Evolves*:

> *"In most of the species that have ever lived "mental" causation has no need for, and hence does not evolve, any elaborate capacity for self-monitoring. In general,* ***causes work just fine in the dark****, without needing to be observed by anybody, and that is as true of causes in animals' brains as anywhere else. So however "cognitive" an animal's faculties of discrimination might be, the capacity of their outputs to cause the selection of appropriate behavior does not need to be experienced* by anything or anybody*."* (Dennett, 2003, mijn vetjes, cursief in origineel)

Hier lijkt hij wel te suggereren dat het bewustzijn ("*experience*") toch nog iets bovenop het normale cognitieve functioneren is, iets dat ervoor zorgt dat dit zich niet "*in the dark*" afspeelt. Vooralsnog blijft het niet geheel duidelijk welke positie Dennett nu verkiest.

4.1.2 Causaliteit, determinisme en de vrije wil

Dennetts argumentatie in *Freedom Evolves* is opgebouwd uit drie grote delen. In het eerste (en belangrijkste) ervan, stelt hij dat een deterministisch universum niet hoeft te impliceren dat alles onvermijdelijk is. Hiervoor bedient hij zich lustig van retoriek en neologismen, zoals wanneer hij het woord vermijdelijk ("*evitable*") introduceert om indien niet een metafysische, dan toch een talige opening te krikken in een causaal gedetermineerde wereld.

"*Determinism is the thesis that 'there is at any instant exactly one physically possible future'*", herinnert Dennett ons. (Van Inwagen, 1983, p. 3, geciteerd door Dennett, 2003, p. 25) Toch wil dit volgens hem niet zeggen dat alles onvermijdelijk is, in tegendeel: we zijn voortdurend bezig met het vermijden van dingen die ons schade zouden kunnen berokkenen. Daarom zijn we zo'n succesvolle organismen, zo stelt hij. Juist doordat de wereld gedetermineerd is (door fysische wetten), zijn we in *staat* om te vermijden wat er *zou gebeurd zijn* als we niet hadden ingegrepen! (Dennett, 2003, p. 58-59) Voor wie daarbij de wenkbrauwen optrekt, voegt Dennett daar snel aan toe: "*I guess it all depends on what you mean by 'going to happen'.*"

Dennett baseert zich voor zijn concept van ‘mogelijkheid’ (‘possibility’) op de ‘mogelijke-werelden’-theorie van David Lewis (1973b). Het klopt dat als onze wereld in al zijn microscopische toestanden perfect overeenkomt met een andere, deze twee werelden ook nooit meer van elkaar zullen verschillen, zegt Dennett. Toch zullen we nooit helemaal zeker kunnen zijn welke van de ondenkbaar talrijke mogelijke werelden de eigenlijke wereld is, zelfs niet als we de wetten van de fysica perfect kunnen bevatten, tenzij we ook (zoals de demon van Laplace) een perfecte en volledige kennis hebben van de toestand van ieder atoom in ons universum. (p. 69) Dit is relevant voor het beoordelen van tegenfeitelijke beweringen, gaat hij door, beweringen van het type: “Als ik Arthur niet had geduwd, zou hij niet zijn gevallen.” We besluiten dan dat onze bewering over Arthur zal kloppen in onze wereld als Arthur ook valt in elke mogelijke wereld waarin ik hem duw en die voldoende overeen komt met de onze om relevant te zijn (de wetten van de zwaartekracht zijn er hetzelfde, en de kamer is niet gevuld met airbags). Op basis van dergelijke beweringen banen we ons een weg door onze omgeving, en vormen we ons een beeld van wat mogelijk is, en wat de ‘regels’ zijn. Toch kunnen we nooit voor de volle honderd procent zeker zijn in welke mogelijke wereld we ons bevinden, omdat onze kennis ervan altijd beperkt blijft:

> *“Every finite information-user has an epistemic horizon; it knows less than everything about the world it inhabits, and this unavoidable ignorance guarantees that it has a* subjectively *open future.”* (Dennett, 2003, p. 91, cursief in origineel)

Het is duidelijk wat Dennett hier mee bedoelt: vanuit het standpunt van een alziende demon mag de toekomst dan wel vast liggen, maar vanuit ons eigen standpunt is dat niet zo. Bovendien is het aantal mogelijke werelden (of mogelijke toestanden van de wereld) zo onbevattelijk groot, dat het idee dat een volgende toestand op de één of andere manier voorspelbaar zou zijn, ons absurd overkomt. Hier en daar lijkt Dennett te verwijzen naar de chaostheorie als hij erop wijst dat “minieme variaties” op de startcondities een hele andere toekomst teweeg zouden brengen. Toch is dit geen argument voor ‘vermijdelijkheid’: het verwart immers determinisme met voorspelbaarheid. Dat het voor ons onmogelijk is om de toekomst te voorspellen, is van geen belang voor de demon van Laplace, die over een oneindig

bevattingsvermogen beschikt. Die zou dus geen probleem mogen hebben met een systeem dat zo complex is dat het zich chaotisch gaat gedragen.

Door in zijn argumentatie te wisselen van perspectief, kan Dennett volhouden dat determinisme (op metaperspectief) niet leidt tot onvermijdelijkheid (vanuit het subject gezien). We proberen immers voortdurend een aantal 'mogelijke toekomsten' te vermijden, zoals in zijn favoriete voorbeeld waarin we een vliegende baksteen ontwijken. Vanuit het standpunt van de demon zou de baksteen nooit ons hoofd hebben geraakt: we zouden hem altijd hebben ontweken. Maar vanuit ons eigen standpunt zou hij tegen ons hoofd zijn gevlogen, en *juist daarom* hebben we hem ontweken!

We kunnen gemakkelijk volhouden dat dit alles weinig impact heeft op wat *echt* mogelijk is (en niet alleen maar mogelijk vanuit een subjectief gezichtspunt), maar dit legt Dennett naast zich neer: "*To say that if determinism is true, your* future *is fixed, is to say ... nothing interesting*," stelt hij.

Maar wat dan met causaliteit? Dennett weigert zich vast te pinnen op één bepaalde causaliteitsopvatting. Het is realistischer om een formele analogie te ontwikkelen die ons toestaat om op een meer heldere manier na te denken over de wereld, zegt hij.[33] (Dennett, 2003, p.71) Voor wie ietwat vertrouwd is met Dennett is dit natuurlijk al meteen een beetje een verdachte beweging; door geen partij te kiezen in het hedendaagse causaliteitsdebat, kan hij de visies van verschillende auteurs met elkaar combineren, zonder dat hij voor specifieke posities hoeft te argumenteren – het is immers maar een illustratie.

Dennett baseert zich losjes op Lewis (2000, geciteerd door Dennett, 2003, p. 72 e.v.) waarbij hij een paar gekende voorbeelden opsomt van probleemgevallen van causaliteit. Daarbij introduceert hij ook een aantal concepten bij de lezer: causale noodzakelijkheid ('*causal necessity*'), voldoende grond ('*causal sufficiency*'), onafhankelijkheid ('*independence'*), en temporele prioriteit ('*temporal priority*').

[33] Zijn letterlijke woorden zijn: *"We should mistrust any informal arguments that masquerade as 'proofs' validating or debunking particular causal doctrines." "These are fighting words to some philosophers, of course,"* zo gaat hij verder in een voetnoot, *"Fine; we happily shift the burden of proof to them."*

Hierbij moet weer worden opgemerkt dat Dennett geen van deze factoren ziet als voorwaarden voor causaliteit: hij stipt enkel aan dat ze “een rol lijken te spelen” bij onze beoordeling ervan.

Vervolgens stelt hij dat determinisme een doctrine is over ‘voldoende grond’ (‘*causal sufficiency*’):

> “*If S_0 is a (mind-bogglingly complex) sentence that specifies in complete detail the state description of the universe at* t_0*, and S_1 similarly specifies the state description of the universe at a later time* t_1*, then determinism dictates that S_0 is* ***sufficient*** *for S_1 in all physically possible worlds. But determinism tells us nothing about what earlier conditions are* **necessary** *to produce S_1 or any other sentence for that matter.*” (Dennett, 2003, p. 84, mijn vetjes, cursief in origineel)

Dennett bedoelt dat, welke historische feiten ook aan S_1 vooraf gaan, we misschien ook op een andere manier tot S_1 gekomen zouden kunnen zijn. We kunnen niet volhouden dat, als S_0 er niet was geweest, S_1 er ook niet was geweest. Theoretisch gezien is het mogelijk (voor een alwetende demon van Laplace) om vanuit een perfecte en volledige kennis van de huidige toestand van het universum iedere volgende toestand te voorspellen. Dat is echter niet mogelijk in de omgekeerde richting. Als we terug proberen te gaan naar het verleden, kunnen we niet zomaar alles afleiden dat vooraf ging. De toekomst ligt dus vast, maar het verleden blijft onzeker, grapt Dennett.

Dennett schetst als voorbeeld een wereld waarin, net als in de onze, John F. Kennedy vermoord wordt in 1963. “John F. Kennedy is vermoord in 1963,” noemt hij deelzin C. Deze wereld is atoom per atoom identiek aan de onze, zo stelt hij, alleen verplaatsen we John F. Kennedy op het moment in kwestie één millimeter naar links. We veranderen dus iets aan de geschiedenis. “John F. Kennedy is neergeschoten in 1963,” is nog steeds waar, zegt Dennett, maar met een minieme variatie in de omstandigheden die ervoor zorgen dat het waar is. Wat blijkt nu? Als we vertrekken van *deze* S_1, en de “film” terugspoelen helemaal tot aan het ontstaan van het universum, komen we S_0 helemaal niet tegen!

> "*There are highly similar possible worlds in which Kennedy is killed but S0 is not the case, so the state of the universe described by S0 is* not *the cause of Kennedy's assassination*." (Dennett, 2003, p. 84, cursief in origineel)

Wat Dennett hier op een ietwat onhandige manier tracht te illustreren, is dat er vele mogelijke werelden zijn waarin Kennedy vermoord wordt (en die dus in dat opzicht gelijk zijn aan de onze), maar waarin S_0 geen deel uitmaakt van de geschiedenis.
De vraag is natuurlijk of S_1 ook informatie bevat over de geschiedenis van atomen. In dat geval kunnen we vanaf die net iets andere moord op Kennedy immers niet meer tot S_1 komen, want S_1 zal van deze minieme variatie de sporen blijven dragen.
Dit lijkt een beetje op Mackies oplossing voor het verhaaltje van Billy en Suzy die elk met een steentje naar de ruit gooiden. Het klopt dat als Billy zijn steentje niet had gegooid, de ruit nog steeds gebroken zou zijn, maar dat zou dan *op een andere manier* zijn gebeurd. Net zo is één millimeter naar links staan een andere manier van vermoord te worden. Het hangt er dus maar vanaf welke informatie je opneemt in de beschrijving van de wereld.

Maar Dennett beweert in dit deel nog iets meer. Door te verdedigen dat de vorige toestand van het universum (S_0) wel voldoende, maar niet noodzakelijk was voor onze huidige toestand S_1, wil hij de lezer doen besluiten dat S_0 niet echt de *oorzaak* was van S_1. Om van een oorzaak te spreken, hebben we immers noodzakelijkheid nodig, stelt hij:

> "*Hence, since causation generally presupposes necessity, the truth of determinism would have little, if any, bearing on the validity of our causal judgments*." (Dennett, 2003, p. 84)

Dit hangt er natuurlijk maar van af welk causaliteitsconcept je precies wil verdedigen. In feite is het zelfs zo dat David Lewis, waar Dennett zich al een aantal bladzijden op baseert, dit zelf niet als een voorwaarde beschouwde. Noodzakelijkheid impliceert causaliteit, stelt die, maar het omgekeerde is niet zo, omwille van de mogelijkheid van causale ketens ('*causal chains*', cfr. supra). Daarmee lijkt Lewis te zeggen dat de wereld die S_0 beschrijft niet noodzakelijk hoeft te zijn voor de wereld die beschreven wordt door $S_{1,}$ zo lang er maar toestanden van de wereld zijn die voor de daarop volgende

toestand noodzakelijk zijn, en zo een keten vormen van de wereld van S_0 naar die van S_1. De 'noodzakelijkheid' waar Lewis het over heeft is tegenfeitelijke afhankelijkheid. Er bestaan natuurlijk nog tal van andere vormen van noodzakelijkheid. Bovendien doet niet elke causaliteitstheorie een beroep op noodzakelijkheid. In hetzelfde verhaal van Billy en Susan, kunnen we zeggen dat het steentje van Billy de oorzaak was van de gebroken ruit, hoewel het er niet noodzakelijk voor was. Om dit vol te houden, hoeven we enkel een procestheorie zoals die van Salmon en Dowe (cfr. supra) aan te hangen in plaats van een tegenfeitelijke theorie.

Hoe kadert dit alles in Dennetts conceptie van de vrije wil? Hij wil aantonen dat zelfs in een deterministische wereld, er nog steeds mogelijkheden zijn "*in an important everyday sense of the word.*" (Dennett, 2003, p. 77) Maar dan nog is het moeilijk om te zien wat de relevantie is van zijn betoog in het grotere geheel. Het lijkt erop dat Dennett hier via retoriek onze notie van determinisme wil afzwakken, terwijl hij zelf eerder toegaf dat de toekomst enkel open is vanuit het subjectieve gezichtspunt (Dennett, 2003, p. 91). Dat we het verleden niet perfect kunnen afleiden, verandert hier niets aan. En waarom zouden we ons hier zorgen om maken, aangezien dat toch alleen maar een zaak is van die Laplaciaanse demon? Wat ons werkelijk aanbelangt, is hoe we ons dat subjectieve gezichtspunt precies moeten voorstellen, en hoe het komt dat het ons een kans biedt op vrijheid.

4.1.3 Het 'ik' is geen speldenprik

De meest waardevolle bijdrage van Dennett aan het vrije-wil-debat zit volgens mij in het ene zinnetje *"If you make yourself really small, you can externalize virtually everything"* (Dennett, 1984, p. 143). Dit was de belangrijkste zin in zijn boek *'Elbow Room'*, zo zegt hij zelf in een voetnoot in *Freedom Evolves* (p. 122), waarin hij onderstreept dat het hem natuurlijk om het tegendeel te doen is: het rechtzetten van de misvattingen die het gevolg zijn van het idee van het 'zelf' als een speldenprik. Of in zijn eigen woorden: *"You'd be surprised how much you can internalize, if you make yourself large."*

De passage in *Freedom Evolves* waar deze voetnoot bij hoort, maakt deel uit van een lang hoofdstuk waarin Dennett zich richt tegen het libertarisme van Robert Kane, in

zijn boek *The Significance of Free Will* (1996). Dennett geeft er een uitgebreide analyse van Kanes filosofie, en wijst erop dat diens pogingen om een moment van indeterminisme (waar libertaristen aan vasthouden) binnen te smokkelen in ons beslissingsproces, gebaseerd zijn op de foute veronderstelling van een duidelijk afgebakend en gelokaliseerd 'ik', waarin dit beslissingsproces plaatsvindt (zie: het 'Cartesiaans theater'). Dit zal natuurlijk irrelevant blijken te zijn aan het einde van zijn betoog, aangezien indeterminisme, of het nu bestaat of niet, helemaal geen hulp biedt voor vrije wil of verantwoordelijkheid.

> *"... if the decision is undetermined – the defining requirement of libertarianism – it isn't determined by you, whatever you are, because it isn't determined by* ***anything****."* (Dennett, 2003, p. 123, mijn vetjes)

Wat wel relevant is, is hoe deze opvatting van het gecentraliseerde 'ik' - een erfenis van het 'centrale homunculus'-idee - het vrije-wil-debat in belangrijke mate heeft gecontamineerd. Bovendien lijkt het soms alsof deze opvatting, die eigenlijk gewoon uit de volkspsychologie komt, niet eens in vraag wordt gesteld, terwijl ze toch uitermate problematisch is en verdient om op zijn minst expliciet te worden gemaakt. Toch lezen we bvb. in het boek 'Zonder Vrije Wil' van Jan Verplaetse, wanneer hij uitleg geeft bij het concept 'broncontrole':

> "Het gaat me hier om het woordje 'zelf'. [...] De productie van beslissingen moet in handen zijn van een bron waarover de beslissende persoon controle heeft, *een eigen Ik bijvoorbeeld*. Omgekeerd kunnen oorzaken die vreemd zijn aan de persoon geen beslissingen voortbrengen. Althans geen beslissingen die men vrij noemt." (Verplaetse, 2012, p. 65, mijn cursief)

Het is een beetje onthutsend hoe snel hier over dat "eigen Ik bijvoorbeeld" wordt gegaan (terwijl het blijkbaar wel een hoofdletter verdient), dat toch zo centraal is voor het idee van broncontrole. Wat binnen of buiten dat 'ik' valt, hangt er immers volledig van af hoe we dat gaan definiëren. Door zijn introductie van het 'meervoudige kladmodel' en de 'virtuele Joyceaanse machine' (cfr. supra), levert Dennett alvast enkele hulpmiddelen aan om dat 'ik' anders te denken dan als een centraal punt in ons brein waar alle beslissingen worden genomen. Maar dit is natuurlijk maar één

mogelijke visie. Sommige harde fysicalisten bedoelen met 'ik' gewoon 'mijn lichaam', oftewel 'de atomen waaruit ik besta' (zie bvb. de boeken van Dick "ik ben mijn brein" Swaab). Dat het feit dat we het hebben over 'ons' lichaam een belangrijke aanwijzing is dat dit een foute opvatting is, benadrukken dan weer Wittgensteiniaan Peter Hacker en neuroloog M.R. Bennett (2003). Het laatste dat je kan zeggen over het 'ik' is dat er niet oneindig veel boeken over zijn geschreven.

Wie echter geen zin heeft om zich toe te leggen op Heidegger[34] of andere continentale bespiegelingen over de aard van het 'zelf', kan zich in ieder geval aansluiten bij Dennett als hij poneert dat het 'zelf' eigenlijk niet meer is dan een concept, een abstractie die ons toestaat om bepaalde relaties te denken, net zoals het gravitatiecentrum in de fysica. Zelf noemt hij het dan ook 'het centrum van narratieve gravitatie' *('center of narrative gravity')*. (Dennett, 1993, p. 410)

Maar als het 'zelf' een verhaal is dat we over onszelf vertellen (of het "zich verhouden", zoals Kierkegaard het stelt)[35], wordt het al snel heel problematisch om uit te maken wat daar wel of niet toe behoort. Zijn mijn voorouders een deel van mijn 'ik'? De kat die zonet de weg overstak voor mijn fiets? De stad Gent?

Dennett probeert de zaak te illustreren met een knullig syllogisme waarin wordt bewezen dat zoogdieren niet kunnen bestaan, omdat een zoogdier altijd een ander zoogdier als moeder heeft, en er bijgevolg nooit een eerste zoogdier kan zijn geweest. (p. 126 - hij ontleent dit voorbeeld aan een artikel van David Sanford uit 1975) Waar hij echter op aanstuurt, is dat het, net als met zoogdieren of "de kip of het ei?"-vragen, onmogelijk is om een lijn te trekken die aanduidt waar 'ik' de bron wordt van mijn handelingen, omdat je het 'ik' niet kan afbakenen. De vereiste voor broncontrole is dus zelf inconsistent.[36]

---

[34] Toch lijkt Heidegger me hier zeer relevant, al is het maar omwille van de meesterlijke manier waarop hij de opvatting van het 'zelf' als een soort van object onderuit haalt.

[35] *"The self is a relation that relates itself to itself or is the relation's relating itself to itself in the relation; the self is not the relation but is the relation's relating itself to itself."* (Kierkegaard, 1849, geciteerd in Blattner, 2006, p. 35) - indien nog beargumenteerd moet worden dat het 'zelf' een problematische notie is!

[36] Als we het 'ik' zouden beschouwen als een verhaal, kunnen we er (vanuit een bepaalde visie begrepen) wel oneindig veel dingen aan blijven toevoegen. Zoals de populaire serie *'How I met your Mother'* intussen al acht jaar doet over dat ontmoetingsverhaal, kunnen we in theorie helemaal terug gaan naar het ontstaan van het universum, terwijl we nog steeds 'mijn' verhaal aan het vertellen zijn. Op die manier zou ik de oorzaak én het gevolg zijn van alles dat met mij in verband te brengen valt. Dit is

De vraag is natuurlijk, of dit alles dan niet juist in de kaart speelt van de harde incompatibilisten[37]. Als broncontrole noodzakelijk is voor vrije wil, terwijl dit een notie is die we niet duidelijk kunnen definiëren, is de zaak dan niet bij voorbaat al beklonken? Zelf ben ik die mening toegedaan over dat hele idee van 'vrije wil': ik heb er nog nooit een bevredigende definitie van gehoord die zich niet beroept op even problematische concepten zoals 'broncontrole'. Dennett is echter vastbesloten om niet het kind met het badwater weg te gooien. Om dat kind te redden, doet hij een beroep op de reeds eerder aangehaalde notie van 'memen'.

4.1.4 Memen zijn hemelhaken.

Dennett ontleent die memen aan Richard Dawkins, die ze als een toemaatje introduceerde aan het eind van zijn boek *The Selfish Gene* (1976, p. 206 e.v.). In *Consciousness Explained* spoort Dennett ons aan om het idee van memen letterlijk te nemen:

> *"Once our brains have built the entrance and exit pathways for the vehicles of language, they swiftly become* parasitized *(and I mean that quite literally, as we shall see) by entities that have evolved to thrive in just such a niche:* memes*."* (p. 200, cursief in origineel)

Memen, zo zegt Dawkins, kunnen melodieën zijn, ideeën, slagzinnetjes, kledingstijlen, methodes om potten te bakken of bogen te bouwen. (Dawkins, 1976, p. 206) Ze voldoen aan precies dezelfde eisen als genen dat doen, om van natuurlijke selectie te kunnen spreken: ze kunnen eindeloos variëren, kunnen zich repliceren (door zich te verspreiden in de hoofden van mensen), en kunnen een verschillende '*fitness'* hebben, die (net zoals bij genen) *niet hoeft te correleren met de fitness van het organisme dat er de drager van is*. Dennett geeft hier zelf enkele voorbeeldjes van: samenwerking, muziek, onderwijs en milieubewustzijn zijn memen die zich verspreiden en ons ook ten goede komen. Maar andere memen zoals racisme, vliegtuigkapen, computervirussen

---

helemaal niet wat Dennett in gedachten heeft, maar het zou er wel voor zorgen dat 'ik' de bron zou zijn van al mijn handelingen!

[37] Incompatibilisten stellen dat het bestaan van vrije wil uitgesloten is in een deterministisch universum.

en vandalisme, verspreiden zich even goed, zonder dat we er zelf iets aan hebben. (Dennett, 1993, p. 203) Vanuit het standpunt van de evolutie, is een meme succesvol zolang het erin slaagt om zichzelf te kopiëren.

Op die manier kunnen we het ontstaan van cultuur zien als een eindeloos complex netwerk van memen, dat geïmplementeerd is in de breinen van de menselijke soort, een beetje zoals het internet gedragen wordt door computerservers. Dennett gaat zelfs zo ver als te zeggen dat het bewustzijn *zelf* een virtuele machine is die het resultaat is van *meme-effecten* in ons brein. (Dennett, 1993, p. 210) Dit sluit natuurlijk aan bij zijn functionalisme, en de stelling dat 'bewust zijn' niets meer of minder is dan te kunnen slagen voor de Turing-test (cfr. supra). Hij besluit dan ook dat het 'ik' (in de zin van Descartes' res cogitans), is opgebouwd uit memen: *"Our existence as us, as what we as thinkers are – not as what we as organisms are – is not independent of these memes."* (Dennett, 1993, p. 208)

Tot nu toe belooft dit niet veel goeds voor morele verantwoordelijkheid: als ons bewustzijn (en ons 'ik') alleen een soort computerprogramma is dat geïmplementeerd wordt door de ééntjes en nulletjes van vurende neuronen, hoe ontsnappen we dan aan de argumenten van de incompatibilisten?

In een poging om te laten zien dat zij het bij het verkeerde eind hebben, introduceert Dennett - zoals hij wel vaker doet - een antagonist in zijn eigen boek, een fictief personage dat hem in dialoogformaat de bezwaren van de tegenpartij voor de voeten werpt. Daarop gaat hij verder door deze overduidelijke stroman te overtuigen dat net zoals romantische liefde niet minder reëel wordt door alles wat we kunnen leren over de onderliggende microbiologische processen, we uit ons inzicht in de mechanica van gedrag niet hoeven te besluiten dat de vrije wil een illusie is. Daar kan weinig tegen in worden gebracht, behalve dat dit natuurlijk niet is wat de incompatibilisten willen zeggen.

Het probleem met Dennetts benadering van de vrije wil, is dat hij dit probleem vanuit verschillende perspectieven te lijf gaat - zijn hoofdstukken over determinisme zijn daar één van - maar geen enkel van deze benaderingen doorzet tot het eind. Het gevolg hiervan is dat je als lezer altijd met je vragen blijft zitten op het moment waarop het interessant wordt.

Zo eindigt zijn bespiegeling over memen even abrupt als die over de demon van Laplace (waarbij we het moesten doen met de gedachte dat de toekomst niet vaststond “vanuit het subjectieve gezichtspunt”, cfr. supra), en zet hij in de daarop volgende hoofdstukken een boom op over het morele instinct en bevindingen uit de speltheorie.

Toch lijkt het me nuttig om bij deze memen nog even te blijven stilstaan. Er is immers een groot verschil tussen genen en memen, namelijk dat we van een gen min of meer precies weten hoe het is opgebouwd. Genen zijn een reductionistische droom: we zijn zeer goed op weg om uiteindelijk iedere stap te kunnen verklaren van het sub-atomaire niveau naar het atomaire, en van daar naar het cellulaire. Nu we het menselijke genoom in kaart hebben gebracht, hebben we ook een schat aan inzicht gewonnen in hoe deze bouwstenen bijdragen aan de werking van die enorm complexe machines die we zelf zijn. Alleen heb je niet veel gezegd over de vrije wil, als je er enkel op kan wijzen dat we machines zijn die mettertijd geëvolueerd zijn tot machines die soms het welzijn van andere machines boven dat van zichzelf stellen. Eventueel wel over moraliteit, maar dat is iets anders.

We kunnen een poging doen om de eindjes die Dennett laat liggen aan elkaar te rijgen, door terug te keren naar dat subjectieve gezichtspunt, en het belang van de notie van het ‘ik’. Memen kunnen hun functie misschien daar vervullen, aangezien een meme enkel betekenis kan hebben op dat subjectieve niveau. Om betekenis te hebben, heeft een meme immers een interpretator nodig, een subject. En dat ‘ik’ wordt, samen met die memen, gerealiseerd door de dataprocessen in ons brein - Dennett stelt zelfs dat het ‘ik’ is opgebouwd uit memen. We blijven echter met een gat zitten als we willen overgaan van het mechanische niveau van het brein naar een niveau waarop gedachten (memen) betekenis hebben. En dit is geen onbelangrijk gat, aangezien memen de sleutel lijken te vormen voor het soort van vrije wil dat Dennett wil verdedigen: onze handelingen moeten op zijn minst het gevolg kunnen zijn van onze gedachten, *op een niveau waarop ze betekenis hebben*. Dit is de enige manier waarop Dennett kan ontsnappen uit het zuiver mechanische, en bovendien erkent hij een subjectief niveau waarop causaliteit werkzaam is (cfr. supra). Hij weet alleen niet hoe hij er moet geraken. Zo blijven memen binnen zijn filosofie de hemelhaken die hij zo

graag wilde vermijden. Om dit gat te dichten zouden we ons moeten wenden tot de filosofen die zich bezig houden met de metafysica van informatie (zie bvb. de pogingen van David Chalmers). Maar ook door zijn causaliteitsconcept verder uit te werken zou Dennett misschien de overgang kunnen maken die hij nodig heeft. Hij hoeft het zelfs niet verder te zoeken dan de tegenfeitelijke theorie van David Lewis die hij zelf al aanhaalde. Hoe hij deze zou kunnen aanwenden heeft David Chalmers al geïllustreerd (cfr. supra). Zo zagen we dat binnen alle mogelijke werelden waarin een gedachte onlosmakelijk deel uitmaakt van een bepaalde neurologische toestand, onze daaropvolgende handeling evenzeer het gevolg zal zijn van de gedachte als van die fysieke toestand. Dit vooronderstelt natuurlijk een tweevoudige aspect-theorie (*double aspect theory*). Maar ook Jaegwon Kim komt vanuit een hele andere hoek tot vergelijkbare conclusies voor wat betreft die mentale toestanden die functioneel analyseerbaar zijn. We zouden van Dennetts hemelhaken dus kranen kunnen maken. Of we daarmee de vrije wil hebben gered, is nog een andere vraag. Maar zoals Dennett waarschijnlijk zou antwoorden: dat hangt er maar van af wat je bedoelt met ‘vrije wil’.

4.2 Jaegwon Kim

Een filosoof die zich gedurende zijn carrière expliciet heeft toegelegd op het probleem van mentale veroorzaking, is Jaegwon Kim. Hierbij steunt hij op een zorgvuldig uitgewerkt metafysisch begrippenkader. Zijn werk rond superveniëntie en causaliteit is er dan ook vaak specifiek op gericht om het probleem van mentale veroorzaking op te lossen, maar is ook buiten deze context van grote waarde door de nauwgezetheid waarmee hij niet alleen zijn argumentatie, maar ook dat metafysische kader steen voor steen opbouwt tot een bouwwerk dat je niet zomaar aan het wankelen krijgt.

Door de jaren heen heeft Kim diverse bewustzijnstheorieën verdedigd, zonder echter ver af te wijken van wat hij in één van zijn latere werken *"Physicalism, or something near enough"* noemt. Toch is hij een filosoof die, zoals Chalmers zou zeggen, "het bewustzijn serieus neemt", en naarstig naar een oplossing probeert te zoeken waarbij we niet alleen een fysicalistisch wereldbeeld, maar ook de waarde van ons innerlijke leven kunnen behouden. Hiervoor is het noodzakelijk dat we de verbinding kunnen maken tussen de twee, en zullen we dus moeten kunnen verdedigen dat mentale veroorzaking mogelijk is. Dat dit geen gemakkelijke taak is, is duidelijk. Het belang ervan schetst Kim zelf treffend in zijn *Mind in a Physical World*:

> *"First, the possibility of human agency evidently requires that our mental states – our beliefs, desires, and intentions – have causal effects in the physical world: in voluntary actions our beliefs and desires, our intentions and decisions, must somehow cause our limbs to move in appropriate ways, thereby causing the objects around us to be rearranged. [...] Second, the possibility of human knowledge presupposes the reality of mental causation: perception, our sole window on the world, requires the causation of perceptual experiences and beliefs by physical objects and events around us. Reasoning, by which we acquire new knowledge and belief from the existing fund of what we already know or believe, involves the causation of new belief by old belief; more generally, causation arguably is essential to the transmission of evidential groundedness. Memory is a complex causal process involving interactions between experiences, their physical storage, and retrieval in the form of belief.*

> *If you take away perception, memory and reasoning, you pretty much take away all human knowledge. [...] The problem of determinism threatens human agency, and the challenge of skepticism threatens human knowledge. The stakes seem even higher with the problem of mental causation, for this problem threatens to take away both agency and cognition."* (Kim, 1998, p. 31-32)

Er staat dus heel wat op het spel.

Kims meest bekende ideeën in het debat rond mentale veroorzaking betreffen de causale exclusie van het mentale (*causal exclusion*), het superveniëntie-argument dat dit onderbouwt (*supervenience-argument*), en in mindere mate het idee van het 'leeglopen' van causale krachten (*causal drainage*) dat hier volgens critici (zie: Block, 2003) het gevolg van zou zijn.[38] Zoals Kim zelf benadrukt, is superveniëntie op zich nog geen bewustzijnstheorie. De superveniëntie van het mentale wordt immers geponeerd door vele, soms tegenover elkaar staande posities, zoals emergentisme, monistisch fysicalisme, en epifenomenalisme. (Kim, 1998, p. 12) We kunnen pas van een bewustzijnstheorie spreken zodra we de vraag proberen te beantwoorden *waarom* de superveniëntierelatie geldt tussen het fysieke en het mentale, en wat de aard is van deze relatie. Toch kan de superveniëntie van de geest over het lichaam dienst doen als een waardevolle scheidingslijn: door haar regel 'geen mentaal verschil zonder een fysiek verschil', definieert ze immers een 'minimaal fysicalisme', aldus Kim. (Kim, 1998, p. 15)

Kim gebruikt zijn superveniëntie-argument vervolgens om te pleiten voor een niet-reductionistisch fysicalisme. Door de causale overdeterminatie die zou optreden indien we aan het mentale causale krachten zouden toeschrijven bovenop deze van haar superveniëntiebasis (*supervenience base*), besluit hij dat alle mentale eigenschappen die functioneel analyseerbaar zijn, ook reduceerbaar zijn tot hun onderliggende fysieke

[38] Kort gezegd zou dit inhouden dat hogere-orde fenomenen nooit causale krachten zouden kunnen hebben die niet al gerealiseerd worden door hun lagere-orde superveniëntiebasis (zoals Kim met betrekking tot het mentale zal stellen). Dit zou een probleem vormen voor alle causaliteit tot op het meest microscopische niveau, en misschien zelfs daar voorbij, als we geen 'minimumniveau' aantreffen. Kim is het hier niet mee eens, en dient Block van antwoord in dezelfde uitgave (Kim, 2003). Het superveniëntie-argument geldt immers niet zomaar voor alle fenomenen die in een micro-macroverhouding staan tot elkaar, maar enkel voor de realisatie-ordes zoals Kim ze onderscheidt. (cfr. infra bij de kritiek van Woodward)

realisatoren (*realizers*). Qualia zijn echter niet op deze manier analyseerbaar, en zorgen ervoor dat het mentale domein nooit volledig gereduceerd kan worden tot het fysieke. (Kim, 2005, p. 29) In wat volgt zal ik Kims superveniëntie-argument en zijn conclusies uitgebreider bespreken.

4.2.1 Kim en Causaliteit

Hoewel Kim er in zijn vroegere werk wel aandacht aan besteedt – *Supervenience and Mind* (Kim, 1993) wijdt er een korte sectie aan – verdedigt hij in zijn latere publicaties geen uitgebreide causaliteitstheorie. Het is dus niet helemaal duidelijk hoe we de 'causale krachten' die door het superveniëntie-argument worden uitgesloten, precies moeten denken.

In *Supervenience and Mind* (Kim, 1993) verwijst hij niet alleen naar Hume, maar ook naar J.A. Foster, Patrick Suppes en J.L. Mackie. *"I find the preceding two accounts of Humean causation (contiguous causation and the account that takes cause as essentially a relational event) attractive,"* besluit hij dan, maar voegt daar meteen aan toe:

> *"It is best, to look upon the tentative accounts of Humean causation in this section not as a full-fledged analysis of causation, but rather as approximations to the broader notion of subsumption of events under a law, an idea that forms the foundation of the Humean, or nomological, approach to causation."* (Kim, 1993, p. 21)

In ieder geval lijkt het idee van noodzakelijkheid een belangrijk onderdeel van causaliteit zoals Kim die in gedachten heeft:

> *"Causation is a preeminent example of what I am calling determinative or* ***dependency*** *relations; apart from those that are logically based, such as entailment, it is the only explicitly recognized and widely discussed relation of this kind. Causes determine their effects, and effects are* ***dependent****, for their existence and properties, on their causes."* (Kim, 1993, p. 54, mijn vetjes)

De nadruk die Kim hier legt op afhankelijkheid doet denken aan de tegenfeitelijke theorie van David Lewis. Die werd ooit door Kim besproken in een korte paper (Kim,

1973) die hij besloot met de bedenking dat hij voor een beter oordeel meer zou moeten weten over de rol van wetten in Lewis' theorie.

Toch blijkt in Kims latere werk dat ook de nomologische benadering niet streng genoeg is om het soort van causaliteit te garanderen waar hij naar op zoek is: zowel Humeaanse als Hempeliaanse causaliteit worden platgemaaid door de zeis van zijn superveniëntie-argument. (cfr. infra) De correlaties die bestaan tussen het mentale en haar superveniëntiebasis zijn immers geen *echte* causaliteit, hoewel ze wetmatig zijn en tegenfeitelijk standhouden. (Kim, 2005, p. 21)

Voor *echte* causaliteit is er dus nog iets meer nodig. De vraag is alleen, wat dan precies? Heeft Kim zich hier bekeerd tot een mechanistische causaliteitsopvatting à la Wesley Salmon? (Psillos, 2002, p. 107 e.v.) In *Physicalism, or Something Near Enough*, vult Kim deze schets verder aan:

> *"The intuitive idea is the idea of an event or state, or a property instantiation, owing its existence to another event or state – or, to put it another way, the idea that one thing is generated out of, or derives its existence from, another. [...] Causation as generation, or effective production and determination, is in many ways a stronger relation than mere counterfactual dependence, and it is causation in this sense that is fundamentally involved in the problem of mental causation."* (Kim, 2005, p. 18)

De ruimte die Kim laat voor interpretatie van zijn causaliteitsconcept (of voor het naast elkaar bestaan van verschillende vormen van causaliteit, zoals het bovenstaande citaat lijkt te suggereren), stelt hem open voor kritiek van o.a. James Woodward, in diens *'Mental Causation and Neural Mechanisms'* (Hohwy en Kallestrup, 2008) en een tot nog toe ongepubliceerde paper waarin hij zich meer specifiek tot Kim richt (Woodward). Maar ook Carl Craver uit in zijn boek bedenkingen bij de onduidelijkheid van de benaming 'causale krachten':

> *"I am adopting here the notion of a 'causal power' because it is part of Kim's argument. Causal powers are sometimes described as if they need to be added*

> *into the world in addition to the entities if change is to be possible.*[39] *They are understood as forces that push and pull, attract and repel, bond and break bonds, restore equilibrium, and so on. Sometimes they are described merely as sufficient causes (as Kim seems to think of them).”*[40] (Craver, 2009, p. 211)

De kritiek van Woodward bespreek ik verderop in dit deel.

4.2.2 Superveniëntie, realisatie en reductie

In *Mind in a Physical World* (Kim, 1998) definieert Kim de superveniëntierelatie tussen lichaam en geest als volgt:

> *“Mental properties* supervene *on physical properties, in that necessarily, for any mental property* M*, if anything has* M *at time* t*, there exists a physical base (or subvenient) property* P *such that it has* P *at* t*, and necessarily anything that has* P *at a time has* M *at that time.”* (Kim, 1998, p. 9, cursief in origineel)

Bovenstaande definitie is ook degene die hij in *Supervenience and Mind* geeft voor *sterke supervenientie*. (Kim, 1993, p. 80) Dat is dan ook de variant waar we ons in de huidige context op zullen concentreren. (Kim, 1998, p. 9)

Zoals we hebben gezien, komt zijn definitie hiervan ongeveer overeen met wat David Chalmers *‘logical supervenience’* noemt. Toch is er hier al meteen een belangrijk verschil. Herinner dat Chalmers stelde dat het mentale slechts ‘natuurlijk’ (*naturally*) supervenieert op het fysische/fysieke, maar niet logisch (*logically*). In een andere mogelijke wereld, met andere wetten, zou het volgens Chalmers kunnen dat het mentale *niet* op het fysische/fysieke supervenieert. (cfr. supra) Kim maakt hier meteen komaf mee:

> *“Mental properties* supervene *on physical properties, in that necessarily any two things (in the same or different possible worlds) indiscernible in all physical*

[39] Dit is wat David Chalmers poneert als hij suggereert dat causale krachten niet logisch superveniëren op het fysische domein, cfr. supra.

[40] Craver kiest hier dus voor een andere interpretatie van Kim dan ikzelf eerder deed. Toch denk ik (om bovengenoemde redenen) dat Kims idee van causaliteit sterker is dan enkel dat van een ‘voldoende’ (*sufficient*) oorzaak.

> *properties are indiscernible in mental respects."* (Kim, 1998, p. 10, cursief in origineel)

Dit hangt natuurlijk samen met zijn stelling dat de (sterke!) superveniëntie van het mentale op het fysieke 'minimaal fysicalisme' garandeert. Als we, zoals Chalmers, zouden volhouden dat het mentale enkel in deze mogelijke wereld supervenieert op het fysische, laten we de deur op een kiertje staan voor (zijn vorm van) dualisme.

Als ik me niet vergis, zou Chalmers ook problemen hebben met de stelling die Kim hierop laat volgen, namelijk dat de laatst gegeven definitie equivalent is aan de stelling: *"No mental difference without a physical difference."* (Kim, 1998, p. 10) Volgens Chalmers zou het immers kunnen dat *P* in een bepaalde wereld *M* realiseert, maar in een andere mogelijke wereld (bvb. in een universum waarin andere wetten gelden) M*. Dit is niet tegenstrijdig met de stelling dat er geen mentaal verschil kan zijn zonder een fysiek verschil binnen deze respectievelijke werelden. De voetnoot waarin Kim stelt dat zij die niet aanvaarden dat de beweringen equivalent zijn, in ieder geval akkoord zullen gaan dat wie de ene bewering aanvaardt, ook de andere zal aanvaarden, is dus niet helemaal juist. Chalmers aanvaardt de tweede bewering over onze wereld, maar niet de bewering die over verschillende mogelijke werelden gaat.

Het verschil tussen beide auteurs zit hem in de gradatie van noodzakelijkheid. Terwijl Kim bereid is om toe te geven dat mentale eigenschappen kunnen variëren wat betreft de modaliteit van hun superveniëntie op het fysieke - intentionele eigenschappen kunnen bijvoorbeeld logisch/conceptueel superveniëren, terwijl fenomenale eigenschappen slechts nomisch[41] superveniëren (Kim, 1998, p. 10) – is de manier waarop ze superveniëren altijd op zijn minst wetmatig. De zwakkere variant waar Chalmers het over heeft, behoort voor Kim niet tot de mogelijkheden voor het mentale.

Zoals Kim zelf opmerkt zegt het aanvaarden van een bepaalde superveniëntierelatie niet noodzakelijk iets over de reduceerbaarheid van het fenomeen in kwestie (cfr. supra). We kunnen sterke superveniëntie aanvaarden en toch verschillende meningen toegedaan zijn over de reduceerbaarheid van verschillende fenomenen – door

[41] Herinner dat we van Chalmers zijn 'logische superveniëntie' als 'nomologisch' mochten opvatten.

bijvoorbeeld te stellen dat biologische eigenschappen reduceerbaar zijn tot microchemische eigenschappen, maar een antireductionistische positie in te nemen met betrekking tot andere eigenschappen (zoals bewustzijn) en/of niveaus. (Kim, 1998, p. 17) Kim zal zijn conclusies over de reduceerbaarheid van mentale eigenschappen baseren op zijn superveniëntie-argument, dat betrekking heeft op de causale krachten (cfr. supra) van deze eigenschappen. Zoals we eerder bij David Chalmers zagen, kunnen we stellen dat de reduceerbaarheid van een fenomeen, afhangt van de functionele analyseerbaarheid ervan.

> *"Functionalism takes mental properties and kinds as functional properties, properties specified in terms of their roles as causal intermediaries between sensory inputs and behavioral outputs, and the physicalist form of functionalism* ***takes physical properties as the only potential occupants, or 'realizers', of these causal roles.****"* (Kim, 1998, p. 19, mijn vetjes)

Dit fysicalistische functionalisme noemt Kim "fysiek realisationisme" (*physical realizationism*). Hogere-orde-eigenschappen worden *gerealiseerd* door lagere-orde-eigenschappen:

> *"*F *is a* second-order property *over set* ***B*** *of base (or first-order) properties iff* F *is the property of having some property* P *in* ***B*** *such that* D(P)*, where* D *specifies a condition on members of* ***B****."* (Kim, 1998, p. 20, cursief en vetjes in origineel)

Een verhelderend voorbeeld heeft betrekking op kleuren: als de set **B** alle kleuren omvat, kunnen we stellen dat 'een primaire kleur hebben' een hogere-orde-eigenschap (*F*) is die hetzelfde is als het hebben van *P* in **B**, waarbij *P* = blauw of *P* = geel of *P* = rood. (Kim, 1998, p. 20)

Merk op dat de eigenschap 'een primaire kleur hebben' meervoudig realiseerbaar is. (cfr. supra) Dit geldt ook voor functionele eigenschappen. *Functionele* eigenschappen zijn hogere-orde-eigenschappen die gedefinieerd worden in termen van causale of nomische relaties tussen lagere-orde-eigenschappen. Elke basiseigenschap met de juiste causale/nomische relaties ten opzichte van andere eigenschappen kan dus een

functionele eigenschap realiseren, ongeacht haar verdere constitutie.[42] Voor zover mentale eigenschappen functionele eigenschappen zijn, zijn ze dus meervoudig realiseerbaar. (Kim, 1998, p. 21)

In dit geval hebben we ook te maken met een superveniëntierelatie:

> *"Suppose that P realizes M in systems of kind S. From the definition of realization, it follows that P is sufficient for M. [...] Thus, if ⟨P_1, . . . , P_n⟩ is a realization of ⟨M_1, . . . , M_n⟩, in the sense that each P_i is a realizer of M_i, it follows that the Ms are supervenient on the Ps."* (Kim, 1998, p. 23)

Fysiek realisationisme biedt ons dus een verklaring voor de superveniëntierelatie: het mentale supervenieert dan op het fysieke omdat mentale eigenschappen hogere-orde-eigenschappen zijn met fysieke realisatoren (en geen niet-fysieke). Het staat ons toe te zeggen dat het hebben van een hogere-orde eigenschap (*F*) "niets meer of minder" betekent dan het hebben van een lagere orde-eigenschap (*P*). Dit is verenigbaar met het meest courante model van functionele reductie in de wetenschap.[43] (Kim, 1998, p. 24)

4.2.3 Causale exclusie en het superveniëntie-argument

De superveniëntierelatie wordt niet alleen verklaard door fysiek realisationisme, ze volgt hier ook uit. Zoals we gezien hebben, zijn er ook andere posities die de superveniëntie van het mentale aanvaarden (zoals bijvoorbeeld emergentisme). Dit is belangrijk, aangezien Kim zijn argumentatie tegen mentale veroorzaking op deze superveniëntierelatie zal baseren. Niet alleen de 'minimale' fysicalisten waar Kim zichzelf toe rekent, maar iedere filosoof die de superveniëntie van het mentale op het fysieke aanvaardt, krijgt dus met dit argument af te rekenen als hij de mogelijkheid van mentale veroorzaking wil onderzoeken. Voor zij die resoluut voor fysicalisme kiezen,

[42] Dit heeft natuurlijk gevolgen voor de mogelijkheid van *'strong artificial intelligence'* (cfr. supra). Het zou immers betekenen dat de functionele eigenschappen in kwestie ook in een artificieel systeem gerealiseerd zouden kunnen worden. Hierbij moeten we natuurlijk de bedenking maken dat Kim het bestaan aanvaardt van fenomenologische bewustzijnscomponenten (qualia) die niet functioneel zijn. Dit zou betekenen dat voor hem, in tegenstelling tot voor Dennett, een machine nooit dezelfde graad van bewustzijn kan bereiken als die van ons.

[43] Het komt ook overeen met Chalmers' model van reductionistische verklaringen (cfr. supra).

zijn er echter alvast enkele principes die zij niet naast zich neer kunnen leggen. Het eerste daarvan is dat de fysische wereld een causaal gesloten domein constitueert:

> *"If a physical event has a cause at time* t*, then it has a physical cause at* t*."* (principle of causal closure, Kim, 2005, p. 15)

Het equivalent hiervan met betrekking tot verklaringen is:

> *"If a physical event has a causal explanation (in terms of an event occurring at* t*), it has a physical causal explanation (in terms of a physical event at* t*)."* (Kim, 2005, p. 16)

Omwille hiervan, spreekt men ook over '*explanatory exclusion'*.

Kim wijst er op dat het principe van causale geslotenheid van het fysische domein perfect compatibel is met een dualistische opvatting over lichaam en geest, zelfs met substantiedualisme. Het enige dat het stipuleert, is dat er geen oorzaken van buitenaf het fysische domein binnen kunnen dringen. Het interactionisme van Descartes wordt er dus wel door uitgesloten. De theorie van Leibniz (cfr. supra) echter niet. (Kim, 2005, p. 16)

Het tweede principe dat Kim vooropstelt is er één dat overdeterminatie uitsluit:

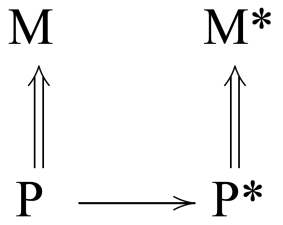
> *"If an event* e *has a sufficient cause* c *at* t*, no event at* t *distinct from* c *can be a cause of* e*."* (principle of causal exclusion, Kim, 2005, p. 17)

Het superveniëntie-argument gaat nu als volgt (ik vat zijn versie samen uit *Physicalism, or Something Near Enough* (Kim, 2005)):

Stel dat een mentale eigenschap *M*, een andere mentale eigenschap *M** veroorzaakt. Dit is een voorbeeld van causaliteit die zich volledig binnen het mentale domein afspeelt, en is dus, zoals we zagen, verenigbaar met het principe van causale geslotenheid van het fysische domein. De superveniëntie van het mentale op het fysieke stelt nu echter dat *M** zich enkel kan voordoen omdat één van de fysieke eigenschappen waarop *M** supervenieert, zich voordoet; noem dit de fysieke

basiseigenschap *P**. Hier komt het principe van exclusie op de proppen[44]: Is het optreden van *M** nu te wijten aan *M*, of aan *P**? Kim stelt voor dat de enige coherente manier om ons dit voor te stellen is dat *M*, *M** veroorzaakt, *door* P*** te veroorzaken*. Hier hebben we dus echt te maken met causaliteit die van het mentale naar het fysieke domein verloopt. M moet echter zelf een fysieke superveniëntiebasis hebben, namelijk P. Als we dit nader bekijken, kunnen we echter vaststellen dat *P* op zich voldoende is voor *P**. Het resultaat hiervan is overdeterminatie van P* door M en P. Omwille van het principe van causale geslotenheid echter, moeten we M uitsluiten als oorzaak van P*, en dus ook van M*.

Visueel voorgesteld[45]:



Figuur 1

> *"The final picture that has emerged is this: P is a cause of P*, with M and M* supervening respectively on P and P*. There is a single underlying causal process in this picture, and this process connects two physical properties, P and P*. The correlations between M and M* and between M and P* are by no means accidental or coincidental; they are lawful and counterfactual-sustaining regularities arising out of M's and M*'s supervenience on the causally linked P and P*. These observed correlations give us an impression of causation; however, that is only an appearance, and there is no more causation here than*

[44] In *Physicalism, or Something Near Enough* breidt Kim het principe van causale exclusie uit met een principe van "determinerende/generatieve exclusie", en het is datgene waar hij hier naar verwijst. Dit is belangrijk omdat de kritiek van Woodward zich vooral hierop zal richten: de relatie van P* naar M* is immers geen causale relatie, maar een superveniëntierelatie. Toch liet ik het eerder weg omdat dit niet is hoe het superveniëntie-argument door de band genomen wordt geïnterpreteerd. Volgens de klassieke interpretatie die enkel rekening houdt met causale exclusie hebben we hier dus nog geen exclusieprobleem, maar pas later.

[45] Ik heb deze figuur overgenomen uit (Kim, 2005), maar waar Kim identieke pijlen gebruikt met de woorden *'causes'* en *'supervenes'* heb ik deze veranderd in respectievelijk enkele en dubbele pijlen, om later beter te kunnen vergelijken met Woodward.

> *between two successive shadows cast by a moving car, or two successive symptoms of a developing pathology."* (Kim, 2005, p. 21)

Het gevolg hiervan is dat het superveniëntie-argument niet alleen causaliteit van het mentale naar het fysieke, maar ook van het mentale naar het mentale domein uitsluit. Alle causaliteit speelt zich af op het fysieke niveau, en het mentale heeft geen causale krachten die het hier aan toevoegt. Terugkerend naar de definities van fysieke realisatie, wil dit zeggen dat we mentale eigenschappen niet anders functioneel kunnen analyseren dan door te verwijzen naar de causale eigenschappen van de onderliggende fysieke basis, en we ze daar volledig tot kunnen reduceren. Het goede nieuws dat daaruit volgt, zo stelt Kim, is dat het mentale (onze wensen en overtuigingen) in dit geval zeer reële causale krachten heeft. Het slechte nieuws, dat deze niets meer zijn dan de causale krachten van het fysieke. (Kim, 1998, p. 118)

Maar het *echte* slechte nieuws, zo gaat hij verder, is dat er bepaalde mentale eigenschappen zijn, namelijk de fenomenale eigenschappen van bewuste ervaring, die aan functionalisatie lijken te weerstaan.[46] Dit wil zeggen dat er geen enkele manier is om hier binnen een fysicalistisch wereldbeeld causale eigenschappen aan toe te schrijven. Toch kan hij niet zeggen dat een vorm van dualisme hier enige hulp biedt. Zelf is hij echter niet vertrouwd met wat zich in de "duistere spelonken van het dualisme verschuilt".[47] (Kim, 2005, p. 120)

Met de titel van Kims volgende boek *Physicalism, or Something Near Enough* mag het echter duidelijk wezen welke richting hij gekozen heeft.

4.2.4 Woodward versus het superveniëntie-argument

In *'Mental Causation and Neural Mechanisms'* (Woodward in Hohwy en Kallestrup, 2008) zet Jim Woodward uiteen waarom hij vindt dat veel van de gangbare argumenten voor de causale inertie van het mentale gebaseerd zijn op een verkeerde

[46] Dit vindt ook Chalmers (cfr. supra). Kim herhaalt zijn standpunt over de niet-reduceerbaarheid (want niet functioneel analyseerbaarheid) van qualia in *Physicalism, or Something Near Enough* (Kim, 2005, p. 22 e.v.)

[47] Bij deze zin verwijst Kim in een voetnoot specifiek naar Chalmers, voor wie hij minstens een zeker respect lijkt te hebben. Ook in zijn volgende boek verwijst hij hier en daar naar hem.

opvatting van wat een relatie causaal maakt, en wat er precies komt kijken bij een mentale verklaring.

> *"These mistaken assumptions involve [...] a conception of causation according to which a cause is simply a condition (or a conjunct in a condition) which is nomologically sufficient for its effect, and the closely associated deductive nomological (DN) conception of explanation according to which explaining an outcome is simply a matter of exhibiting a nomologically sufficient condition for it."* (Hohwy en Kallestrup, 2008, p. 218-219)

Hoewel ook Craver Kims causaliteitsopvatting zo lijkt te interpreteren als enkel een nomologisch voldoende voorwaarde (of voldoende grond), hebben we eerder gezien dat het zou kunnen dat Kim er verschillende causaliteitsconcepten op nahoudt (waarbij de verschillende soorten causaliteit eventueel naast elkaar kunnen bestaan) en in ieder geval een sterker causaliteitsconcept verdedigt in gevallen van mentale veroorzaking (cfr. supra). Toch denk ik niet dat dit het essentiële punt is waarop Woodward en Kim met elkaar verzoend moeten worden. Wel denk ik dat Woodward Kim misschien fout interpreteert met betrekking tot de conclusies die volgen uit het superveniëntie-argument met betrekking tot de causale krachten van het mentale, en vermoed ik dat een interventionistische vertaling van Kims argument dit argument niet weerlegt, maar juist bevestigt. Tenslotte sluit ik me aan bij Carl Craver, die erop wijst dat het belangrijk is om goed het onderscheid te blijven zien tussen 'causale krachten' enerzijds en 'causale relevantie' anderzijds.

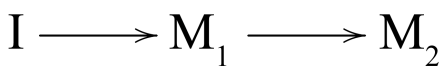
Woodward begint zijn betoog in *'MCNM'* met de suggestie dat een interventionistische causaliteitsopvatting op het eerste gezicht het idee lijkt te bevestigen dat mentale toestanden oorzaken kunnen zijn. We proberen immers voortdurend om de mentale toestanden (overtuigingen, wensen, etc.) van anderen te beïnvloeden om hen daardoor tot een ander gedrag te brengen. Ook in psychologische experimenten is dit de gebruikelijke gang van zaken.

> *"That is, all that is required for changes in a mental state* M_1 *to cause changes in a second mental state* M_2 *(or in behavior* B*) is that it be true that under some intervention that changes* M_1*,* M_2 *(or* B*) will change. Common sense certainly*

supposes that episodes like these are very widespread." (Woodward in Hohwy en Kallestrup, 2008, p. 231)

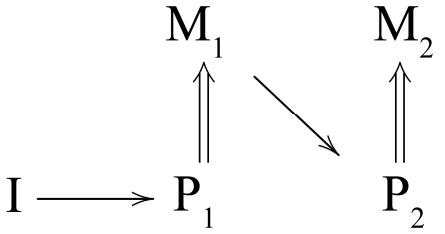
Wie echter nog Kims schematische weergave van dit verhaal in gedachten heeft, ziet onmiddellijk waar het probleem hier zit. De fysieke superveniëntiebasis voor respectievelijk M_1 en M_2, ontbreekt immers in dit plaatje.

Omdat hierdoor de indruk wordt gewekt dat we een rechtstreekse interventie kunnen uitvoeren op M_1, is dit een **foutieve** weergave.[48]:



Figuur 2

Als we deze wel toevoegen zoals Kim ze voorstelt, krijgen we een veel ingewikkelder verhaal:



Figuur 3

De belangrijkste zaken die deze figuur ons leert, zal ook Woodward benadrukken in de rest van zijn paper. Het eerste dat ons opvalt, is dat (a) het onmogelijk is om een interventie uit te voeren op M_1 die niet via P_1 verloopt. Dit is niet alleen te wijten aan het voor de hand liggende feit dat we iemands gedachten enkel via fysieke weg kunnen beïnvloeden (door in te werken op de zintuigen, of door de hersenen rechtstreeks te manipuleren). Ook als we in telepathie zouden geloven, zou het per definitie onmogelijk zijn om een verandering in M_1 tot stand te brengen zonder een verandering in P_1, omwille van het feit dat (b) de relatie tussen P_1 en M_1 geen causale, maar een superveniëntierelatie is. Over deze twee zaken zijn Woodward en Kim het alvast volmondig eens. Dit heeft als gevolg dat (c) we weer het superveniëntie-argument kunnen toepassen, en een pijl trekken van P_1 naar P_2 die de pijl van M_1 naar P_2 overbodig maakt (niet afgebeeld). Met dit laatste is Woodward het echter niet eens,

[48] Ik laat het gedrag *B* hier even buiten beschouwing.

en presenteert dan ook zijn hierop volgende uiteenzetting als een gedeeltelijke kritiek op Kim. Vermoedelijk is dit het gevolg van zijn interpretatie van Kims conclusies, die anders is dan de mijne. Kim beweert immers niet dat het mentale in het geheel geen causale krachten kan hebben, wel dat de causale krachten van het mentale gelijk zijn aan die van haar fysieke superveniëntiebasis:

> *"On a reductionist position of this sort*[49] *, however, the causal powers of mental properties turn out to be* ***just those of their physical realizers****, and there are* ***no new*** *causal properties brought into the world by mental properties."* (Kim, 1998, p. 118, mijn vetjes)

Volgens Woodward wordt dit echter:

> *"As we have seen, many philosophers worry that if there are mental causes, then this would require a bizarre and implausible kind of over-determination – the physical states that realize the causal effects of mental states would be caused both by mental states and by physical states."* (Woodward in Hohwy en Kallestrup, 2008, p. 258)

Wat Kim betreft, is het bovenstaande enkel een correcte weergave van zijn stelling als we na de woorden *'mental causes'* zouden toevoegen *'over and above those of their physical realizers'*.

Woodward geeft hierna twee voorbeelden die hij vrij uitgebreid bespreekt. Het eerste heeft betrekking op het gedrag van een mol gas:

> *"Suppose that a mole of ideal gas at temperature* T_1 *and pressure* P_1 *at time* t_1 *is confined to a container of fixed volume* V*. The temperature of the gas is then increased to* T_2 *by the application of a heat source and the gas is allowed to reach a new equilibrium at time* t_2 *where its pressure is found to have increased to* P_2*."* (Woodward in Hohwy en Kallestrup, 2008, p. 233)

[49] Kim bedoelt een reductie door functionalisering. Herinner dat Kim ruimte laat voor niet-functionaliseerbare eigenschappen (qualia), maar hier geen causale eigenschappen aan toeschrijft. Zijn besluit over het uitblijven van nieuwe causale eigenschappen gaat dus op voor het volledige mentale domein.

Dit voorbeeld heeft als doel ons ervan te overtuigen dat microverklaringen niet altijd superieur zijn aan macroverklaringen: bij een *macroverklaring* van het bovenstaande voorbeeld kunnen we ons immers beroepen op de ideale gaswet die betrekking heeft op macroscopische variabelen zoals druk, volume en temperatuur – een *microverklaring* zou echter het gedrag van individuele moleculen moeten bespreken.
Dit is natuurlijk waar, al is niet precies duidelijk wat de relevantie is van dit voorbeeld. Toegegeven, Woodward schrijft aan Kim in het bijzonder geen principiële voorkeur voor microverklaringen toe. Wel laat hij het voorbeeld voorafgaan door het volgende:

> *"One motivation for skepticism about assigning any causal role to the mental derives from the assumption that mental states are 'multiply realizable' by different neural or physical states, combined with the thought that there is a general preference for detailed or fine-grained or more micro level causal claims/explanations (in this case claims at some physical or neural level) over less fine-grained, more macro (e.g. mental or psychological) claims."* (Woodward in Hohwy en Kallestrup, 2008, p. 232)

Afgezien van het eerder aangehaalde verschil dat Kim het mentale niet *'any causal role'* ontzegt, zit in het bovenstaande citaat naar mijn mening een tweede belangrijke verschil met Kim, dat meteen aantoont waarom het voorbeeld in kwestie irrelevant is, en bovendien niet onderhevig aan het superveniëntie-argument.
Kim maakt immers een onderscheid tussen niveaus van realisatie (wat hij 'ordes' noemt) en micro-macro-niveaus, hoewel Woodward in het bovenstaande enkel rekening lijkt te houden met die gevallen waarin fenomenen op macroniveau worden gerealiseerd door fenomenen op microniveau. Deze gevallen doen zich zeker veelvuldig voor, alleen benadrukt Kim dat het exclusie-argument niet van toepassing is op niveaus die in zo'n micro-macro-verhouding tot elkaar staan. Dit moet hij ook doen, om te voorkomen dat Ned Blocks idee van *'causal drainage'* (cfr. supra) een legitieme vrees zou zijn:

> *"This means that the supervenience argument, which exploits the supervenience relation, does not have the effect of emptying macrolevels of causal powers and rendering familiar macro-objects and their properties causally impotent."* (Kim, 1998, p. 86)

Op het belang van het onderscheid tussen deze verschillende opvattingen van de notie 'niveau' (en de grote variatie in mogelijke betekenissen), wordt uitgebreid ingegaan door Carl Craver in zijn *Explaining the Brain* (2009). Hij vat het verschil met realisatieniveaus (of –ordes) als volgt samen:

> *"In levels of realization, a property or activity at a higher level is realized by a property or activity at a lower level of realization. The item at a lower level of realization is not part of the item at a higher level;* ***the realized and realizing properties are properties of the same thing."***[50] (Craver, 2009, p. 165, mijn vetjes)

Het staat natuurlijk open voor discussie of 'het' mentale zich op macroniveau bevindt terwijl haar realisatie plaatsvindt op microniveau. Toch lijkt me dit geen vruchtbare manier om dit te bekijken, en is Kim het hier zeker niet mee eens.

Het tweede voorbeeld dat Woodward (maar ook Craver) bespreekt, is er één dat hij ontleent aan Stephen Yablo (1992) en waarin een duif geleerd heeft om op een plek te pikken wanneer deze plek rood is. In het geval in kwestie blijkt de duif herhaaldelijke malen te pikken naar een stimulus die een scharlaken kleur heeft. Woodward vraagt ons vervolgens om twee stellingen in overweging te nemen:

(a) De duif pikt omdat de stimulus scharlaken is.

(b) De duif pikt omdat de stimulus rood is.

Dit is een goed voorbeeld omdat hier wel sprake is van de realisatierelatie zoals Kim die in gedachten heeft, getuige zijn eigen voorbeeld dat betrekking had op primaire kleuren (cfr. infra). Hoewel Woodward ook hier een micro-macroverschil in onderscheidt, maakt hij wel duidelijk dat het hier om een verschil in *verklarings*niveaus gaat. Dan vraagt hij de lezer om een keuze te maken tussen de volgende twee zinnen:

(a*) Dat de duif pikt in plaats van niet te pikken is omdat de stimulus scharlaken is in plaats van niet scharlaken.

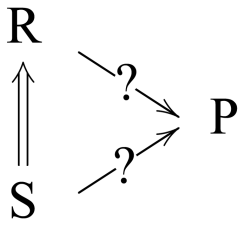
[50] Intuïtief lijkt het mogelijk om al vanuit deze beschrijving te argumenteren dat het mentale en het fysieke geen causale krachten hebben die zij niet met elkaar delen, aangezien ze eigenschappen zijn van dezelfde zaak. Dan komen we echter in neutraal monistisch vaarwater terecht, of misschien wel in dat van Chalmers' pan-proto-psychisme (cfr. supra).

(b*) Dat de duif pikt in plaats van niet te pikken is omdat de stimulus rood is in plaats van niet rood.

Het kan toch niet, zo vervolgt hij, dat we de voorkeur zouden geven aan (a*), enkel maar omdat deze beschrijving specifieker is (dit noemt Woodward het microniveau in dit voorbeeld). Toch is het presenteren van een scharlaken stimulus *voldoende* om de duif te doen pikken. Dit illustreert echter alleen dat er meer nodig is voor het succesvol verklaren van causale claims dan enkel het geven van nomologisch voldoende voorwaarden, aldus Woodward. (Woodward in Hohwy en Kallestrup, 2008, p. 135-6)

Een belangrijke opmerking die we hier moeten maken, is dat, zoals eerder gezegd, Kim het heeft over 'causale krachten', terwijl Woodward en Craver zich bezig houden met 'causale relevantie'. Als we het voorbeeld in Kims terminologie beschouwen, zullen we ons dan ook afvragen of de rode, dan wel de scharlaken stimulus causale krachten bezit die de andere niet heeft. Woodward lijkt er bij dit voorbeeld van overtuigd dat het 'rood zijn' van de stimulus relevanter is dan het 'scharlaken zijn'.

Beide denkers zijn het met elkaar eens dat in dit geval het 'rood' zijn van de stimulus wordt gerealiseerd door haar 'scharlaken zijn'. Wat dat betreft is het dus een goede analogie met Kims beschrijving van de relatie tussen het mentale en haar realisatoren, en kunnen we het voorbeeld weergeven in een gelijkaardig diagram:

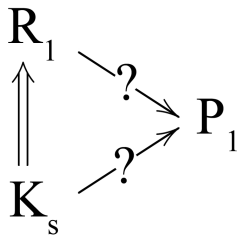


Figuur 4

De vraag die Woodward hier stelt, is: "Wordt het pikken *P* veroorzaakt door het 'scharlaken zijn' *S* of door het 'rood zijn' *R*?"

Om deze vraag op te lossen, kunnen we geen beroep kunnen doen op een principe van causale geslotenheid om één van beide kandidaat-oorzaken uit te sluiten. Als we bij Kim blijven, hebben we dus eventueel een probleem.

Als we het bovenstaande diagram willen 'vertalen' naar de theorie van Woodward, moeten we een aantal belangrijke aanpassingen doen om duidelijk te maken dat de vraag in het bovenstaande verkeerd wordt gesteld. Zoals we eerder zagen, werkt Woodward immers met variabelen. 'Rood' en 'scharlaken' zijn in het bovenstaande voorbeeld echter geen variabelen, maar *waarden van variabelen*. Ik stel zelf even een figuur op die hier rekening mee houdt:

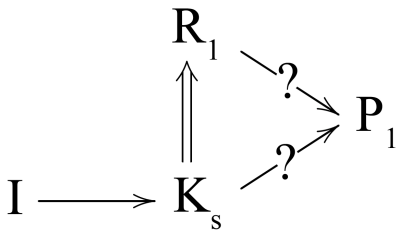


Figuur 5

In ons gecorrigeerde diagram blijkt dat *R* twee waarden kan aannemen: 'rood' (R_1) en 'niet rood' (R_0). Dit komt omdat de duif ook maar twee toestanden heeft die voor ons voorbeeld relevant zijn, namelijk 'pikken' (P_1) of niet pikken (P_0). *K*, of 'kleur', kan echter een oneindig aantal waarden aannemen, over het volledige kleurenspectrum. Dit komt omdat *beide waarden van R* meervoudig realiseerbaar zijn: zo zijn 'turkoois' (K_t), 'lila' (K_l) en 'oker' (K_o) enkele mogelijke realisatoren van 'niet rood' (R_0), en zijn 'vermiljoen' (*Kv*), 'scharlaken' (*Ks*) en 'bordeaux' (*Kb*) enkele mogelijke realisatoren van 'rood' (*R1*). Bovendien is de superveniëntiebasis van *R*, namelijk het kleurenspectrum *K*, in principe oneindig, aangezien het een continuüm betreft.

In dit geval heeft *K* de waarde 'scharlaken'. *R* supervenieert op *K*; dit wil zeggen dat we niet kunnen veranderen of een stimulus rood of niet-rood is zonder de concrete kleur van de stimulus te veranderen.[51] Het directe gevolg hiervan is dat het onmogelijk is om een directe interventie te doen op *R*, die niet via *K* verloopt (we kunnen immers niet de 'roodheid' van een stimulus veranderen, zonder eerst de concrete kleur van de stimulus te veranderen):

[51] Dit maakt ons voorbeeld tot een goede analogie voor dat waarin het mentale supervenieert op het fysieke, met dit verschil dat *M* natuurlijk geen binaire variabele is, en ook ons gedrag niet beperkt is tot al dan niet pikken!



Figuur 6

Dit begint alweer sterk te lijken op het diagram dat we hadden in figuur 3. Hoe weten we nu of *P* veroorzaakt wordt door *R* of door *K*? Volgens Woodward moeten we dit uitvissen door een interventie te doen op de ene variabele, terwijl we de andere constant houden. Alleen kunnen we dit enkel uitvoeren voor *K*, maar niet voor *R*: omwille van de superveniëntierelatie tussen *K* en *R* is dit laatste onmogelijk.

Zoals Woodward uiteenzet in zijn latere paper (Woodward), hoeft dit ook geen probleem te zijn, aangezien het een foute veronderstelling is dat we bij een interventie op een variabele moeten proberen om haar superveniëntiebasis constant te houden:

> *"[...]one can't simply assume that because it is appropriate to control for ordinary confounders in cases in which no non-causal dependency relations are present, it must also be appropriate to control for factors like supervenience bases which do represent non-causal dependency relations."* (Woodward, p. 29)

Het wordt op deze manier wél moeilijk om op een zinnige manier een onderscheid te maken tussen de causale invloed van *K* en *R*, aangezien we, juist omdat de relatie ertussen niet causaal is, we deze relatie niet 'uit elkaar kunnen trekken' (Woodward spreekt over *'arrow breaking'* – in dit geval kunnen we zeggen dat we de pijl van *K* naar *R* niet kunnen breken omdat het een dubbele pijl is!). Woodward ziet dit als een argument tegen Kim: aangezien het zinloos is om te 'controleren' op superveniëntierelaties, is het zinloos om te stellen dat een superveniërende eigenschap geen extra causale krachten kan hebben. Volgens mij is het echter even zinloos om te stellen dat ze die wél heeft. De kern van het probleem zit hier echter in het onderscheid tussen 'causale krachten' en 'causale relevantie'.

Als we het in termen van causale relevantie willen stellen, is er volgens Carl Craver (2009, p. 206 e.v.) maar één manier om na te gaan welke de meest relevante eigenschap is (of volgens Craver: wat het *'switch point'* zal zijn om over te gaan van

'niet pikken' naar 'pikken'): we zullen de duif een grote hoeveelheid stimuli in verschillende kleuren moeten voorleggen en het gedrag van de duif statistisch analyseren. Als de duif zonder onderscheid blijkt te pikken bij stimuli die vermiljoen, scharlaken of bordeaux zijn, maar niet bij blauwe of groene stimuli, zullen we besluiten dat 'rood zijn' in dit geval het *'switch point'*, of de meest relevante factor is. Als de duif echter enkel pikt bij scharlaken- maar niet bij vermiljoenkleurige stimuli, zullen we besluiten tot het andere geval: de perfecte oplossing. Wat hieruit blijkt, is dat 'causale relevantie' de eigenschap is die we nodig hebben als we over variabelen spreken. Als we het hebben over 'Causale krachten', komen we tot bovenstaande spraakverwarring. Dat komt omdat dit begrip betrekking heeft op het causale 'werk' dat wordt uitgevoerd in *elk individueel geval*, of voor elke individueel aangeboden stimulus. Daar kunnen we ons immers de vraag blijven stellen of het nu de scharlakenheid was, die de duif deed pikken, of de roodheid. Dit is voor Craver echter geen probleem:

> *"A multilevel mechanist need only hold that higher-level phenomena are causally relevant, not that they exercise novel causal powers. Belief in 'causal powers', not only at higher levels but also at fundamental levels, is an additional metaphysical commitment beyond the manipulationist account of causal relevance."* (Craver, 2009, p. 198)

Het experiment met de duif kunnen we ook perfect opzetten met betrekking tot mentale gewaarwording. We zouden een scala aan stimuli moeten selecteren waarvan sommige gepaard gaan met een bewuste gewaarwording en andere niet (bvb. prikjes op verschillende plaatsen op het lichaam). Daarna vragen we een proefpersoon om enkel op een knop te drukken als hij zich bewust wordt van een stimulus. Hieruit zal (een beetje redundant) blijken dat het de bewuste gewaarwording is die causaal relevant is voor het drukken op de knop. Alleen blijft ook hier de vraag of in het individuele geval de bewuste gewaarwording een extra causale impact heeft op het drukken, afgezien van die van het elektrisch stroompje dat vanaf het prikje via de hersenen naar de vinger loopt. Op een dergelijke vraag kan een interventionistische (of manipulistische) theorie geen antwoord geven. Of we die vraag willen stellen, hangt echter af van hoe ver onze metafysische nieuwsgierigheid rijkt.

5. Conclusie

In de voorgaande tekst heb ik uit de uitgebreide literatuur enkele bewustzijnsfilosofen geselecteerd die tot verschillende kampen behoren. Ze verschillen niet alleen van elkaar wat betreft de oplossingen die zij aanreiken voor kwesties als mentale veroorzaking, de mogelijkheid van reductie, en de vrije wil, maar ook wat betreft hun houding ten opzichte van het concept van causaliteit en de verschillende causaliteitstheorieën. In hun werk ben ik op zoek gegaan naar hoe zij causaliteit gebruiken in hun argumentatie, hoe zij dit eventueel zelf trachten te definiëren of, in andere gevallen, of zij verwijzen naar specifieke theorieën of denkers. Ik wilde aantonen wat de gevolgen waren van hun causaliteitsconcept op de rest van hun filosofie.

Dit was niet altijd even gemakkelijk. Soms gaven filosofen in hun werk wel blijk van een uitgebreide kennis van de verschillende mogelijke posities (Chalmers), maar hadden ze het verderop toch hoofdstukkenlang over 'causaliteit' en 'causale netwerken' als een generisch begrip (alweer Chalmers maar ook Dennett). Soms deden ze smalend over de mogelijkheid om een robuuste causaliteitstheorie op de bouwen die weerstaat aan alle tegenvoorbeelden (ook weer Dennett), en speelden dus leentjebuur bij verschillende concepten en theorieën al naargelang het hen paste. Maar ook een doorwinterde metafysicus zoals Jaegwon Kim bleek niet altijd zijn gebruikte concepten even duidelijk te definiëren, en stond daardoor open voor kritiek.

In wat volgt zet ik mijn voornaamste bevindingen nog even op een rij.

Het hoofdstuk over Descartes was een warmlopertje. Ik ben niet in staat geweest om al zijn werk na te pluizen op zoek naar een grondige analyse van causaliteit. Toch ben ik er vrij gerust in dat ik die niet gevonden zou hebben, juist gezien de vele speculatie die nu nog wordt gevoerd over Descartes idee hiervan. Ik gebruikte hem in de eerste plaats als springplank om in het probleem van mentale veroorzaking te duiken, aangezien hij dit zo bekend maakte. Een onstoffelijke geest kan immers geen contact maken met een stoffelijk lichaam, zo wist prinses Elizabeth.

De visie op causaliteit die contact hiervoor als een voorwaarde ziet, is echter maar één mogelijke positie. We treffen ze bijvoorbeeld aan in de *Conserved Quantity Theory* van

Phil Dowe. Flage en Bonnen (1997) wijzen er in hun paper echter op dat Descartes net zo goed een Hempeliaanse verklaringsopvatting aangehangen zou kunnen hebben waardoor de correlatie van lichaam en geest het logische gevolg zouden zijn van een overkoepelende natuurwet. Ook een Humeaanse causaliteitsopvatting zou dit probleem op een vergelijkbare manier oplossen.

Volgens David Chalmers supervenieert het bewustzijn op een onderliggend causaal netwerk. Vanaf het moment dat een georganiseerd systeem een bepaalde graad van complexiteit bereikt, zou hieruit een bewustzijn kunnen emergeren. Dit zou het gevolg zijn van de intrinsiek proto-fenomenologische-eigenschappen van de relata.
Een dergelijke theorie snakt misschien wel bij uitstek naar een duidelijke invulling van het begrip causaliteit. Toch moeten we deze bij Chalmers ontberen. Wanneer hij bijvoorbeeld stelt dat fenomenologische bewustzijnsinhouden niet definieerbaar zijn volgens de causale rol die zij vervullen (een vereiste voor functionele reductie), of wanneer hij bij het uiteenzetten van zijn informatietheorie gebruik maakt van het concept van causale werking, wordt dit gemis duidelijk voelbaar.
Chalmers geeft weliswaar een overzicht van manieren waarop een bepaalde visie op causaliteit een antwoord zou kunnen bieden op het probleem van de causale werkzaamheid (*causal efficacy*) van (mentale) eigenschappen. Daarin stipt hij zowel een regulariteitstheorie als de mogelijkheid van overdeterminatie aan. Misschien is het deze drang tot volledigheid die ervoor zorgt dat Chalmers geen keuze wenst te maken tussen deze verschillende mogelijkheden, en enkel hier en daar een bepaalde voorkeur uitdrukt. Als hij zijn bovengenoemde theorie over causale relata uiteenzet, blijft echter de vraag hoe we de causale relatie tussen deze (protofenomenologische) elementen nu precies moeten denken. Zijn ze noodzakelijk of enkel voldoende voor elkaar? Is fysiek contact nodig, of zijn ze tegenfeitelijk van elkaar afhankelijk? Hierop gaat Chalmers niet altijd voldoende in.

Natuurlijk moet er (zeker in de metafysica) altijd een plek zijn waar *'the buck stops'*: als we ons bezig houden met de fundamenten van het zijn, blijkt er immers altijd wel een laag te zijn onder die waar we mee bezig zijn, en één daar onder. Toch geeft Chalmers

zelf al aan wat de mogelijkheden zijn voor elke specifieke causaliteitstheorie. Hij werkt ze alleen niet altijd verder uit. Dit gaat misschien enkel ten koste van zijn theorie als je hem specifiek op dit punt onderzoekt. Het feit dat hij bijvoorbeeld wel ingaat op de metafysica van informatie en aan dit begrip een inhoud probeert te geven, heeft hij dan weer voor op Daniel Dennett. Die heeft immers naar eigen zeggen geen kaas gegeten van wat hij 'causale doctrines' noemt. Het is beter, ons bezig te houden met intuïtieve noties, zodat we allemaal nog weten waar we het over hebben. Hier ben ik het helemaal niet mee eens: soms hebben we metafysica juist nodig om eerst te *bepalen* waar we het over hebben! Opvallend bij Dennett is ook dat hij na deze laatdunkendheid toch wel zeer dicht bij David Lewis blijft in de daaropvolgende tekst. Misschien spreekt hier toch een zekere voorkeur uit? In ieder geval kon ik in de tekst telkens aangeven waar een draad die Dennett laat hangen verder afgewikkeld had kunnen worden door iets dieper in te gaan op bepaalde metafysische begrippen.

Jaegwon Kim, tenslotte, is iemand die zich aan dit laatste in ieder geval niet schuldig maakt. Toch is het opmerkelijk dat causaliteits- en verklaringstheoretici zoals James Woodward en Carl Craver in staat zijn om hun respectievelijke kritieken te uiten op het superveniëntie-argument en de notie van causale krachten. Niet dat een filosoof bij voorbaat ingedekt zou moeten zijn tegen alle kritiek, wel omdat het een beetje verbluffend is dat Kim enerzijds zoveel aandacht besteedt aan het begrip superveniëntie, en het in al haar mogelijke gedaantes en aspecten ontleedt (hij schreef er een boek over), maar verder niet dezelfde analyse maakt van het begrip 'causale krachten'. Dit zorgt natuurlijk voor verwarring. Craver geeft zelf aan niet goed te begrijpen wat Kim hier precies mee bedoelt, en Woodward schijnt te denken dat Kim hier hetzelfde mee bedoelt als 'causale relevantie'. Toch mag uit het werk van Kim duidelijk worden dat hij hier iets heel anders mee bedoelt – alleen is niet helemaal duidelijk wat!

Tenslotte valt het me zwaar om zelf een positie te kiezen in het debat. De theorieën van Woodward en Craver zijn bij uitstek toepasbaar in een concrete context en ontdaan van problematische noties. Ze kunnen omgaan met diverse tegenvoorbeelden

zoals wanneer zaken toevallig gebeuren maar toch ergens het gevolg van zijn, causaliteit op afstand of 'negatieve' causaliteit. In principe zouden zij dus mijn voorkeur wegdragen. Toch blijkt uit Woodwards bespreking van Kim dat zijn theorie toch niet helemaal in staat blijkt om het probleem te vatten dat Kim in zijn superveniëntie-argument aan de kaak wil stellen. Dit probleem zouden we misschien kunnen oplossen door voor een andere theorie te kiezen, maar dat lijkt me een beetje vals spelen. Iemand die Kims argumenten wil weerleggen, zal moeten kunnen aantonen dat het probleem zich niet voordoet gegeven diens eigen causaliteitsopvatting, of dat zijn causaliteitsopvatting op de één of andere manier niet wenselijk, of niet coherent is, en het probleem dus niet relevant. Het lijkt me dus het beste om de keuze te laten afhangen van de context.

De reden voor dit onderzoek was mijn ontzag voor de waarde van causaliteits- en verklaringstheorieën bij de analyse van metafysische theorieën. Aan het eind ervan is dat ontzag alleen nog maar toegenomen. Zelfs als we zoals Chalmers geen uiteindelijke keuze maken, staan ze ons toe om zeer nauwkeurig uit te drukken waar de pijnpunten zitten. Bovendien werkt alleen het bestaan ervan als een soort van prisma: door het inzicht dat één zo'n fundamenteel begrip, causaliteit, op zoveel verschillende manieren kan worden opgevat, worden meteen ook een heel aantal nieuwe perspectieven op de werkelijkheid 'vrijgespeeld'. Dit laatste, het aanreiken van andere gezichtspunten, is naar mijn mening nu juist de grootste waarde van de filosofie...

Bibliografie

The Blue Brain Project [Online]. EPFL. Beschikbaar: http://bluebrain.epfl.ch/ [Geraadpleegd 27/04/2013].

ALIVISATOS, A. P., CHUN, M., CHURCH, G. M., GREENSPAN, R. J., ROUKES, M. L. & YUSTE, R. 2012. The Brain Activity Map Project and the Challenge of Functional Connectomics, *Neuron,* 74**,** 970-974.

BENNETT, M. R. & HACKER, P. M. S. 2003. *Philosophical Foundations of Neuroscience,* Oxford, Blackwell Publishing.

BLATTNER, W. 2006. *Heidegger's Being and Time,* Londen, Continuum International Publishing Group.

BLOCK, N. 1978. Troubles with functionalism, *Minnesota Studies in The Philosophy of Science,* 9**,** 261–325.

BLOCK, N. 2003. Do Causal Powers Drain Away?, *Philosophy and Phenomenological Research,* 67.

CHALMERS, D. 1995. Facing Up to the Problem of Consciousness, *Journal of Consciousness Studies,* 2**,** 200-219.

CHALMERS, D. 1997. *The Conscious Mind,* Oxford, Oxford University Press.

CHALMERS, D. 2010. *The Character of Consciousness,* Oxford, Oxford University Press.

CRAVER, C. F. 2009. *Explaining the Brain,* Oxford, Oxford University Press.

CUMMINS, R. 1975. Functional Analysis, *The Journal of Philosophy,* 72**,** 741-765.

DAWKINS, R. 1976. *The Selfish Gene,* Oxford, Oxford University Press.

DENNETT, D. 1985. *Elbow Room: The Varieties of Free Will Worth Wanting,* New York, Oxford University Press.

DENNETT, D. 1993. *Consciousness Explained,* Londen, Penguin Books.

DENNETT, D. 1995. *Darwin's Dangerous Idea,* Londen, Penguin Books Ltd.

DENNETT, D. 2003. *Freedom Evolves,* Londen, Penguin Books.

DESCARTES, R. *Correspondance avec Élisabeth* [Online]. Wikisource. Beschikbaar: http://fr.wikisource.org/wiki/Correspondance_avec_%C3%89lisabeth [Geraadpleegd 25/03/2013].

DESCARTES, R. 1634. *Méditations Métaphysiques,* Paris, Éditions Flammarion.

DESCARTES, R. [1649]. *Les Passions de l'âme* [Online]. Wikisource. Beschikbaar: http://fr.wikisource.org/wiki/Les_Passions_de_l%E2%80%99%C3%A2me/%C3%A9dition_de_1649/Premi%C3%A8re_partie [Geraadpleegd 27/04/2013].

FLAGE, D. & BONNEN, C. 1997. Descartes on Causation, *The Review of Metaphysics,* 50.

GAUKROGER, S. (ed.) 2004. *René Descartes: The World and Other Writings,* Cambridge: Cambridge University Press.

HOHWY, J. & KALLESTRUP, J. (eds.) 2008. *Being Reduced,* Oxford: Oxford University Press.

HUME, D. [1748] 2007. *An Enquiry concerning Human Understanding,* New York, Oxford University Press.

KIM, J. 1973. Causes and Counterfactuals, *The Journal of Philosophy,* 70**,** 570-572.

KIM, J. 1993. *Supervenience and Mind,* New York, Cambridge University Press.

KIM, J. 1998. *Mind in a Physical World,* Cambridge, MIT Press.

KIM, J. 2003. Blocking Causal Drainage and Other Maintenance Chores with Mental Causation, *Philosophy and Phenomenological Research,* 67.

KIM, J. 2005. *Physicalism, or Something Near Enough,* Princeton, Princeton University Press.

KIRK, R. 1974. Zombies versus materialists, *Aristotelian Society,* 48 (supplement)**,** 135-152.

LEVINE, J. 1983. Materialism and qualia: the explanatory gap, *Pacific Philosophical Quarterly,* 64**,** 354-361.

LEWIS, D. 1973a. Causation, *The Journal of Philosophy,* 70**,** 556-567.

LEWIS, D. 1973b. *Counterfactuals,* Cambridge, MA, Harvard University Press.

NAGEL, T. 1974. What is it like to be a bat?, *Philosophical Review,* 4**,** 435-450.

OPPY, G. & DOWE, D. 2011. *The Turing Test* [Online]. The Stanford Encyclopedia of Philosophy. Beschikbaar: http://plato.stanford.edu/archives/spr2011/entries/turing-test [Geraadpleegd 26/04/2013].

PSILLOS, S. 2002. *Causation & Explanation,* Chesham, Acumen Publishin Limited.

ROBB, D. & HEIL, J. *Mental Causation* [Online]. The Stanford Encyclopedia of Philosophy. Beschikbaar: http://plato.stanford.edu/entries/mental-causation/#AscPro [Geraadpleegd 27/04/2013].

ROSENBERG, G. 1996. Consciousness and Causation: Clues toward a double aspect-theory. Indiana University.

ROSENBERG, G. H. 2004. *A Place for Consciousness,* New York, Oxford University Press USA.

SEARLE, J. 1980. Minds, Brains and Programs, *Behavioral and Brain Sciences***,** 417–457.

SEARLE, J., DENNETT, D. & CHALMERS, D. 1997. *The Mystery of Consciousness,* New York, New York Review of Books.

SEARLE, J. R. 2004. *Mind,* New York, Oxford University Press.

SHANNON, C. E. 1948. A mathematical theory of communication, *Bell Systems Technical Journal,* 27**,** 379-423.

VERPLAETSE, J. 2012. *Zonder Vrije Wil,* Amsterdam, Uitgeverij Nieuwezijds.

WEBER, E. & DE VREESE, L. 2012. Causation. Gent: Universiteit Gent.

WEBER, E., VAN BOUWEL, J. & DE VREESE, L. 2012. Scientific Explanation. Gent: Universiteit Gent.

WOODWARD, J. Interventionism and Causal Exclusion. Pittsburgh: University of Pittsburgh.

WOODWARD, J. 2003. *Making Things Happen,* New York, Oxford University Press.

YABLO, S. 1992. Mental Causation, *Philosophical Review,* 101**,** 245-80.

YOO, J. 2007. *Mental Causation* [Online]. The Internet Encyclopedia of Philosophy. Beschikbaar: http://www.iep.utm.edu/mental-c/ [Geraadpleegd 25/03/2013].