**cursus philosophy of Mind (op basis van ppt notities van deheer (en docent) P. Adriaens met aanpassingen, verduidelijkingen en suggesties van student Koen Dorleman)**

1. Inleiding
   1. **Het rijk der geesten**

Probleem: heeft de mens enkel een brein, met andere woorden is de mens zijn brein. Of heeft de mens zijn geest en indien ze beiden hebben wat is hun verhouding?

Er zijn drie brede categorieën die de fenomenen die we als geestelijk omschrijven indelen.

* **Attitudes (attitudes):** dit zijn houdingen, meerbepaald propositionele houdingen. Zo’n propositionele houding bestaat uit een bepaalde houding ten opzichte van een gedachte inhoud. De gedachte inhoud is de propositie de houding die men er tegenaan heeft is de attitude. Bij één propositie of gedachte inhoud kan men verscheidene houdingen aannemen. Voorbeeld: solliciteren. Men kan uitkijken naar een sollicitatiegesprek, men kan angst hebben ten opzichten van het sollicitatie gesprek of men kan hopen dat een sollicitatiegesprek goed verloopt.

* **Ervaringen: (expriences):** Men wordt ze gewaar, men schenkt er aandacht aan. Ze zijn op een hele directe manier toegankelijk voor onszelf. Voor andere is het zeer moeilijk om onze ervaringen te zijn. Ze zijn ook zeer moeilijk te beschrijven ( zie qualia van de ervaringen). Voorbeelden zijn: pijn hebben, plezier hebben, gevoelens hebben, emoties hebben.
* **Acties, handelingen ( actions):** Hier heerst een groot debat over wat nu precies mentale handelingen zijn. We zijn vrij overtuigd dat een rekensommetje maken een mentale handeling is, bij een schilderij maken is de overtuiging allicht kleiner. Voorbeelden: rekensommetje maken, schilderij maken, keuzes maken.

Wat als mentaal wordt gezien, kan via een conceptuele analyse tot stand komen. Wat een conceptuele analyse is, heeft Peter Strawson misschien wel het duidelijkst geformuleerd: “to establish the connections between the major structural features or elements of our conceptual scheme - as a coherent whole whose parts are mutually supportive and mutually dependent - to do this may well seem [...] the proper, or at least the major, task of analytical philosophy.”

Het onderscheid van bovenstaande categorieën komt tot stand aan de hand van vier criteria (zie ook tekst waarop dit gebaseerd is => toledo, pagina 6 in notities).

* **Observeerbaarheid (observability):** dit kan je het duidelijkst formuleren als observeerbaarheid door derde. Zoals reeds gezegd zijn ervaringen voor andere moeilijk observeerbaar, denk aan de tandartsstoel waar je jouw hand omhoog moet steken als je pijn voelt. Handelingen zijn dan weer wel observeerbaar. Je kan makkelijk zien of iemand een schilderij maakt. Handelingen hebben een effect in de fysieke werkelijkheid. Natuurlijk zijn ook hier weer gradaties, namelijk een rekensommetje maken in je hoofd is wederom moeilijk te observeren voor derden. Attitudes of houdingen zijn dan weer soms observeerbaar.
* **Toegankelijkheid (accesability):** Dit is het best begrijpbaar in de zin van toegankelijkheid voor je zelf. Attitudes zijn niet meteen zichtbaar en blijven soms verborgen zoals daar zijn seksuele verlangens. Ervaringen zijn dan weer zeer toegankelijk zonder dat we weten hoe of wat, ze zijn moeilijk te beschrijven (wederom zie qualia).We kunnen moeilijk pijn lijden zonder de pijn niet te ervaren. Acties zijn soms toegankelijk en soms ook weer niet. Denk maar aan dromen daar zijn we soms niet zo zeker over als dat we zeker zijn over ervaringen of attitudes.
* **Communiceerbaarheid (communicatability):** ervaringen zijn moeilijk te communiceren, want hoe beschrijf je jouw pijn. De eigenschappen van ervaringen (qualia) zijn moeilijk te duiden. Het object van onze pijn ( hoofd bij hoofdpijn) kunnen we wel beschrijven, maar hoe het voelt die pijn te hebben is moeilijk. Bij attitudes is dit heel wat eenvoudiger, men herhaalt de propositie. Een handeling is ook communiceerbaar; men kan een kind uitleggen wat je aan het doen bent indien je een gat in de muur aan het boren bent.
* **Intentionaliteit of gerichtheid (aboutness):** Waarop zijn onze handelingen ervaringen en propositionele attitudes gericht? Bij de laatste weten we dat ze gericht zijn op iets, maar op wat, de propositie?. Handelingen zijn gericht, maar hebben een andere gerichtheid dan propositionele houdingen. intentionaliteit neemt af in respectievelijke volgorde als hierboven beschreven.

**Qualia:** term die verwijst naar kwalitatieve aspecten van onze ervaringen – ‘hoe het is’ of ‘hoe het voelt’ om deze of gene ervaring te hebben. Of nog: de kwalitatieve dimensie van een mentale toestand. Ze lijken bij hun beschrijving zich te ontrekken aan een fysicalistische beschrijving.

Bovenstaande criteria toont een mogelijkheid aan om een onderscheid te maken tussen de verschillende mentale dingen. Maar wat is nu de gemeenschappelijke deler van mentale activiteiten? Deze vraag is nog niet bevredigend beantwoord in het hele debat hierover.

Een definitie van mentaal zou bepalen wie of wat een geest heeft. De definitie zou dan oordelen dat god al dan niet een geest ( In de middeleeuwen dachten ze van wel) zou hebben of dieren (volgens Descartes waren dieren automaten) al dan niet een geest hebben.

Het probleem van extended minds stelt de mogelijkheid dat de geest iets meer is dan ons lichaam en ons brein, zelfs meer dan dat wij de drager zijn van die geest. Wat maakt de geest eigenlijk tot de geest? Men spreekt hier van de Halmark of the mental.

* **Oncorrigeerbaarheid (incorrigibility):** Betekend dat niemand dat niemand kan aantonen dat mijn ervaringen onwaar zijn. Via introspectie (in de eigen geest kijken) kan ik meer zekerheid vinden over mijn ervaringen ( introspectie geld ook voor PH en handelingen) in tegenstelling tot de ervaringen van dieren of andere mensen. Op deze manier kan ik onmiddellijk meer kennis verwerven over de eigen geest. Nadeel aan deze halmark: geldt alleen voor ervaringen en niet voor PH of handelingen
* **Intentionaliteit ( intentionality, aboutness):** Geldt zeker voor attitudes, maar niet voor ervaringen. Er zijn echter filosofen die beweren dat ervaringen wel intentioneel zijn (pijn gericht op schade aan orgaan), maar dan krijgen we een brede, lege definitie van gerichtheid. Wederom een nadeel dat bij deze halmark er enkele van de drie brede categorieën niet ondervallen. (probleem van het single property view)
* **Bewustzijn (consciousness):** lijkt te gelden voor zowel ervaringen als attitudes (en zeker handelingen). Maar wat dan met onbewuste mentale processen? Er zijn mensen die zeggen iets is mentaal als het potentieel bewust is, wanneer ze dus via intensieve therapie bewust worden gemaakt.

Wat we duidelijk zien is dat bij zo’n definitie sommige dingen waarvan we dachten dat ze mentaal waren dan niet mentaal bleken, je besluit dus niet het volledige domein.

* 1. **Wat is philosophy of mind? (bekijk de 4 “onderdelen” bij elke positie en probeer hun uiteenzetting te kaderen onder de verschillende onderdelen, bekijk ook de problemen die onder deze vier onderdelen staan en hoe de verschillende posities (bijvoorbeeld materialisme) zich daartoe verhouden.)**

Metafysisch

Ben ik mijn brein of is er ook een geest? Dualisten zeggen neen, brein is lichaam en de geest is wat mijn eigenheid maakt. Materialisten zouden zeggen dat ik enkel mijn brein ben, maar ook daar bestaan verschillende posities in (bekijk hoofstukken materialise pro en contra). De idealisten zeggen dat er enkel een geest bestaat.

Er zijn echter enkele problemen: de qualia of de kwalitatieve eigenschappen van de ervaringen, zijn deze enkel breinprocessen? Ja dan zouden we ze fysicalistisch moeten kunnen bespreken, maar dat kunnen we niet (ook niet met een een psychologische beschrijving). Een ander probleem vormt zich met mentale causaliteit (psychopathologie zal hier een belangrijke rol spelen: ‘Psychopathology is a domain of scientific inquiry into human behavior in which there is a presumption that consciousness makes a behavioral difference’ (Graham)). Het belang va common sense is ook een probleem. (nadat alle posities behandeld zijn probeer de vorige problemen duidelijker te situeren bij elke positie)

Epistemologisch

Er zijn overlappingen met ‘de leer van de kennis’. Het probleem situeert zich rond de al dan niet zekerheid over het bestaan van andere geesten. Met welke zekerheid en hoe kan ik weten of andere mensen (of objecten: computers) geen geesten hebben?

Methodologisch

Op welke manier moet ik onderzoek doen in dit debat. Via leunstoelfilosofie, namelijk uitgaan van bepaald intuïties zonder info en dat verder uitwerken via gedachten experimenten. Andere maken dan weer gebruik van experimentele filosofie omdat ze twijfelen over de werkbaarheid van intuïties. Ze baseren zich dan deels op empirische wetenschappen, maar welke moeten we dan gebruiken? Psychiatrie, ontwikkelingspsychologie, neurowetenschappen, computerwetenschappen,…

Alleszins hebben gedachten experimenten belang. Voorbeeld het malin genie van Descartes. We leren er iets over de werkelijkheid enkel beroep doend op gedachten.

Voorbeeld van gedachten experiment van Lucretius: ga naar de rand van het universum, neem een speer mee, en werp die tegen de rand. Ofwel gaat de speer door de rand, en eindigt de ruimte dus niet; ofwel komt de speer terug, en dan moet er iets zijn buiten de rand van het universum, een of andere kosmische muur, die ook deel uitmaakt van het universum. Er is bijgevolg geen einde aan het universum.

Andere voorbeelden: de kat van Schrödinger, Descartes’, Searle’s Chinese room, Jackson’s Mary the Colour Scientist, enzovoort

Ethisch

Dit gaat rond de vraag van vrije wil. Indien we zeggen dat we enkel ons brein zijn (Dick Swaab) moet die verplicht positie innemen of er dan ook een vrije wil is.

1. Dualisme pro ( tekst substance dualism van T.J. Mawson)
   1. **Bij wijze van inleiding**

In de common sense zijn wij dualisten. Wij zien lichaam en zielen als verschillende dingen. Wij kijken naar de werkelijkheid op twee verschillende manieren; in termen van materiële objecten (ook ons eigen lichaam) en in termen van niet materiële zielen. ( Paul Bloom)

Onze intuïtie neigt naar dualisme en volgt uit een fenomenologische ervaring. Het gevoel dat we allen vleesmachines zijn (Marvin Misky) is tegen intuïtief. We spreken van Mijn lichaam, mijn brein omdat we ervan uitgaan dat we ons lichaam bezitten en bezetten, maar wat we zijn is meer dan dat, we voelen ons nauw verbonden met ons lichaam zoals we dat doen bij een auto, we zeggen mijn auto, maar we zijn meer dan die auto.

Uit die dualistische common sense kan men allerlei zaken verklaren over persoonsidentiteit, namelijk het feit dat we ons allerlei situaties kunnen inbeelden waar het lichaam enorm veranderd, maar wij het zelfde blijven. Voorbeeld hiervan is het boek van Frans Kafka waarin hij beschrijft dat een man des avonds gaat slapen en het des ochtends wakker wordt als een kakkerlak, maar nog steeds die man is. We kunnen ook begrijpen dat mensen beweren dat hun ziel naar een parallelle wereld gaat, het een ander lichaam bezet en dergelijke terwijl het lichaam afsterft. Dit is enkel mogelijk wanneer we ziel en lichaam loskoppelen.

Buiten persoonsidentiteit verklaart onze dualistische intuïtie ook morele en politieke kwesties. De vraag die hier centraal staat is of het ding in kwestie een ziel heeft of niet, zo ja (precious individual) is het de moeite waard om het te beschermen indien niet is het gewoon een fysiek object ( mere thing). Mooi voorbeeld hiervan is klonen. Het vaticaan bracht als argument tegen dat personen die gekloond worden niet dezelfde ziel zouden hebben. Ander voorbeeld is wanneer wetenschappers vragen wanneer het leven precies begint. Ze verwachten hier een objectief antwoord op, maar of dat er kan komen is echter de vraag, het gaat hier niet om biologie, maar om het moment dat een hoopje cellen leven wordt.

* 1. **Wat is dualisme?**

Dualisme in een eerste definitie is de stelling dat het universum opgebouwd is uit twee totaal verschillende ( irruducibly different) soorten ‘spul’.

Geestesspul en fysisch spul (materie) zijn zo verschillend dat ze niet tot elkaar kunnen gereduceerd worden. Men kan dus de een niet herleiden tot de andere en de andere niet herleiden tot de een.

Fysisch spul verstaan we het brein, hersenen, lichaam of objecten onder. In een vorige eeuw werd fysisch spul vooral materie genoemd door wetenschappers, maar die wetenschappers zijn nu bezig over materie dat omzetbaar is in energie, quarks hyperdimensionale snaren om de meer gewone objecten die wij in het dagelijkse leven ervaren te beschrijven. Fysisch spul wordt verstaan als drie-dimensionaal. Het is gebonden aan tijd (eindig) en aan ruimte (uitgebreid). Probleem met fysisch spul tegenwoordig is dat ze voornamelijk uit lege ruimte blijken te bestaan. Fysisch spul is objectief en publiek toegankelijk. Vragen die we ons stellen wanneer we fysisch spul bespreken zijn, hoe groot is het, waar bevindt het zich, welke vorm heeft het,…

Geestesspul is een tweede soort spul, namelijk dat spul dat nooit volledig in een wetenschap van objecten kunnen beschreven worden. Dit omdat het eigenschappen heeft die niet behoren tot die wetenschap en dus onmogelijk volgens die wetenschap kunnen gereduceerd worden. Geestesspul is niet gebonden aan tijd noch aan ruimte, is privé en subjectief (= moeilijk toegankelijk voor derden). De vragen die gesteld worden wanneer we fysisch spul bespreken , lijken raar als we ze bij geestesspul stellen. Probeer de vorm van emoties eens te beschrijven.

* 1. **Varianten van dualisme**

Wat alle varianten van dualisme gemeen hebben is zijn de centrale vragen (opgedeeld in 2) die gesteld worden.

* Hoe verhouden geestesspul en fysisch spul zich tot elkaar, is er sprake van uitwisseling of wisselwerking en is dat dan eenrichtingsverkeer of tweerichtingsverkeer? Indien er wisselwerking is spreken we van interactionistisch dualisme.
* Hoe verhouden fysisch spul en geestesspul zich tegenover ons persoon-zijn. Substantie-dualisten zullen zeggen dat ons persoon-zijn louter bepaalt worden door onze geest. Eigenschapsdualisten stellen dat onze persoon twee soorten eigenschappen heeft die overeenkomen met de twee soorten spul in de werkelijkheid, namelijk fysieke en mentale eigenschappen.
  + 1. **Substantie dualisme**

Definitie van een substantie volgens Descartes: een ding dat kan bestaan dat niets anders nodig heeft dan zichzelf om te bestaan. Om het anders te formuleren substantie versus eigenschap, want eigenschappen hebben wel een ander ding nodig om te bestaan, namelijk een substantie. Verschil tussen essentiële en accidentele eigenschappen.

Let op: dualisten claimen vaak, maar niet altijd, dat wij als persoon verbonden zijn met een lichaam, maar we hebben dat lichaam niet nodig om onszelf te identificeren. ‘I have no hair, for instance, no face, no hands, fingers, or limbs. I have no heart, no brain – no organs at all. I am not capable of being hit by a bus, or being burned in a fire. These are all physical attributes, and if substance dualism is true, I have no physical attributes’ (J 37). Ik ben geen menselijk organisme, en mensen die mij omringen zijn geen personen.

1. Als we zonder lichaam kunnen bestaan, dan kunnen we geen lichaam zijn
2. We kunnen zonder lichaam bestaan 🡪 twee argumenten: a) conceivability en b) cartesiaanse redenering (denken/ervaren is mijn enige essentiële eigenschap)
3. DUS: we kunnen geen lichaam zijn

Descartes Zijn drijfveer was voornamelijk de onenigheid die bestond tussen alle filosofen. Hij ging op zoek naar een vast archimedisch vertrekpunt. Hij Heeft een radicale conclusie in verband met het probleem geest, persoon ⬄ hersenen of lichaam, namelijk  *Ik* heb geen lichaam. De vraag is echter of we zonder lichaam kunnen bestaan. Het radicale bij Descartes bestaat erin dat het persoon-zijn volledig afhankelijk is van ons geestesspul We zijn verbonden met een lichaam, maar onze identiteit licht in ons geestesspul. Descartes argument bestaat in twee delen/ bewegingen.

1. Op zoek gaan naar een vertrekpunt
2. Kennis opbouwen vanuit dit standpunt

Hoe kunnen we overtuigingen produceren die onverwerpelijk zijn? Descartes heeft drie soorten twijfels om op zoek te gaan naar een vertrek punt. Zijn eerste twijfel is de natuurlijke twijfel hierbij twijfelt hij aan de zintuigen (vb. optische illusies). Tweede twijfel is de hyperbolische twijfel ook van deze twijfel zijn de zintuigen voorwerp, alleen gaat het hier om de gedachte dat zintuigen ons permanent bedriegen (hallucinaties, levendige droom). Zijn derde twijfel is de metafysische twijfel. Voorwerp van deze twijfel zijn de wiskunde en andere intuïties. Voorbeeld malin genie. Je kan je makkelijk inbeelden dat je eigenlijk een pot hersenen bent op een verre planeet en een alliën zorgt ervoor dat ik voorstellingen heb.

Descartes conclusie.

De eerste zekerheid is de volgende: ik denk dus ik ben. Zolang je twijfelt, denk je en zolang je denkt besta je. Denken is gelijk aan ervaren. Ervaren is normaal gekoppeld aan lichaam, denken is beperkt tot twijfelen, fantaseren,… Als ik stop met denken houd ik op met bestaan. Met deze redeneringen zijn de premisse aangetoond. Voor Descartes is het bewustzijn dus het kenmerk voor het mentale; ‘Res cogitans is a thing which doubts, understands, affirms, denies, wills, refuses, which also imagines and feels’. Descartes denken is een zeer breed begrip en houdt zowel ervaringen en propositionele attitudes als mentale handelingen in.

John Eccles is een Australisch neurofysioloog die nobelprijs voor fysiologie en geneeskunde won met werk over verbindingen tussen hersencellen, toch is hij er van overtuigd dat er meer is dan alleen hersenen. Hij is een dualist en in vervolg met Karl Popper een interactionistisch trialist. Ze gaan beiden uit van 3 werelden:

* Wereld 1: fysieke objecten en statussen (materie en energie, levende wezens en hun acties en artefacten)
* Wereld 2: states of counsciousness: (outer sense: expiriences; inner sense (PA); en pure ego/self/soul)
* Wereld 3: kennis in objectieve zin: resulaten van wereld 2 in wereld 1 met name theorieën, literatuur, resultaten van intellectuele ontspanningen.

Alles wat bestaat in de wereld of in onze ervaringen kan geclassificeerd worden in een van de werelden.

* + 1. **Eigenschapdualisme**

Het verschil tussen eigenschapdualisme (E) en substantiedualisme (S) is dat bij S enkel personen mentale eigenschappen hebben en enkel lichamen materiële eigenschappen. Bij E hebben personen zowel lichamelijke als mentale eigenschappen. Daarom dat men aan eigenschapdualisme vaak de term dual attribute theories (DAT) geeft.

Eigenschapdualisten beweren net als substantiedualisten (en tegen materialisten) dat sommige mentale fenomenen niet beschreven kunnen worden met behulp van de fysica, maar net als materialisten beweren ze dat als er mentale eigenschappen bestaan, ze alleen begrepen kunnen worden door te verwijzen naar lichamelijke processen.

Er zijn verschillende soorten DAT, de verschillen zijn afhankelijk van de visie op de relatie tussen fysische en mentale fenomenen. Waar er geen onenigheid tussen beide theorieën is dat zowel epifenomenalisme als emergentisme beweren dat fysieke fenomenen mentale fenomenen veroorzaken of beschrijven, zo ook mentale fenomenen als geloof, pijn en verlangen.

Epifenomenalisme (zie ook eliminatevistisch fysicalisme)

Ze gebruiken verschillende metaforen om de geest te beschrijven: Idle wheels (radertjes die worden aangedreven, maar zelf niets aandrijven), Fluittoon van een stoomlocomotief, aristocraten,…

Ze stellen dat emergente mentale fenomenen causaal impotent zijn met andere woorden dat mentale fenomenen geen invloed hebben op de fysische werkelijkheid. Mentale fenomenen worden gezien als restverschijnselen of randfenomenen van fysische processen. Ze nemen niet deel aan de causale keten van input (informatie) => output (gedrag)

Epifenomenalisten zijn interactionistisch, maar het verkeer werkt maar in één richting. Hun stelling lijkt in te gaan tegen de intuïtie van de common sense, namelijk wanneer wij pijn voelen (mentaal fenomeen, informatie) zelf onze hand wegtrekken (gedrag).

Emergentisme (zie verder)

* + 1. **Occasionalisme**

Een van de centrale argumenten in het dualisme is het argument van de common sense dat we zelf greep hebben op ons lichaam. Occasionalisme (en ook parallellisme) omzeilt dit probleem ze stelt met name dat er geen interactionistisch model is.

Nicolas de Malebranche stelt dat er geen causaal verband is tussen lichaam en geest, meer nog hij stelt dat het lichaam geen impact kan hebben op de oneindige, perfecte geest.

Het is god die ervoor zorgt dat veranderingen in mijn brein gevolgd worden door veranderingen in de geest.

(probleem: als god de verantwoordelijkheid neemt voor elke wisselwerking, is hij ook verantwoordelijk voor elke kwade mens.)

Occasionalisme en parallellisme zijn varianten van non-interactionistisch dualisme. Substantiedualisme en emergentisme zijn vormen van interactionistisch dualisme. (verklaar de verschillende stromingen aan de hand van de theorie = goede oefening).

* + 1. **Parallellisme**

Bekende vertegenwoordiger van deze stroming is Gottfriend Leibniz. Hij ziet mentale fenomenen (geest) en lichamelijke fenomenen ( fysische wereld, brein,…) als twee parallelle werelden waartussen de illusie bestaat dat er causaliteit is.

* 1. **Argumenten voor dualisme**

Religie

Sinds mensenheugenis geloven mensen in leven na de dood, dit wil zeggen dat een persoon in de zin van een ziel, blijft voort leven, de ziel is dus onsterfelijk. Het gaat hier niet alleen om de macht van de massa (7 billion people can’t be wrong), maar ook om het verleden. Intuïtief is dit een heel aantrekkelijk argument omdat het makkelijk in te beelden is.

Kant trekt een ander gevolg aan de onsterfelijkheid van de ziel, met name dat we enkel goed handelen als we geloven in die onsterfelijke ziel omdat we in een ander leven dan beloond worden voor ons goed gedrag. Indien we dat niet zouden doen zouden we allemaal free-riders zijn.

Het religie argument is eerder een motivatie dan een sluitend argument, je kan met name iets geloven zonder goede redenen/argumenten te hebben om er in te geloven. Daarbij is reeds bewezen dat de massa het wel eens mis kan hebben, denk maar aan geocentrisme. Religie is bovenal iets intuïtief en vloeit voort uit de common sense, namelijk religie is een naturalistisch fenomeen dat voortkomt uit de geesten van de mensen op een bepaalde plaats in een bepaalde tijd, intuïties van die mensen kunnen dus weerspiegeld worden door mensen. Veel christenen zijn minder radicaal dan Descartes, Het is uiteindelijk de bedoeling dat die onsterfelijke ziel opnieuw belichaamd wordt, denk maar aan de verrijzenis van Christus.

Bijna-doodervaring

Een bijna doodervaring kan beschreven worden aan de hand van volgende kenmerken: men ervaart dat men het lichaam verlaat en leeft buiten het lichaam, men kan getuigen van buiten het lichaam, men kan bewegen rond de onmiddellijke fysieke omgeving van het lichaam, men ervaart een compleet andere wereld omgeving en een terugkeer naar het lichaam vanuit die andere wereld omgeving.

Indien substantiedualisme de waarheid spreekt, moet het mogelijk zijn verder te leven als mijn lichaam ligt te rotten onder de grond, vraag is welke argumenten er kunnen gegeven worden. Het BDE-argument is een argument dat ook gegeven wordt door mensen die niet religieus zijn. A.J. Ayer een filosoof die zo een ervaring heeft gehad, is na die ervaring ervan overtuigd dat het leven niet stopt na zijn dood al zou hij dat wel willen. Zijn bijna dood ervaring heeft hem enige angst bezorgd voor de dood.

Men kan zich de vraag stellen bij de getuigenissen of die wel oprecht zijn en niet eerder hallucinaties doormiddel van verdoving zoals men bij het gebruik van drugs eventueel roze olifantjes kan zien.

Sterkere tegenargumenten zijn echter dat BDE-ervaringen beschreven worden als lichamelijke ervaringen zijn. Men wandelt door een tunnel ervaart licht,… Dit lijkt te suggereren dat een BDE-ervaarder toch belichaamd is. Men kan hier een tegenargument oproepen dat dit in het taalgebruik ligt en dat dit de enige manier is om te getuigen van zulke ervaringen.

BDE-ervaringen getuigen soms van ontmoetingen en communicatie met andere personen in hun ervaring, maar hoe is het mogelijk te communiceren zonder stemapparaat, zonder gehoorapparaat,.. Hoe is het mogelijk te spreken van interactie hier?

Common sense

Substantiedualisme doet beter dan fysicalisme omdat het enkele intuïties uit de commen sense in rekening brengt (Mawson p. 85). We stonden in het begin al even stil bij Paul Bloom en zijn argument rond commen sense. We zijn allemaal geboren dualisten en onze intuïties zijn doorspekt van dualisme.

Common sense is een verzamelnaam voor een waaier aan intuïties die wij als mens hebben. Een eerste intuïtie is dat we ondeelbaar zijn (zie ook volgende argument), we kunnen namelijk niet zeggen dat we de ene moment doodgemarteld worden en op hetzelfde moment genieten van een plezant drankje (mawson p. 87). Mawson redenering gaat als volgt: a) onze intuïtie …, b) een goede theorie over mind/body moet onze intuïties een plaats kunnen geven, c) fysicalisme heeft geen plaats voor onze intuïtie dat… (WAAROM NIET? Is in deze fase misschien nog niet helemaal begrijpelijk, maar punt is dat fysicalisme persoon identificeert met ofwel psychologische eigenschap ofwel deel van het lichaam ofwel combinatie van beide – MAAR zowel eigenschappen als delen van het lichaam zijn meervoudig realiseerbaar, en dat geldt niet voor personen); d) DUS fysicalisme is geen goede theorie.

Een tweede intuïtie: wij zijn vrije personen die keuzes kunnen maken en bijgevolg moreel verantwoordelijk zijn. Indien fysicalisme waar is, is vrije keuze onmogelijk, want een fysisch universum is gesloten en bepaald door voldoende en noodzakelijke oorzaken. Bijgevolg houdt dualisme meer rekening met deze intuïtie met name de geest maat keuzes om dit te doen en niet dat. (Mawon p. 88)

Derde intuïtie: er zijn zaken die ontsnappen aan een fysische beschrijven met name qualia van ervaringen (zie tekst What mary didn’t know)

De vraag is echter wat het belang is van common sense in de filosofie en common sense in het algemeen. Veranderen intuïties niet doorheen de tijd en zijn intuïties bij alle bevolkingsgroepen of leeftijdsgroepen het zelfde?

Ondeelbaarheid

Fysisch spul kan men opdelen en dat blijft fysisch spul. Geestesspul kan men niet opdelen die hetzelfde geestesspul blijven, het is alles of niets. Wanneer het ik denkt over zijn eigen bestaan, bestaat het of bestaat het niet.

Probleem bestaat echter bij Multiple personality dissorder (MPD, of dissociatieve identiteitsstoornis) die tonen aan dat geesten wel kunnen opgedeeld worden in verschillende persoonlijkheden. Kritische analyse van MPD toont echter dat het vaak om een opgezette constructie waarin deze patiënten geloven waarbij therapeuten die wouden bewijzen dat MPD echt bestaat ervoor zorgde dat ze verschillende persoonlijkheden ontwierp. Een andere opwerping is echter dat mensen beweren zichzelf niet te zijn of dat ze veranderd zijn.

Denkbaarheid (conceivability)

het argument van denkbaarheid is een van de sterkste argumenten voor dualisme. Het wordt ook een modaal argument genoemd omdat het een bepaalde invulling geeft van mogelijkheid

Wanneer we Descartes redenering bekijken is dat een eenvoudige ponendo ponens (als p dan q, p dus q)

1. Als we kunnen bestaan zonder lichaam (p), dan kunnen we niet ons lichaam zijn (q)
2. We kunnen bestaan zonder lichaam (p)
3. DUS: We kunnen niet ons lichaam zijn (q)

Dualisten zeggen dat het mogelijk is te bestaan zonder lichaam. Ze zeggen dit omdat we ons dat kunnen voorstellen/inbeelden (concievability) en als we het ons kunnen voorstellen dan is het ook mogelijk (possebility).

Waarom zouden we aannemen dat we ons kunnen voorstellen dat we kunnen bestaan zonder lichaam. Twee argumenten (Descartes malin génie en BDE-argument) zijn al besproken we concentreren ons verder op het denk-experiment van William Hart.

William Harts denkexperiment gaat als volgt: Op een bepaald moment sta je voor de spiegel, gooit wat water in je gezicht, kijkt opnieuw in de spiegel en in plaats van ogen te zien die terugstaren naar jezelf, zie je enkel oogkassen zonder iets in. Na wat tasten om zeker te zijn snij je jouw schedel open en om te zien of er een brein is en je merkt, er is geen brein. J ziet dus zonder brein en ogen. Uiteindelijk is er geen lichaam in de spiegel, maar toch kijk je. Het komt erop neer dat wanneer je kan inbeelden dat je zonder lichaam leeft, dat tenminste metafysisch mogelijk is en daardoor geest en lichaam niet geheel hetzelfde zijn.

Er zijn echter twee tegenargumenten die we kunnen opwerpen. In de eerste plaats gaat het niet over elke vorm van conceivability. Wanneer iemand onder invloed van drugs een delirium ondergaat en zich allerlei waanvoorstellingen kan inbeelden is dat niet iets wat ook mogelijk is. Dualisten hebben het echter over een speciale vorm van conceivability, met name een nuchtere vormen van voorstellen zoals Descartes idee claires et distinctes. Alleen dingen die helder en welonderscheiden zijn kunnen ook mogelijk zijn. Ik kan me helder en welonderscheiden voorstellen dat mijn lichaam vasthangt aan mijn geest, maar ook dat mijn geest als iets denkend welonderscheiden en helder iets afzonderlijk is, hierdoor besluit Descartes tot het onderscheid tussen lichaam en geest.

Een tweede argument dat we kunnen opwerpen is wat er precies met mogelijk bedoeld wordt, want niet alles wat men zich kan voorstellen is daarom mogelijk in de echte wereld. Er zijn drie types van mogelijkheid, met name fysische, technische en metafysische. Er zijn dingen die vroeger technologisch onmogelijk, maar niet fysisch onmogelijk waren. Denk aan radio-communicatie. Sommige dingen zijn fysisch onmogelijk (voor mensen althans), zoals 100 meter lopen in 1 seconde. ‘Fysisch onmogelijk’ betekent dat zulke toestanden incompatibel zijn met de wetten van de fysica. Anders gezegd: in ons universum is het onmogelijk dat een menselijk wezen 100 meter loopt in 1 seconde. Maar dat betekent niet dat dat metafysisch onmogelijk is. Misschien zijn er universa waar andere wetten heersen, en waar het wel mogelijk zou zijn, zelfs voor een gewone mens, om 100 meter in 1 seconde te lopen. Maar zelfs in zulke universa is het onmogelijk dat 5+5=17, of dat een vierkant rond is. Zulke dingen zijn met andere woorden metafysisch onmogelijk. Er is geen enkele wereld, geen enkel universum waar 5+5=17. We kunnen namelijk geen zinnige beschrijving geven van hoe zoiets mogelijk zou zijn.

De vraag die zich dus stelt is of het metafysisch echt mogelijk is om ons voor te stellen dat een geest zonder lichaam zou kunnen bestaan en of we daar een coherente beschrijving van kunnen geven. De gedachten experimenten laten uitwijzen dat zoiets wel degelijk mogelijk is. Vele dualisten zijn zich echter bewust dat in onze wereld het brein aan de geest gekoppeld is, maar is hun hele argument van denkbaarheid dan geen speilerei. Bij de dualisten gaat het er echter om de radicale claim van de materialisten dat mentale toestanden in alle mogelijke werelden gelijk zijn aan hersentoestanden. Dualisten vinden deze claim veel te sterk.

1. Dualisme contra
   1. **Argumenten tegen dualisme**
      1. **Interactie**

Het argument van denkbaarheid ondersteund de intuïtie dat de werkelijkheid uit twee soorten spul bestaat. Dualisten hebben echter een ander probleem, namelijk in rekening brengen van de intuïtie dat er tussen het geestesspul en het fysisch spul interactie bestaat zoals Ryle het definieert: ‘What the mind wills, the legs, arms and the tongue execute; what affects the ear and the eye has something to do with what the mind perceives; grimaces and smiles betray the mind’s moods and bodily castigations lead, it is hoped, to moral improvement’.

Een andere manier om dit probleem te illustreren is aan de hand van het cartesiaans theater dit gaat ook onder de naam van homunculus-probleem (Daniel Bennet). De homunculus is een klein mens-achtig mannetje dat in het verleden een rol speelde in allerlei alchemische en biologische theorieën. In deze visuele voorstelling van het Cartesiaans theater zijn er verschillende homunculi (verkleed als wetenschappers), en die zijn er om de gegevens die aangeleverd worden door het lichaam te interpreteren, en om te zetten in wilsbeslissingen die op hun beurt doorgespeeld worden naar het lichaam. De levensgeesten van Descartes zijn ook een soort homunculi.

Wat is het probleem met homunculi? Ten eerste zijn het fictieve wezens. En ten tweede verplaatsen ze het probleem van interactie. Want een homunculus lijkt ook uit een geest en lichaam te bestaan. En dus stelt het probleem van interactie zich opnieuw.

De intuïtie van interactie lijkt moeilijk te rijmen met substantiedualisme, want hoe kan een niet-fysische substantie een fysische substantie beïnvloeden, om het met een mooie metafoor te verwoorden: hoe kan een trein op een spoor een andere trein aanrijden op een totaal ander spoor.

Indien dualisme waar is en er interactie bestaat tussen de twee, staat het in contradictie met de fysische wet dat een universum gesloten is. Dan zou wanneer het lichaam de geest beïnvloedt energie uit dat universum verdwijnen en wanneer de geest het lichaam beïnvloedt energie in het universum gepompt worden. Dit is het probleem van causale geslotenheid. Een ander probleem dat waaruit dan volgt dat als dualisme bestaat interactie onmogelijk wordt, maar toch ervaren wij raakpunten tussen de twee werelden.

Sommige dualisten brengen hier tegenin dat er levensgeesten bestaan of dat er verbondenheid is via een pijnappelklier. Descartes bracht zelf nog een andere mogelijkheid in, namelijk dat we de stelling van interactie als een soort axioma moeten aannemen de materialisten doen dit ook wanneer zij zeggen dat interactie mogelijk is tussen twee fysische entiteiten.

Een ander iets wat er tegenin kan gebracht worden is het argument van kwantumsprongen dat toeval mogelijk maakt in een deterministisch universum. De openingen die veroorzaakt worden door de sprongen worden dan opgevuld door zielen. Dit gaat nog steeds in tegen de causale geslotenheid, maar er worden geen natuurwetten meer geschonden.

Een andere bemerking die we zouden kunnen maken is de vraag of wetenschap dan zoveel beter is in de vraag of er geesten zijn of niet. Mawson formuleert het als volgt: deterministische fysica is ‘just a paradigm of interpretation; there would be no necessity that the substance dualist follow it’ (p.84). Indien een substantiedualist dit argument aanneemt en bijgevolg het wetenschappelijk paradigma niet volgt, moet hij ook de gevolgen aanvaarden, namelijk dat hij door andere filosofen en wetenschappers niet serieus wordt genomen.

Een andere bemerking is de volgende: waarom zou interactie onmogelijk zijn in een substantiedualisme, kolonialen en autochtone bevolking hadden niets gemeenschappen en toch bestond er een primitieve vorm van interactie. Echter dan zouden we de interactie tussen geestesspul en fysisch spul die uitermate complex is aannemen tot iets elementair, maar de interactie is juist complex.

* + 1. **Eenvoud**

Filosofen zijn het erover eens dat eenvoudige theorieën verkozen worden boven complexe ingewikkelde theorieën. Eenvoud is een van de theoretische deugden naast verklaringskracht en voorspellende kracht.

Onder eenvoud kan men twee dingen verstaan, namelijk 1) syntactische eenvoud of elegantie. Een zo laag aantal mogelijk basisprincipes die eenvoudig te beschrijven zijn. Substantiedualisme heeft geen probleem om de fysische substantie te beschrijven dat doet de fysica voor hun, maar de geestessubstantie beschrijven is een andere zaak. Ryle bemerkte al dat Deschartes de geest beschrijft aan de hand van woordenschat uit de fysica en mechanica, maar dan gewoon met negatieve verwoording van die termen. Descartes doet dit juist om het contrast duidelijk te maken dat geestesspul en fysisch spul totaal verschillend zijn. 2) Ontologische eenvoud is de andere eenvoud. Hieronder wordt verstaan dat men, kort en krachtig, het simpel moet houden. Men moet zich, maar houden aan een klein aantal entiteiten die gepostuleerd worden. Dit principe hangt zeer nauw samen met het immers bekende gezegde van William van Ockham, gekend onder het gezegde van het scheermes van ockham: Entia non sunt multiplicanda praeter necessitatem. Vertaald: je mag entiteiten nooit nodeloos vermenigvuldigen. Mawson lijkt hiermee akkoord te gaan wanneer ze zegt dat substantie dualisme ontologisch extravagant is.

Mawson lijkt dit argument tegen het dualisme het enige argument om waarde aan te hechten. Hij stelt zich echter de vraag of de deugden van het dualisme, met name rekening houden met de common sense, niet sterker weegt. Uiteindelijk vindt ze die deugd sterker. Goede lezers van deze samenvatting/cursus hebben juist gelezen dat dualisten geen rekening houden met de intuïtie dat de twee spullen communiceren met elkaar. De vraag die we ons ook moeten stellen is common sense belangrijker dan de andere theoretische deugden?

* + 1. **Andere geesten**

De toegang tot onze eigen geest is zeer geprivilegieerd, keerzijde van deze medaille is dat we geen toegang hebben tot de geesten van andere personen. Geen rechtstreekse (perceptie), want geest is immaterieel, maar ook geen onrechtstreekse toegang via het lichaam, want het zou evengoed kunnen dat een lichaam verbonden is met verschillende geesten (MPD). Als dualisme waar is kunnen we nooit zeker weten dat andere geesten bestaan, want we kunnen dat niet afleiden uit lichamelijke fenomenen (zoek ook in tekst Ryle).

Wederom situeert het probleem zich rond een intuïtie die we hebben. Namelijk dat we weten dat ze bestaan en dat we er een zekere toegang tot hebben. Goede theorieën dienen rekening te houden met intuïties, SD doet dat niet, dus SD is geen goede theorie.Tegen de probleemstelling van andere geesten kan men het argument van analogie aanbrengen: wij weten dat er andere geesten bestaan, en in welke toestanden die geesten zich bevinden, omdat a) ik een geest en een lichaam heb, en b) geesten en lichamen van andere individuen gelijkaardig (of analoog) zijn aan mijn geest en mijn lichaam (deze hypothese is zuiniger dan de hypothese dat andere individuen radicaal anders georganiseerd zijn). Probleem: is mijn veralgemening wel gerechtvaardigd?

Mawson stelt echter dat het andere geesten argument in het voordeel van dualisten spreekt, zijn redenering is de volgende: hoe weten we dat iemand anders een geest heeft, of een mentale toestand? Fysicalist heeft een brainoscope, en heeft zo rechtstreeks toegang tot geest/brein van iemand anders. Het Mary-argument (een gedachte-experiment eigenlijk) toont echter aan dat zo’n brainoscope niet alle mentale toestanden weergeeft – qualia bijvoorbeeld zou je er niet mee kunnen zien. Het dualisme is compatibel met dit gedachte-experiment. Het is compatibel met de gedachte dat inderdaad niet alle mentale toestanden toegankelijk zijn voor derden – dat we bijgevolg niet met zekerheid weten of iemand een bepaalde mentale toestand heeft, en bij uitbreiding of iemand een geest heeft. Eigenlijk geeft Mawson toe aan een zeker pessimisme: we zullen nooit met zekerheid weten of iemand anders ook een geest heeft. Er is geen echte oplossing voor dat probleem.

* + 1. **Geestige dieren**

De vraag of dieren net als mensen geest hebben is op zich al een debat waard binnen philosophy of mind. Panpsychisten hebben geen moeite om dieren een geest toe te schrijven, ze doen dit ook met thermometers, vulkanen,… Descartes en de substantiedualisten begeven zich aan het hele andere uiteinden. Voor hen zijn dieren louter automaten die geen ziel kunnen hebben en bijgevolg geen pijn of andere mentale fenomenen kunnen ervaren.

De vraag of mensen niet gewoon een dier onder de dieren zijn en bijgevolg ook automaten (indien Descartes en SD gelijk zouden hebben) steunt op een bepaalde verwarring, met name anachronisme. In de vroeg-moderne tijd was het niet bepaald populair om te stellen dat mensen ‘gewoon’ dieren zijn. Het is wachten op Darwin om die gedachte te lanceren en als vanzelfsprekend te maken ook al had De lamettrie in de 18de eeuw een poging ondernomen.

Hume stelde echter al dat dieren ook een geest konden hebben. Hij maakte gebruik van het argument van analogie; als wij mensen als intelligent beschouwen hebben we geen problemen om hen een geest toe te schrijven, wanneer dieren intelligent gedrag vertonen zouden we dit toch ook moeten doen.

Een wetenschappelijk argument (zoek de verdere verklaring elders = niet duidelijk in ppt)

Volgens Donald Davidson ligt de zaak echter iets complexer. Overtuigingen stelt hij, zijn pas mentaal indien ze in een netwerk van overtuigingen liggen. Dieren hebben zulke netwerken niet waardoor we kunnen stellen dat dieren geen geest hebben. Indien ze toch een geest hebben is die fundamenteel verschillend van de onze.

Als we toch geloven dat dieren een geest hebben en overtuigingen hebben zoals geloven, moeten we kunnen aantonen in wat ze geloven, maar hoe doen we dat? We kunnen niet spreken met dieren. (er zijn filosofen die stellen enkel omdat dieren niet kunnen spreken, ze niet kunnen denken, want om te denken moet je een taal hebben waarin je kunt denken, bijgevolg hebben dieren geen geest).

* + 1. **Varia**

Antionia Damasio

Antonio Damasio schrijft in zijn boek *Descartes vergissing* dat het lichaam en geest wel met elkaar verstrengeld zijn, ze kunnen niet strikt gescheiden worden. Alle gedachten zijn volgens Damasio gegrond in lichaamvertegenwoordigde neurale structuren en uit die structuren krijgen gedachten de kracht om ons in beweging te zetten. Descartes beweerde ten onrechte dat alleen het geestelijke zo een dynamiek had om ons in actie te zetten en op onverklaarbare wijze het lichaam mee in actie kon zetten.

Damasio vergelijkt Descartes geest met zijn patiënten. Zij hebben neurologische schade waardoor zij uiterst rationeel zijn, maar geen emoties kunnen ervaren. Descartes zou deze patiënten als filosofische ideaal zien voor de geest.

Wanneer Damasio stelt dat de geest niet zonder lichaam kan bestaan, is de vraag of hij dit fysisch onmogelijk acht of metafysisch onmogelijk. Het is duidelijk dat de huidige mens ondersteund word door zijn emoties en emoties lichamelijk fenomenen zijn, maar geld dit in alle mogelijke werelden (metafysisch onmogelijk)?

Sigmund Freud

De vraag die men kan stellen is of de geest wel zo toegankelijk is. Freud maakt een einde aan die gedachte door te stellen dat het bewuste een louter oppervlakte fenomeen gestuurd en gecontroleerd is door een massieve onpersoonlijke macht (meer zie: psychopathologie)

Gilbert Ryle

Ryle verwijst ook naar Freud in zijn tekst, op p. 2. Volgens Descartes, zegt Ryle, zouden mensen geen illusies mogen koesteren over hun eigen verlangens; zouden ze hun eigen overtuigingen en motieven niet mogen miskennen. Maar dat doen ze toch. Repliek van de S-D: ‘in normal circumstances…’ Maar wat zijn ‘normale omstandigheden’? Een van Freuds belangrijkste stellingen is dat ieder van ons een onbewuste heeft, dat ieder van ons bepaalde voorstellingen en verlangens verdringt, en dat ieder van ons met andere woorden met neurotische symptomen te kampen heeft. Normaliteit bestaat niet, zegt Freud.

* 1. **Tekst Gilbert Ryle: Descarts’ Mythe**

Eerst en vooral is er een onderscheidt in behaviorisme. Dit is belangrijk om aan te duiden tot welk soort behaviorisme Gilbert Ryle behoord. Burrhus Skinner (1904-1990) & John Watson (1878-1958): methodologisch behaviorisme. Reactie op introspectie-psychologie; nadruk op publiek observeerbare data. Visie op menselijke geest als black box: ondoorzichtige doos met complexe neurale mechanismen die op de een of andere manier ervoor zorgt dat bepaalde stimuli leiden tot bepaalde responsen. Verhaal van Kleine Albert (Watson).

Gilbert Ryle (1900-1976): logisch (of filosofisch of analytisch) behaviorisme. Het probleem van lichaam en geest berust op een groot misverstand – een misverstand dat door middel van een goede conceptuele analyse aan het licht kan worden gebracht. Relatie tussen een mentale toestand en gedrag is niet een relatie van wederzijdse veroorzaking. Niet een relatie tussen twee substanties. Het gaat om een conceptuele relatie. Hoofdpijn, bijvoorbeeld, staat in een conceptuele relatie tot bepaalde soorten gedrag (in medicijnkast naar dafalgan zoeken, kreunen, etc). Als je het gedrag begrijpt, begrijp je ook de hoofdpijn. Het gedrag is eenvoudigweg de hoofdpijn – er is niet anders dan gedrag. ‘Elke zin over een mentale toestand kan geparafraseerd worden zonder verlies van betekenis, in een lange en complexe zin over wat het observeerbaar gedrag van iemand zou zijn, als de persoon in kwestie in bepaalde omstandigheden zou verkeren’ (Dooremalen, 22). Vraag: kan je angsten en verlangens van Rattenman beschrijven in logisch-behavioristische termen? Het dreigt een lange lijst te worden. ‘Ryle lijkt te ontkennen dat we een werkelijk innerlijk leven hebben, wat gewoonweg zeer onwaarschijnlijk is: het lijkt er toch echt op dat we dat hebben’.

De tekst is opgedeeld in 3 luiken, met name: schets substantiedualisme en kritiek, centrale probleem substantie dualisme; de categoriefout, oorsprong van de categoriefout

Schets substantiedualisme en kritiek

Niets verrassend in het eerst deel. Ryle schets het substantiedualisme door een mooi beeld te vormen van de twee verschillende werelden (geest en lichaam). Denk aan het vorige voorbeeld van de twee treinen.

Zijn kritiek geeft hij door het bespreken van een aantal tegen-argumenten van substantiedualisme. Eerste argument dat hij aanhaalt is dat van interactie via het beeld van een pluimpje (shuttlecock) toont hij aan dat interactie van een andere soort moet zijn dan zowel lichaam als geest. (mind Descartes levensgeesten). Tweede argument dat hij bespreekt is het argument van andere geesten. Hij definieert het aan de hand van robinson crusoe-geesten. Allen op een apart mentaal eiland, maar niemand toegang tot elkaars mentale eiland.

Zijn conclusie is merkwaardig en de moeite om even bij stil te staan. De eerste premisse gaat als volgt: wij zijn allen in staat om met enige precisie mentale toestanden te beschrijven. Tweede premisse: substantiedualisme beschrijft mentale als een soort spook eilanden. Conclusie Substantie dualisme maakt het onmogelijk om iets te zeggen over geesten.

Merkwaardig hieraan is dat eerste premisse en de conclusie een paradox vormen. Op basis van die paradox stelt Ryle dat substantiedualisme een slechte theorie is.

Categoriefout

Ryle is duidelijk: substantiedualisme is een grote misvatting. De fout die ze maakt is een categoriefout. Men maakt een categoriefout als men een bepaald iets onder een logisch type, maar eigenlijk onder een ander logisch type zou moeten staan. Duiden van de verschillende categorieën in een voorbeeld: Soldaat => regiment => bataljon => compagnie => leger

Denk aan de jongen die langsgaat op een universiteit, een ganse rondleiding krijgt langs de gebouwen,… en dan vraagt, maar waar is de universiteit nu? (categoriefout, hij ziet de universiteit als een alternatieve substantie die in de werkelijkheid een apart object is, los van wat hij al gezien heeft)

Waarom begaat Descartes dan een categorie-fout? Hij spreekt over de geest als een ‘ding’ (res cogitans) of een ‘substantie’, en mentale toestanden als ‘processen’, terwijl die toestanden eigenlijk moeten begrepen worden als disposities om bepaalde handelingen te stellen. Wanneer Descartes over het mentale spreekt gebruikt hij gewoon de taal van de wetenschappen van zijn tijd, die erg succesvol waren en daarom een aantrekkelijk paradigma (zie sectie 3). Daarom spreekt Descartes over het mentale met termen als ‘ding’, ‘spul’, proces’, ‘verandering’, ‘oorzaak’, ‘effect’, etcetera (zie laatste zinnen van sectie 2 van Ryle’s tekst). Maar die manier van kijken is verwarrend, en leidt tot allerlei problemen. Ryle’s voorstel is om mentale toestanden te omschrijven als een verzameling van specifieke handelingen. Verdriet is huilen, het hoofd in de handen nemen, in elkaar stuiken, jammeren, enzovoort. Deze handelingen zijn het mentale. Naast deze handelingen is er NIET nog zoiets als een mentale toestand ‘verdriet’, die deze handelingen veroorzaakt. Verdriet IS deze handelingen.

* Universiteit | allerlei gebouwen & diensten
  + Geest | allerlei handelingen en uitspraken
  + Ryle: Geest = handelingen en uitspraken

Oorsprong categoriefout

‘Obverse’ betekent ‘omgekeerd’ of ‘omgedraaid’. Geest wordt beschreven als een substantie zoals materie, maar dan gewoon in negatieve termen. Ryle: ‘The workings of minds had to be described by the mere negatives of the specific descriptions given to bodies; they are not in space, they are not motions, they are not modifications of matter, they are not accessible to public observation. Minds are not bits of clockwork, they are just bits of not-clockwork’ (6). Hij noemt de geest dan ook ‘een spookmachine’ (a spectral machine’).

Ryle redeneert als volgt (onderaan 6): filosofen moeten taalgebruik van gewone man respecteren of corrigeren. Gewone man kan onderscheid maken tussen rationeel en irrationeel gedrag. Descartes verschijnt op het toneel, probeert rekening te houden met intuïties van gewone man, maar gaat in zijn enthousiasme stap te ver. Zijn oplossing laat ons immers niet meer toe om onderscheid tussen rationeel en irrationeel gedrag te maken. Want hoe weet ik of de ander rationeel is? Ik heb toch geen toegang tot zijn geest?

1. Verschillende soorten materialisme: pro en contra
   1. **Reductionistisch fysicalisme**
      1. **wat is fysicalisme?**

Een eerste definitie

De stelling dat het universum opgebouwd is uit een enkel soort spul, namelijk fysisch spul.

Anders geformuleerd: alles is fysisch. Het is mogelijk om alle kenmerken en handelingen van een organisme te beschrijven met behulp van de fysica. Meer nog: deze beschrijving maakt alle andere beschrijvingen overbodig. Er valt verder niets te vertellen. We hebben met andere woorden geen psychologische beschrijvingen nodig, zoals verlangens, overtuigingen, redenen, pijn, enzovoort. Bijkomende vragen: a) wat bedoelen we met ‘alles’ in ‘alles is fysisch’? Ook God? Abstracte concepten? Getallen? Eigenschappen? En b) wat bedoelen we met ‘is’ in ‘alles is fysisch’? Sommigen zeggen: het is een verklaringsrelatie. X=Y betekent dat X verklaard kan worden met behulp van Y. Anderen zeggen: het is een ontologische relatie. X=Y betekent dat alle eigenschappen van X ook eigenschappen van Y zijn. Of nog: dat alle eigenschappen van X gedetermineerd zijn door eigenschappen van Y.

Bij fysicalisme is een hoofdrol weggelegd voor de fyisica. Naturalisten blijven vaag in het concept fysica en zeggen dat het iets natuurlijk is.

Probleem is echter welk soort fysica men bedoelt. Er zijn verscheidene fysische theorieën. Sommige theorieën zijn nog niet volledig. Andere theorieën zijn nog niet compatibel. Mawson haalde de kwantumfysica aan om het dualisme nog een kans te geven. Vraag die dus centraal staat is over welke fysica we het precies hebben.

Intermezzo

Powers of Ten’, een film van de bekende meubeldesigners Charles & Ray Eames, is een mooie illustratie van het geloof in de vooruitgang en het wonder van de wetenschappen. De film (1977) is gebaseerd op het boekje van de Nederlander Kees Boeke, *Wij in het Heelal, Een Heelal in Ons* (1959). Een Engelse versie van het boekje vind je hier: http://www.vendian.org/mncharity/cosmicview

De tekst van Smart (‘Sensations and brain processes’) dateert ook van 1959, en draagt eenzelfde geloof in de wetenschappen uit.

Het hele universum (planeten, sterren, melkwegstelsels,…) is opgebouwd uit eenzelfde bescheiden set van bouwstenen: electronen, protonen, quarks, enzovoort. De fysica beschrijft het gedrag van deze bouwstenen, en de wetten die ze volgen wanneer ze samenklitten tot grotere eenheden: atomen, moleculen, cellen, enzovoort. Met behulp van deze kennis over bouwstenen en wetten kan de fysica zelfs voorspellen hoe deze bouwstenen zich zullen gedragen in specifieke situaties.

Historische achtergrond

Waar we voor de moderne tijd een organisch wereldbeeld hadden, komt in de moderne tijd het mechanische wereldbeeld op. Dit wil zeggen dat het hele universum als een machine is. Het is een fysische werkelijkheid. Verschil met de premoderne wereld is het verdwijnen van doeloorzaken. Dit wil zeggen dat bijvoorbeeld een zaadje in zijn essentie heeft zicht te actualiseren. Dit potentieel is een boom worden. Zo heeft elk organisme, artificieel en natuurlijk een eigen doel in zich.

Wanneer we iets willen begrijpen komt dat op het volgende neer: in staat zijn om iets te demonteren, de verschillende delen te ontleden en de mechaniek te begrijpen, net zoals je een raderwerk of een bermbom zou ontleden. Voor alle duidelijkheid: zowel dualisten als materialisten hebben zo’n mechanistische kijk op de werkelijkheid ( note: waarin verschillen ze dan?).

Descartes had verschillende redenen om de geest te vrijwaren van een mechanisch beeld. De eerste materialisten moesten dus rekening houden met Descartes zijn bezorgdheden. Dit zijn onder andere de subjectieve aspecten van de geest, mentale veroorzakingen, intentionaliteit,…

Eerste aanzet tot materialisme kwam van Hobbes en De la Mettrie. Bij die laatste werd ook de geest gepoogd om te materialiseren. De redenering van De la Mettrie is alsvolgt. De geest wordt samen met ons geboren, groeit samen met ons op, takelt samen met ons af en sterft met ons (ons in de zin van ons lichaam).

Drie soorten fysicalisme

* Reductionistisch fysicalisme (deze paragraaf)
* Niet-reductionistisch fysicalmisme (2de paragraaf)
* Eliminatevistisch fysicalisme (3de paragraaf)

Verschil tussen drie soorten hangt voornamelijk af van hun opvatting over het belang niet-fysische wetenschappen bij de beschrijving en verklaring van fenomenen. Voorbeelden van niet-fysische wetenschappen: folk psychology, wetenschappelijke psychologie, biologie, enzovoort. Denk terug aan voorbeeld uit eerste les: smetvrees. Anders geformuleerd: hoe belangrijk zijn niet-fysische beschrijvingen van mentale fenomenen? Er zijn drie antwoorden op die vraag, en elk van die antwoorden weerspiegelt een specifieke episode in de geschiedenis van de wetenschappen:

Eliminativistisch F: mentale fenomenen bestaan eenvoudigweg niet. We dwalen wanneer we ons gedrag, en het gedrag van anderen, proberen te vatten met behulp van termen als ‘pijn’, ‘verlangens’ of ‘overtuigingen’. Eliminativisten zijn hardcore, en vergelijken de psychologie (zowel folk psychology als wetenschappelijke psychologie) met mythologische verklaringen uit de antieke oudheid. ‘Trying to describe and explain human behavior by appeal to mental states is like trying to describe and explain the weather by appeal to the Greek gods’ (J 72). Vroeg of laat zal blijken dat een psychologisch discours volstrekt overbodig is. Moeilijker geformuleerd: psychologie heeft beschrijvende noch verklarende legitimiteit.

Overeenstemmende gebeurtenis uit de geschiedenis van de wetenschappen: flogiston-theorie over brandbaarheid van materialen. Hoe komt het dat sommige materialen brandbaar zijn en andere niet? Verklaring: brandbare materialen bevatten flogiston, een kleurloze, geurloze, smaakloze en massaloze substantie. Asse is dan het brandbare materiaal minus de flogiston. Toen aan het eind van de negentiende eeuw bleek dat zuurstof een voldoende verklaring bood voor de brandbaarheid van materialen, zeiden chemici niet dat flogiston en zuurstof naar eenzelfde (fysische) substantie verwezen, dat het gewoon twee verschillende beschrijvingen waren van eenzelfde fenomeen. Integendeel: flogiston werd gewoon in de vuilbak gegooid. De flogiston-theorie werd eenvoudigweg geëlimineerd.

Reductionistisch F: mentale fenomenen bestaan, maar vroeg of laat zal blijken dat ze netjes overeenkomen met fysische gebeurtenissen. Onze gewone psychologische (*folk psychology*) beschrijvingen van pijn, verlangens, en overtuigingen, zullen uiteindelijk overeenkomen met fysische beschrijvingen. Het zijn gewoon twee verschillende raamwerken of vocabulaires om naar eenzelfde (fysisch) fenomeen te verwijzen. In de tekst van Smart: ‘talking of an experience in terms appropriate to physical processes is not a change in our present rules for the use of experience words but an addition to them’ (152). (Reductionistisch betekent: uiteindelijk zullen we alles kunnen beschrijven en verklaren met behulp van de fysica)

Overeenstemmende gebeurtenis uit de geschiedenis van de wetenschappen: ontdekking van chemische structuur van water. We gebruikten al eeuwenlang de term ‘water’ om te verwijzen naar een drinkbare, vloeibare, doorzichtige substantie. Toen bleek dat moleculaire structuur van water H²O is, hebben we niet beslist om te stoppen met spreken over water. Integendeel: we zijn ervan overtuigd dat ‘water’ en ‘H²O’ naar eenzelfde fenomeen verwijzen, namelijk H²O.

Niet-reductionistisch F: vanaf de jaren 1970 begonnen filosofen te beseffen dat verhouding tussen fysica en de andere wetenschappen (en folk psychology) misschien toch wat ingewikkelder is dan reductionistische F dachten. Voor alle duidelijkheid: ook niet-reductionistische F zijn ervan overtuigd dat alles, werkelijk alles kan beschreven en verklaard worden door middel van de fysica. Maar: er zijn verschillende manieren om de wereld te beschrijven, afhankelijk van waarin we precies geïnteresseerd zijn. Als we op zoek gaan naar ijzeren wetten waaraan een fenomeen gehoorzaamd, dan kan de fysica ons helpen. Maar soms zijn we geïnteresseerd in andere aspecten van een fenomeen. En dan kunnen we een beroep doen op andere wetenschappen. Deze wetenschappen hebben ons iets te bieden dat de fysica ons niet kan bieden, en kunnen daarom niet tot de fysica gereduceerd worden.

Voorbeeld van smetvrees. A) Er bestaat niet zoiets als smetvrees; het is gewoon een manier van spreken. Er beantwoordt niets in de werkelijkheid aan het woord ‘smetvrees’. B) Uiteindelijk zal blijken dat smetvrees netjes overeenkomt met een of ander fysisch fenomeen in onze hersenen (een gebied of een gebeurtenis in onze hersenen). C) een psychologische beschrijving van smetvrees levert ons iets op dat een fysische beschrijving niet kan bieden.

* + 1. **Varianten van reductionistisch fysicalisme**

Filosofisch behaviorisme

De term behaviorisme heeft verschillende betekenissen.

In de psychologie bestaat er enerzijds methodologisch behaviorisme. Anderzijds bestaat er een radicalere vorm van behaviorisme. De algemene claim bij beide is dat psychologen zich bezig moeten houden met observeerbare fenomenen, namelijk inputs en outputs, niet wat zich daartussen afspeelt. (input= zintuigen => [MIND] => output= gedragingen). Achterliggende reden is dat de psychologie geen vooruitgang boekte met de fenomenologische aanpak in de dertigerjaren. De methodologische behavioristen pleiten voor een komaf met de methode van introspectie en pleiten voor werken met objectieve, publieke en meetbare data. De radicale behavioristen pleiten niet alleen voor eenzelfde aanpak, ze beweren dat er tussen input en output werkelijk niets is. Mentale fenomenen bestaan dus niet (Link met eliminativisten).

In de filosofie spreekt men echter van analytisch of logisch behaviorisme. Waarnaar verwijzen we wanneer we mentale concepten gebruiken? Waarnaar verwijzen ervaringen (*experiences*), propositionele attitudes en mentale handelingen (actions)? Ryle maakte al duidelijk dat het een categoriefout is om te veronderstellen dat mentale concepten verwijzen naar een afzonderlijke ‘mentale realiteit’ (de geest, om te begrijpen denk aan voorbeeld universiteit). Psychologische beschrijvingen zijn eigenlijk afkortingen van fysische beschrijvingen van actueel of mogelijk gedrag. Grappig voorbeeld uit J 105. (om dit duidelijker te begrijpen gebruik zeker de tekst van J.J Smart)

Behaviorisme is fysicalistisch om dat men zegt dat gedragingen vertalingen zijn van mentale toestanden. We kunnen ze dus beschrijven als fysisch fenomeen.

Argumenten pro FB

1. Net zoals de aanhangers van de identiteitstheorie zich beroepen op het succes van de wetenschappen, in het bijzonder de fysica, verwijzen logisch behavioristen graag naar het succes van het logisch positivisme. De basisgedachte van het logisch positivisme is dat de logica ons kan helpen om aan te tonen dat filosofische problemen eigenlijk gebaseerd zijn op conceptuele verwarring. Ryles oplossing voor het probleem van geest en brein past in deze strategie (‘it’s a category mistake, stupid!’). Strijdpunt van de logisch positivisten: betekenis = verificatie (meaning = verification). Opnieuw Ryle: de betekenis van woorden voor mentale toestanden (‘verlangen’, ‘overtuiging’,…) = begeleidend verifieerbaar gedrag.
2. We zagen eerder al dat zowel substantie-dualisten als fysicalisten (zie de tekst van Mawson) moeite hebben om te verklaren hoe het komt dat we er van overtuigd zijn dat andere mensen ook een geest hebben, en dat we in het algemeen erg goed zijn in het toeschrijven van mentale toestanden aan andere mensen. Door andere mensen te observeren kunnen we met enig succes te weten komen hoe ze zich voelen, wat ze denken, enzovoort. Nochtans hebben filosofen altijd beweerd dat het een essentieel kenmerk is van mentale toestanden dat ze subjectief zijn, dat wil zeggen enkel toegankelijk voor het individu dat de toestand vertoont. Het behaviorisme lost dit probleem op, zij het op een nogal botte manier. Volgens het behaviorisme is er helemaal niets subjectief aan mentale toestanden. Integendeel: je kan zulke toestanden eenvoudigweg vertalen naar publiek observeerbare, fysische gedragingen.

Argumenten contra FB

Allerlei problemen met basisprincipes van het logisch positivisme, waaruit het logisch behaviorisme voortgekomen is. Een voorbeeld: het principe dat betekenis verificatie is, is een claim die zelf niet verifieerbaar is. Andere argumenten tegen het behaviorisme:

1. Het herformuleren van mentale toestanden als gedragsdisposities is een eindeloos karwei. Ervaringen (*experiences*) laten zich nog relatief eenvoudig vertalen – denk maar aan toestanden als plezier, pijn, en de emoties. Maar wat met propositionele attitudes? Bij beschrijvingen van overtuigingen en verlangens wordt het vertaalwerk een heuse uitdaging. Daar is een goede reden voor, die we al eens aangekaart hebben ivm de vraag of dieren overtuigingen hebben: overtuigingen (en bij uitbreiding: mentale toestanden) doen zich meestal in netwerken voor. Om ene overtuiging te duiden hebben je andere overtuigingen of verlangens nodig, die op hun beurt… Vergelijking met opzoeken van woord in woordenboek. ‘In practice, then, it would appear to be impossible for us to translate even a single psychological expression into a statement purely about actual and potential behavior’ (J 109).
2. Het probleem van a) wordt pijnlijk duidelijk wanneer we ons toespitsen op die derde categorie van mentale toestanden: de mentale handelingen (of *actions*). Sommige van deze handelingen zijn toestanden die zich voordoen zonder fysische uitlopers. Een rekensom, bijvoorbeeld, kan ik uitvoeren zonder daarbij mijn handen of andere ledematen te gebruiken. Als er geen gedragingen zijn om te beschrijven, betekent dat dan dat er zich geen mentale toestand kan voordoen?
3. Het behaviorisme beweert dat het mentale kan hertaald worden met behulp van gedragsdisposities. Die positie is in zekere zin plausibel wanneer we de mentale toestanden van andere mensen proberen in te schatten. Daarbij baseren we ons namelijk op wat ze doen – op hun gedrag. Onze eigen mentale toestanden zijn daarentegen wel onmiddellijk toegankelijk. Ik heb mijn gedrag niet nodig om te weten dat ik pijn heb. Door het subjectieve tussen haakjes te plaatsen (*einklammerung*), maakt het behaviorisme zich echter kwetsbaar voor kritiek. Je kan je namelijk situaties voor de geest halen waarin iemand doet alsof (een clown, een acteur, de presentator van een candid camera-programma), zodat het lijkt alsof hij of zij een bijhorende mentale toestand heeft. In zo’n geval brengt het gedrag ons op een dwaalspoor.

Probleem van causaliteit: mentale toestanden kunnen niet de oorzaak zijn van lichamelijke toestanden volgens het behaviorisme. Mentale toestanden vallen immers samen met lichamelijke toestanden. Het behaviorisme ontkent de mogelijkheid van causaliteit, terwijl wij wel ervaren dat onze geest ons lichaam in beweging brengt (en omgekeerd).

excursus

Behaviorisme: a priori, door middel van conceptuele analyse. Wanneer we de concepten ontleden die we dagelijks gebruiken om onze eigen mentale toestanden en de mentale toestanden van anderen te beschrijven en te verklaren, dan kunnen we enkel besluiten dat die concepten verwijzen naar ons gedrag, en het gedrag van anderen

Identiteitstheorie

Identiteitstheorie: a posteriori, door middel van empirische wetenschappen (hersenwetenschappen). Vroeg of laat zullen de hersenwetenschappen ons in staat stellen om alle mentale toestanden te ‘vertalen’ naar gebeurtenissen in of eigenschappen van de hersenen. Net zoals de chemie ons geleerd heeft dat ‘water’ eigenlijk verwijst naar een samenvoeging van twee waterstofatomen en een zuurstofatoom: H²O. Interessant is dat deze wetenschappelijke ontdekking het woord ‘water’ niet overbodig maakte – we kunnen het met een gerust hart blijven gebruiken, want ‘water’ en ‘H²O’ verwijzen naar precies dezelfde fysische entiteit, namelijk H²O.

c) Theoretische identificatie: een identificatie van twee elementen waarbij minstens één element behoort tot een theorie. In ons geval: fysische toestanden zijn theoretische elementen, want ze behoren tot de jurisdictie van de fysica. Een kleine nuance: het identificeren van mentale toestanden met fysische toestanden kan op twee manieren gebeuren. Sommige fysicalisten *kiezen* om M met F te identificeren. Die keuze is een keuze voor eenvoud (denk even terug aan het scheermes van Ockham). Dit is de piste die Smart volgt in zijn tekst. ‘Why do I wish to resist the suggestion [that the report “I am in pain” refers to something irreducibly psychical]? Mainly because of Occam’s razor’ (142). Andere fysicalisten menen dat de identificatie niet zozeer een keuze is, maar wel een implicatie van de logica van wetenschappelijk onderzoek. Volgens hen kiezen we er niet voor om M en F te identificeren – die identificatie is gewoon de logica zelve! Neem water als voorbeeld: je kan water definiëren als ‘het spul in deze fles’. De wetenschapper kan dat spul identificeren als H²O. Dus de wetenschapper identificeert water met H²O. Het principe achter deze redenering is dat identiteit transitief is. Als x=y en y=z, dan x=z. Toegepast op ons probleem, het probleem van geest en brein. A) mentale toestanden (x) zijn toestanden die veroorzaakt worden door … en die … veroorzaken (y). B) Y = hersentoestanden. C) Mentale toestanden = hersentoestanden.

Argumenten pro IT

We kaderen hier drie argumenten in het voordeel van identiteitstheorie.

Intermezzo: niet te vergeten het succes van materialisme is omgekeerd evenredig met dualisme. Argumenten tegen dualisme spelen met andere woorden in het voordeel van materialisme (note: probeer de argumenten tegen dualisme te plaatsen voor welke vorm van fyisicalisme ze een voordeel zijn en probeer specifieker te kaderen in de varianten van die vormen= goede oefening).

Voorbeelden:

b) Het scheermes van Ockham leert ons dat wanneer twee theorieën eenzelfde verklaringskracht hebben, eenzelfde coherentie, enzovoort, dan verkiezen we de theorie met de meest eenvoudige ontologie. De ‘meest zuinige’ theorie. Het probleem met dit argument is enerzijds dat het onduidelijk is of F inderdaad even coherent is als SD en ED, en of F eenzelfde verklaringskracht heeft als SD en ED, en anderzijds dat het argument van zuinigheid geen doorslaggevend argument is. Het is best mogelijk dat een eenvoudiger theorie niet beter is dan een ingewikkelde theorie.

a) mentale veroorzaking (*mental causation*) is een probleem voor substantie-dualisme en eigenschapsdualisme, maar niet (of toch niet op het eerste gezicht) voor fysicalisme. Het is een probleem voor SD en ED omdat mentale toestanden of eigenschappen zich per definitie onttrekken aan een fysisch universum, en dus ook niet kunnen communiceren met fysische toestanden. F beweert dat mentale toestanden identiek zijn aan fysische toestanden, waardoor het mogelijk wordt dat geest en hersenen informatie uitwisselen.

Eerste argument: succes van de wetenschappen

Alle pogingen uit het verleden om fenomenen te beschrijven met behulp van niet-fysische entiteiten of eigenschappen, zijn mislukt. En alle pogingen uit het verleden om fenomenen te beschrijven met behulp van fysische entiteiten of eigenschappen, zijn gelukt. Voorbeelden zijn legio: magnetisme (mysterieuze geesten versus electromagnetische krachten), geestesziekten (duivelse bezetenheid of ontregelde neurotransmitters), leven (vitalistische krachten versus metabolische processen in de cel). De fysica heeft een buitengewone verklarende en voorspellende kracht. Andere wetenschappen spiegelen zich aan deze discipline, en baseren zich op ontwikkelingen uit de fysica om indrukwekkende technologieën te ontwikkelen, het menselijk lichaam in kaart te brengen, en ziekten te genezen die de mens al duizenden jaren teisteren. Het fysicalisme (in het debat over lichaam en geest) spiegelt zich aan de wetenschappen, en heeft aan die bewondering ten dele haar populariteit te danken. Net als zoveel andere fenomenen zal de geest vroeg of laat kunnen gevat worden door de fysische wetenschappen.

Tekst van Smart: ‘That everything should be explicable in terms of physics (…) except the occurrence of sensations seems to me to be frankly unbelievable’ (142).

MAAR: het argument dat fysicalisme gelijk heeft omdat de wetenschappen in het verleden vaak gelijk bleken te hebben, is enkel een inductief argument. Het geeft ons weliswaar redenen om te denken dat het fysicalisme gelijk heeft, maar geen doorslaggevende redenen. Inductieve argumenten kunnen gebaseerd zijn op juiste premissen, maar toch foute conclusies opleveren. Een voorbeeld. De premisse is: alles zwanen die we tot dusver hebben kunnen observeren, zijn wit. Dus alles zwanen zijn wit. De premisse is waar, en toch is de conclusie onwaar: er zijn namelijk ook zwarte zwanen, maar die zaten verstopt in Australië. Een substantie-dualist zou op een gelijkaardige manier kunnen tegenwerpen dat wetenschappers inderdaad erg succesvol zijn in het bedenken van fysische verklaringen voor fysische fenomenen. Maar deze fenomenen vormen een eerder beperkt domein. Het succes van de wetenschappen in het verleden sluit niet uit dat er niet-fysische fenomenen bestaan.

Intermezzo: Sommige psychiaters betwijfelen of het ‘hersenverhaal’ over geestesziekten een echte vooruitgang is ten opzichte van het ‘heksenverhaal’. Antipsychiater Thomas Szasz, bijvoorbeeld, argumenteerde dat het concept ‘geestesziekte’ (als een fysische stoornis) een mythe is, net als het concept ‘hekserij’. I have tried to show that the notion of mental illness has outlived whatever usefulness it might have had and that it now functions merely as a convenient myth. As such, it is a true heir to religious myths in general, and to the belief in witchcraft in particular; the role of all these belief systems was to act as social tranquilizers, thus encouraging the hope that mastery of certain specific problems may be achieved by means of substitutive (symbolical-magical) operations’ (Szasz 1960, 118).

Tweede argumenten pro identiteitstheorie: aansluiting bij common sense

*What you see is what you get*. Ons gezond verstand zegt ons dat de dingen die we kunnen observeren ook echt zijn. Omgekeerd hebben we, in het algemeen, moeite om geloof te hechten aan dingen die we niet kunnen waarnemen (in de brede zin van het woord: met onze zintuigen). Hetzelfde geldt voor mysterieuze Cartesiaanse geesten – waarom zouden we aannemen dat zulke entiteiten bestaan als we ze niet kunnen waarnemen? ‘Materialism represents (as dualism claims to do) nothing more than the drawing out of the unavoidable implications of some home-spun common sense’ (F 52).

MAAR: het lijkt misschien alsof de wetenschappen in het verlengde liggen van het gezond verstand, en het gezond verstand verder begronden, maar niets is minder waar. Materiële objecten zijn samengesteld uit kleine bouwstenen die door de fysica beschreven kunnen worden, maar hun soliditeit is een illusie. Eigenlijk gaat het om zwermen van onobserveerbare deeltjes, die elk minder ruimte innemen dan de lege ruimte tussen de deeltjes. We zien zulke zwermen niet, en toch hechten we er geloof aan. Waarom doen we dat dan niet bij Cartesiaanse geesten? (note: vergelijk ook het argument pro dualisme in verband met de common sense)

Derde argument pro IT (argument van denkbaarheid)

Argument van denkbaarheid (fysicalistische variant): teletransportatie. Lees F 59 (dit is de tekst die op.

Wat is verschil met dualistisch argument van denkbaarheid? Denk aan gedachte-experimenten van Descartes, van W. Hart & het fenomeen van bijna-dood ervaringen. Dualistische versie had een negatief doel: aantonen dat geest en brein gescheiden substanties zijn. Fysicalistische versie heeft een eerder positief doel: aantonen dat geest en brein identiek zijn (of beter: dat mentale toestanden identiek zijn aan fysische toestanden). De vraag is of de fysicalistische versie van het argument een belangrijke rol kan spelen. De gedachte dat een identieke kopie van mijn hersenen ook gepaard gaat met een geest, toont niet keihard aan dat mijn hersenen en mijn geest identiek zijn – hoe nauw ze ook verbonden zijn, zoals de neurowetenschappen ons leren. Verhaal van vuur en rook. Waar er rook is, is ook vuur. Maar dat betekent niet dat rook identiek is aan vuur.

Argumenten contra IT

Ook hier zullen we drie argumenten postuleren. Eerste argument is het argument van de zombies, tweede het dilemma van Hempel en als laatste argument het kennisargument van Jackson( zal besproken worden onder eliminatevistisch fysicalisme. Probeer het argument te kaderen waarom het ook tegen Identiteitstheorie zou zijn.)

Eerste argument contra: de filosofische zombies.

Wanneer we spreken over zombies zijn dat niet de zombies die in horrorfilms opduiken. Integendeel, het zijn fysieke identieke kopieën van individuen die identiek dezelfde uitspraken en handelingen stellen. Die identieke kopieën hebben echter geen qualia. Om het met een voorbeeld te duiden, De zombie Vergouwen spreekt op dezelfde manier over zijn Collega De pater, schrijft op dezelfde manier met zijn blauwe stift op een bord over Logica, maar weet niet hoe het voelt de wind door zijn vettige haren waait of hoe het is om een foto van zijn gerookte longen te laten nemen. Men spreekt soms wel van qualiazombies.

De redenering loopt als volgt. Als filosofische zombies mogelijk zijn (mind de theorie over de verschillende vormen van mogelijkheid in vorige hoofdstukken, we spreken hier over metafysisch), dan is fysicalisme onwaar. Filosofische zombies zijn mogelijk dus fysicalisme is onwaar.

De voorgaande redenering is plausibel om de volgende redenen: Waarom moeten we de premisse “filosofische zombies zijn mogelijk” geloven? Opnieuw een argument van denkbaarheid (*conceivability*): het is denkbaar dat zulke zombies bestaan, dus zijn ze ook metafysisch mogelijk. Dit argument is kwetsbaar voor dezelfde kritiek als het argument van denkbaarheid dat voorstanders van het substantie-dualisme aanhangen (supra). Ten eerste kan je de band tussen denkbaarheid en mogelijkheid in vraag stellen. En ten tweede is het de vraag of je je werkelijk kan voorstellen dat er zoiets bestaat als qualia zombies. Het blijft uiteindelijk ‘maar’ een gedachte-experiment.

Waarom moeten we “als filosofische zombies mogelijk zijn dan is fysicalisme onwaar” geloven? Draaien we de zaak even om. Als fysicalisme waar is, dan zijn er maar twee mogelijkheden: ofwel bestaan qualia niet, ofwel zijn ze gewoon identiek aan hersentoestanden. Laten we even aannemen dat qualia bestaan (Daniel Dennett zal het tegendeel beweren, maar dat is voor later). Dan moeten ze identiek zijn aan hersentoestanden. Maar het zombie-argument toont net aan dat het mogelijk is dat er wezens bestaan die identiek dezelfde hersentoestanden hebben als ik, en toch geen qualia. Qualia kunnen dus geen hersentoestanden zijn, en bijgevolg is het fysicalisme onwaar.

Tweede argument pro: het dilemma van Hempel.

Wat betekent het precies dat alles kan beschreven worden en verklaard worden met behulp van de fysica? Over welke fysica gaat het dan precies? De fysica van Heisenberg en Einstein verschilt radicaal van die van hun voorgangers, Galileo en Newton, en er zijn mensen die geloven dat de fysica in de toekomst nog allerlei gedaanteverwisselingen en koerswisselingen zou kunnen ondergaan. Fysicalisten zitten met een dilemma, zegt Hempel, wanneer ze deze vraag beantwoorden. Ze hebben twee mogelijkheden:

1. De fysicalist: ‘Het gaat om de fysica in haar huidige vorm’. Maar, zegt Hempel, in dit geval is het fysicalisme onwaar. Waarom? Net als vele andere wetenschappen is de fysica voortdurend in ontwikkeling. Vroegere fysische theorieën worden gefalsifieerd door nieuwe experimenten, en vervangen door nieuwe fysische theorieën, enzovoort. Het gevolg is dat de zeventiende-eeuwse fysica grondig verschilt van de negentiende-eeuwse fysica, die op haar beurt verschilt van de huidige fysica. Wanneer de fysicalist beweert dat alles exhaustief beschreven en verklaard kan worden door de fysica, dan moet ‘ie aangeven welke versie van de fysica hij verkiest. Wanneer de keuze valt op de huidige fysica, dan kunnen we enkel besluiten dat het fysicalisme onwaar is. Waarom? Omdat we weten dat de huidige fysica minstens ten dele nog gefalsifieerd zal worden door latere versies van de fysica. Bijgevolg is het onwaar dat alles exhaustief beschreven kan worden met de huidige fysica. En dat probleem zal blijven bestaan tot er een definitieve fysische theorie geformuleerd wordt –

een theorie die niet meer gefalsifieerd kan worden.

1. De fysicalist: ‘Het gaat om de fysica in haar afgeronde, ideale vorm’. Maar, zegt Hempel, in dit geval is het fysicalisme een lege theorie. Want de eerlijkheid gebiedt de fysicus te zeggen dat hij of zij geen flauw benul heeft van hoe die finale fysica er zal uitzien. Het houdt met andere woorden geen steek om te zeggen dat de finale fysica een exhaustieve beschrijving en verklaring kan geven van alles wat is, precies omdat we niet weten wat die finale fysica precies inhoudt. Om het punt wat kracht bij te zetten: er zijn dualisten die beweren dat het mogelijk is dat er ooit een versie van de fysica zal bestaan die niet-fysische entiteiten of eigenschappen zal postuleren. David Chalmers, bijvoorbeeld, die u zo dadelijk te zien krijgt op de foto met Jack Smart, is één van hen. Hij noemt zichzelf een naturalistische dualist. En waarom ook niet? We hebben immers geen idee hoe de fysica zich verder zal ontwikkelen. Smart kan enkel mopperen dat nieuwe wetten (of entiteiten) zeer ongewoon zouden zijn: ‘Such ultimate [psychophysical] laws would be like nothing so far known in science. They have a queer “smell” to them’ (143). Het is eigenlijk niet meer, zoals Smart zelf zegt, ‘a confession of faith’ (143).

Derde argument contra: kennisargument van Jackson (zie eliminativistisch fysicalisme + Tekst van Jackson *what mary didn’t know”*

* + 1. **Tekst sensations and brain processes van J.J. Smart**

Gilbert Ryle die een logisch behaviorist is (besproken in kader van contra dualisme) poneerde enkele argumenten tegen het dualimse. De crux van zijn argument tegen het dualisme bestond erin dat dualisten spreken over de geest alsof die tot eenzelfde logische categorie behoort als het lichaam. Het zijn overigens niet alleen dualisten die zich bezondigen aan deze categoriefout, maar ook gewone mensen. De verwarring situeert zich op het niveau van het spreken. Ryle’s analyse is ook het vertrekpunt voor Smart: we moeten op onze hoede zijn voor beweringen als ‘ik heb pijn’ of ‘ik verlang een zeldzame radio te vinden’, want zulke beweringen lijken te verwijzen naar een substantie voorbij de fysische processen van pijn en verlangen.

Smart deelt echter Ryle’s oplossing niet – ‘it will not quite do the trick’ (144; KANTELPUNT van de tekst!). Volgens Ryle verwijzen beweringen als ‘ik heb pijn’ louter en alleen naar gedragsdisposities. Volgens Smart gebeurt er meer bij zulke beweringen. Hij wil het over een andere boeg gooien: ‘Why should not sensations just be brain processes of a certain sort?’ (144). Net als Lamettrie, die zijn positie als vraag formuleerde: ‘Laten we eens kijken hoever we komen wanneer we aannemen dat M=F’. De claim van Smart: wanneer we over ervaringen spreken (zoals pijn etc), dan verwijzen we eigenlijk naar hersenprocessen. Voorbeeld van bliksem. Smart had ook voorbeeld van water kunnen gebruiken, of licht. Punt is dat zowel ‘bliksem’ als ‘electrische ontlading’ naar eenzelfde fysisch fenomeen verwijzen. Idem dito voor ‘pijn’ en ‘het vuren van neuronen in deel van de hersenen’.

Het is niettemin duidelijk dat Smart rekening wil houden (in tegenstelling tot de behavioristen) met die aspecten van het mentale leven die Ryle buiten spel zet, zoals de subjectiviteit van mentale toestanden. De uitdaging is die aspecten een plaats te geven in een fysicalistisch raamwerk. Veel objecties kaarten precies dit probleem aan.

Merk op: Smarts fysicalisme is reductionistisch (in ppt over reductionistisch fysicalisme staat dat Smarts fysicalisme niet-reductionistisch is, echter in de ppt van niet-reductionistisch schijnt dat JJ smart reductionistisch is. Als we de objecties lezen blijkt het tweede te kloppen. Hij stelt namelijk dat ervaringen van pijn hersentoestanden zijn. Dit wil zeggen dat hij het concepten van mentale toestanden gelijkstelt aan concepten van hersentoestanden in de zin van waarnaar ze refereren. Smart behoort dus tot de aanhangers van de Identiteitstheorie Indien bemerkingen posten in de facebook groep!!!). Het alledaags taalgebruik is niet overbodig of verward – de alledaagse (of pre-wetenschappelijke) beschrijving en de wetenschappelijke beschrijving verwijzen allebei naar eenzelfde fysisch fenomeen.

De tekst is als volgt opgebouwd. Smart poneert enkele objecties waarom hersenprocessen en ervaringen niet identiek zouden zijn. In zijn replies is het echter de bedoeling om die te weerleggen. Hij wil niet zeggen dat bijvoorbeeld pijn een hersenproces is, maar wel de ervaring van pijn een hersenproces is. In tegenstelling tot wat Ryle doet, wanneer die zegt dat pijn louter gedragsdisposities zijn en niets anders. Hij spreekt van een conceptuele ervaring. Wanneer wij spreken, volgens Ryle over pijn dan is dit perfect vertaalbaar in (reductionistische visie, meerbepaald logisch behaviorisme) analytisch-behavioristische taal zonder verlies van betekenis.

Objectie 1

Meeste mensen kunnen ervaringen beschrijven. Veel mensen weten niets over hersenprocessen. Dus ervaringen en hersenprocessen zijn niet identiek.

Smarts repliek:

Achtergrond: een taalfilosofische kwestie. Link tussen taalfilosofie en *philosophy of mind*. De kwestie is het onderscheid tussen *Sinn* & *Bedeutung* van Gottlob Frege uit 1892. Vertaald als ‘Sense & meaning’ in het Engels. Ik vertaal het in het Nederlands als ‘Betekenis & verwijzing’. Er is namelijk een onderscheid tussen aan de ene kant de **betekenis** van een woord of een zin, en aan de andere kant de toestand of stand van zaken waar het woord of de zin naar **verwijst**. Er is duidelijk een verschil (in betekenis) tussen de naam ‘Tom Boonen’ en de bijnaam ‘de bom van Balen’, maar beide uitdrukkingen verwijzen naar dezelfde persoon – Tom Boonen. Hetzelfde geldt voor de planeet Venus. Mensen verwijzen naar deze planeet met twee verschillende uitdrukkingen: de ochtendster en de avondster.

Waarom haalt Smart het onderscheid van Frege aan? De redenering is de volgende: zinnen met een verschillende betekenis kunnen toch eenzelfde referent hebben, en zo is het ook met het verschil tussen E-zinnen (zinnen over ervaringen, zoals pijn) en HP-zinnen (zinnen over hersenprocessen). Uiteraard verschillen beide soorten van zinnen van elkaar, dat wil zeggen hun betekenis is verschillend, maar dat belet niet dat ze naar eenzelfde proces verwijzen, namelijk een hersenproces.

Objectie 2

(Smarts repliek steunt op het zelfde onderscheid van betekenis en verwijzing.)

Link tussen ervaringen en hersenprocessen is contingent. Daarom kunnen beiden niet identiek zijn. Contingent is de tegenstelling van noodzakelijk. Met andere woorden iets is contigent als het even goed anders had kunnen zijn.

Smart repliceert op deze objectie dat wanneer we bijvoorbeeld pijn proberen te verwoorden dat niet hetzelfde betekend als zeggen de neuronen in mijn hersenen zijn aan het vuren. Echter het wil niet zeggen dat beide zaken niet verwijzen naar eenzelfde fenomeen namelijk dat pijn hersenprocessen zijn. (zelfde repliek als bij objectie 1)

Zulke redenering is gebaseerd op een fido-fido theorie, dit wil zeggen dat de betekenis van een zin volledig vervat ligt in hetgeen naar waar het verwijst.

Objectie 3

Steunt op de wet van Leibniz. Dit argument is een moeilijk argument en Smart geeft zelf aan dat dit het hardste argument was om te weerleggen. Smarts reply echter lijkt niet helemaal te kloppen.

Achtergrond: Verschil tussen primaire en secundaire kwaliteiten.

Onderscheid komt van John Locke (Essay Concerning Human Understanding; 1690). Materiële objecten hebben vijf primaire kwaliteiten: uitgebreidheid, vorm, beweging (of rust), aantal, en stevigheid. Daarnaast hebben materiële objecten ook secundaire kwaliteiten: kleur, smaak, geur, geluid & temperatuur. Secundaire kwaliteiten omschrijft Locke als krachten van het object om bepaalde zintuiglijke effecten bij ons teweeg te brengen. Een van de belangrijkste verschillen tussen primaire en secundaire kwaliteiten is dat secondaire kwaliteiten erg afhankelijk zijn van de ‘eigenaardigheden’ van het subject. Met andere woorden: in vergelijking met primaire eigenschappen zijn secundaire eigenschappen erg subjectief.

Ervaringen (experiences of sensations) zijn een categorie van mentale toestanden die onder andere de fenomenen van gewaarwording (waarneming in de brede zin), emotie en pijn bevatten. Bij gewaarwordingen worden we ons bewust van secundaire kwaliteiten – we zien kleuren, horen geluiden, snuiven geuren op, enzovoort. Deze secundaire kwaliteiten hebben qualia – je kan er namelijk de vraag over stellen hoe het is om ze te ervaren. Hoe is het om geel te ervaren? Hoe is het om de geur van versgemaaid gras te ruiken? Hoe is het om een madeleine-koekje te proeven? Smart noemt deze qualia ‘phenomenal properties’. Als het waar is dat mentale toestanden, zoals gewaarwordingen, eigenschappen hebben die materiële objecten, zoals het menselijk brein, niet hebben, dan zit Smart met een probleem (beste formulering: 149, voorlaatste paragraaf).

Waarom? Enter de wet van Leibniz. Als x werkelijk identiek is met y, dan betekent dat dat alle eigenschappen van x ook eigenschappen van y zullen zijn. De vraag is dan: hebben hersentoestanden ook qualia? De assumptie (WAAROM?) is dat dat niet het geval kan zijn. Daarom kunnen mentale toestanden (sensations) niet identiek zijn aan hersentoestanden.

Reply:

kwestie is dat wanneer we een uitspraak doen over een quale (of over meervoud: qualia), bijvoorbeeld de uitspraak ‘ik heb pijn’, dan ligt in die uitspraak de claim vervat dat mijn ervaring een *mentale* toestand is. Smart claimt daarentegen dat dat niet zo hoeft te zijn. We kunnen de uitspraak in kwestie (‘ik heb pijn’) ook herformuleren zodanig dat die neutraal is over de aard van wat ervaren wordt. (lees de tekst pag 149-150 waar hij zijn replay uitlegt.)

objectie 4

Objectie 3 stelt dat ervaringen (sensations) eigenschappen hebben die hersenprocessen niet hebben, namelijk qualia. In deze objectie houdt Smart eenzelfde koers aan, en werpt hij op dat, omgekeerd, hersenprocessen eigenschappen hebben die ervaringen niet hebben. HP zijn namelijk gesitueerd in de ruimte, terwijl dat niet het geval is bij ervaringen.

Reply: Ignoratio elenchi is een drogreden waarbij de conclusie van een argument niet volgt uit de premissen. Lijkt een beetje op een non sequitur, maar met de engere betekenis van een conclusie die niet alleen niet volgt uit de premissen, maar ook irrelevant is ten opzichte van het thema van het debat. Een ad hominem argument is dan een voorbeeld van een ignoratio elenchi. Voor Smart betekent het: een argument dat misschien op zichzelf geldig is, maar niet aan de orde is in de huidige context. Smart beweert namelijk niet dat ‘pijn’ een hersenproces is, maar wel dat de ervaring van pijn een hersenproces is.

Objectie 5

Voorbeelden van primaire kwaliteiten: hersenprocessen kunnen snel zijn of traag (BEWEGING), rond of vierkant (VORM), stevig of slap (STEVIGHEID),… Zulke eigenschappen kan je niet toeschrijven aan mentale toestanden, zoals pijn. Is pijn snel? Of vierkant? Stevig of slap? Het lijkt onzinnig om die vragen te stellen.

Smarts repliek is tweevoudig:

1. ‘Pijn’ en ‘neurobiochemisch proces X’ zijn twee verschillende vocabulaires om naar eenzelfde fysisch fenomeen te verwijzen – denk aan Frege’s onderscheid tussen betekenis en verwijzing. Het zou best kunnen, zegt hij, dat uitspraken over ervaringen (sensations) helemaal anders zijn dan uitspraken over hersenprocessen. Maar dat doet niet ter zake. Soms is het zo dat de betekenis van twee uitspraken heel erg verschilt (avondster-ochtendster), en dat ze toch naar eenzelfde realiteit verwijzen (de planeet Venus). Zo zou het best kunnen dat ik op een heel andere manier spreek over ervaringen dan over hersenprocessen, maar dat betekent niet dat beide soorten uitspraken naar verschillende realiteiten verwijzen. Integendeel, zegt Smart: het is erg waarschijnlijk dat ze allebei naar hetzelfde verwijzen, namelijk hersentoestanden.
2. Waarom zou het per definitie onzinnig zijn om pijn te beschrijven met behulp van eigenschappen die typisch zijn voor primaire kwaliteiten (beweging, vorm,…)? Het is niet onwaarschijnlijk dat we vroeg of laat in staat zullen zijn om met behulp van de neurowetenschappen specifieke ervaringen te localiseren in onze hersenen. In plaats van ‘ik heb pijn’ zouden we dan zeggen: ‘er is iets aan de hand met mijn medio-frontale cortex’.

Objectie 6

Wat bedoelt men eigenlijk wanneer men zegt dat ervaringen ‘privé’ (of ‘privaat’) zijn, of ‘privé-toegankelijk’? Er wordt onder meer bedoeld dat zulke ervaringen niet publiek zijn, of niet publiek toegankelijk. Maar er wordt ook meer gezegd, namelijk dat ervaringen

1. altijd iemands ervaringen zijn (een pijn is altijd iemands pijn)
2. niet (op een strikte manier) gedeeld kunnen worden met anderen (ik kan niet eenzelfde pijn hebben als mijn grootmoeder)
3. onzichtbaar zijn (ik kan mijn grootmoeders pijn niet zien)
4. oncorrigeerbaar zijn (ik kan mijn grootmoeders pijn niet betwisten – IS DAT ZO? Wat doen we dan met aanstellers?)

Een bijkomende stelling is dat hersenprocessen totaal tegengesteld zijn aan ervaringen in al deze aspecten. Hersenprocessen zijn zichtbaar (met het nodige materiaal, zoals een scanner), zijn corrigeerbaar (ik kan een bepaalde opvatting hebben over een hersenproces dat vervolgens betwist wordt door een neurochirurg), en zijn ten slotte wel ‘deelbaar’ (zoals in: kunnen gedeeld worden) (ik kan me een situatie inbeelden waarin ik exact hetzelfde hersenproces heb als iemand anders – zoals het vuren van een bepaald neuron).

Reply: lijkt een beetje op de reply die hij formuleert bij objectie 5: de verschillen in onze manier van spreken over ervaringen en hersenprocessen hoeven niet definitief te zijn. Het is mogelijk dat de neurowetenschappen zodanig zullen vorderen dat we individuele hersenprocessen kunnen registreren en lokaliseren, en dan is er geen sprake meer van ‘onzichtbaarheid’ of ‘oncorrigeerbaarheid’.

Note: objectie 3,4,5,6 zijn allemaal gebaseerd op de wet van Leibniz: Indien 2 zaken identiek zijn. Dan zijn de eigenschappen van X ook eigenschappen van Y

Objectie 7

Smart haalt opnieuw het onderscheid tussen betekenis en verwijzing aan, maar hij had ook kunnen tegenwerpen dat niemand zich kan voorstellen steen te zijn en toch pijn te hebben. Het gedachte-experiment is absurd.

Objectie 8= NIET KENNEN (feest! Tijd voor een pauze)

* 1. **Niet-reductionistisch fysicalisme**

Inleiding

*The distinctions introduced by the predicates and terms of the special sciences are not distinctions written in the book of nature as it comes off the press, but instead notes we jot in the margins – our commentary on a text that is written in the language of physics’ (William Jaworski).*

een citaat van William Jaworski dat de eigenheid van het niet-reductionistisch fysicalisme treffend verwoordt. Je kan het vrij vertalen als: ‘de fenomenen die beschreven worden door de categorieën van de speciale wetenschappen [de niet-fysische wetenschappen, inclusief de psychologie] zijn niet de fenomenen die beschreven staan in het grote boek van de natuur, wanneer het vers van de pers komt. Het zijn integendeel kleine notities die we neerschrijven in de kantlijn van het boek – onze commentaar op een tekst die helemaal opgesteld is in de taal van de fysica’.

Vertrekpunt: kunnen we op een of andere manier ontsnappen aan het vermaledijde reductionisme van het behaviorisme (mentale toestanden zijn NIETS MEER dan gedragingen) en de identiteitstheorie (mentale toestanden zijn NIETS MEER dan hersentoestanden)? Denk even terug aan de smetvrees van de dwangneuroot. Kunnen we op een of andere manier tegemoet komen aan onze intuïtie, en de intuïtie van Damiaan Denys, dat smetvrees MEER is dan het knetteren van neuronen? En kunnen we tegemoet komen aan die intuïtie zonder daarom de hersenen te negeren – zoals dualisten wel eens plegen te doen? Niet-reductionistische fysicalisten beweren van wel.

* + 1. **Wat is niet-reductionistisch fysicalisme?**

Opnieuw wat is fysicalisme?

De verschillen tussen deze drie soorten van fysicalisme hebben te maken met hun opvatting over de relatie tussen de fysica en de niet-fysische wetenschappen, in het bijzonder de wetenschappelijke psychologie en de *folk psychology*.

*eliminativistisch*. Eliminativisten menen dat de niet-fysische wetenschappen vroeg of laat van het toneel zullen verdwijnen, omdat de concepten en categorieën waarmee ze werken niet in de werkelijkheid te vinden zijn. Concepten als ‘pijn’ of ‘verlangen’ zijn waardeloos, of zullen waardeloos blijken te zijn, omdat ze nergens over gaan. Het eliminativisme is een eerder controversiële positie (zie volgende week).

*reductionistisch*. Het verschil tussen eliminativisme en reductionisme is in wezen een nuance-verschil. Beide soorten van fysicalisme zijn ervan overtuigd dat de fysica uiteindelijk het laatste woord heeft, en dat de niet-fysische wetenschappen hoogstens een marginale rol kunnen spelen in de totstandkoming van onze kennis. Reductionisten mogen dan al enige waardering hebben voor de niet-fysische wetenschappen, maar uiteindelijk, zo zeggen ze, zal de fysica alles kunnen beschrijven en verklaren, ook de fenomenen die nu nog beschreven of verklaard worden door de niet-fysische wetenschappen. Uiteindelijk is alles fysisch, remember? Ook mentale toestanden zijn niet meer dan fysische toestanden.

*niet-reductionistisch*. Niet-reductionistische fysicalisten zijn tegen zo’n totale reductie gekant. En die tegenkanting kan je op twee manieren formuleren. Ten eerste menen niet-R fysicalisten dat de relatie tussen mentale toestanden en fysische toestanden (zoals gedragingen of hersentoestanden) veel complexer is dan reductionisten doen uitschijnen. De claim dat mentale toestanden netjes kunnen ‘vertaald’ worden als fysische toestanden bleek niet zo vanzelfsprekend te zijn – zie vorige les. Ten tweede is niet-R fysicalisme ook een claim over de relatie tussen fysische en niet-fysische wetenschappen. Niet-R fysicalisten menen dat de beschrijvingen en de verklaringen van niet-fysische wetenschappen niet zomaar kunnen overgenomen worden door de fysica. Waarom niet? Omdat niet-fysische wetenschappen, zoals de psychologie, ons iets bieden dat de fysica niet kan bieden. Wat dat ‘iets’ precies is moet tijdens deze les duidelijk worden. Laten we eerst even het probleem van reductionisme aankaarten. Wat betekent die term precies, en waarin verschil reductionisme van niet-reductionisme?

Reductionisme

De vraag ‘wat is reductionisme?’ of ‘wat is reductionistisch fysicalisme?’ kan je opdelen in een aantal verschillende vragen. Ik noem er hier drie:

Wat wordt er precies gereduceerd in het R fysicalisme? Eerste antwoord: niet-fysische wetenschappen worden gereduceerd tot de fysica. Meer bepaald: de beschrijvingen en verklaringen van niet-fysische wetenschappen kunnen worden vervangen door **beschrijvingen en verklaringen** van de fysica. En omdat verklaringen eigenlijk deducties zijn van algemene wetmatigheden, zou je kunnen zeggen dat ook de wetten van niet-fysische wetenschappen gereduceerd worden tot de **wetten** van de fysica. En ten slotte is het ook zo dat de **entiteiten** die door de niet-fysische wetenschappen beschreven en verklaard worden, zoals verlangens en overtuigingen, uiteindelijk gereduceerd kunnen worden tot entiteiten die beschreven en verklaard worden door de fysica, zoals atomen, moleculen, en hersencellen. R fysicalisme beweert dus dat de beschrijvingen, verklaringen, wetten en entiteiten van niet-fysische wetenschappen kunnen herleid worden tot de beschrijvingen, verklaringen, wetten en entiteiten van de fysica. Zulke claims lijken misschien weinig opzienbarend op het eerste gezicht, maar de gevolgen zijn behoorlijk radicaal. Denk even terug aan de eerste les, en het onderscheid tussen psychoanalytische en hardcore neurofysiologische beschrijvingen van smetvrees. R fysicalisten beweren dat psychoanalytische beschrijvingen en verklaringen van smetvrees volledig kunnen overgenomen worden door beschrijvingen en verklaringen die we ontlenen aan de fysica. Freud sloeg de bal niet mis met zijn beschrijvingen en verklaringen, zoals eliminativisten zouden claimen, maar uiteindelijk zal zijn werk overbodig blijken te zijn. Waarom overbodig? Omdat alles wat Freud zei ook gezegd kan worden in de taal van de fysica.

Wat betekent het concreet om de ene wetenschap tot de andere te reduceren? Kijken we even naar de verschillende betekenissen van ‘reduceren’ in Van Dale. Ik pik er drie uit die relevant zijn in deze context. ‘Reduceren’ betekent in de eerste plaats ‘verminderen’ of ‘verkleinen’. Zoals wanneer je een verpakking reduceert, of denk aan de ‘reductie-kaart’ van de BGJG, waarmee je korting krijgt op de trein, en in bepaalde winkels. Theoretische reducties, zoals de reductie van de niet-fysische wetenschappen tot de fysica, houden in zekere zin ook een ‘verkleining’ in (zie ook het volgende punt). Na de reductie eindigen we immers met minder beschrijvingen en verklaringen, minder wetten en minder entiteiten. Er is sprake van een vereenvoudiging. Een tweede betekenis van ‘reduceren’ is ‘herleiden’ of ‘omrekenen’. In de numerologie, bijvoorbeeld, worden alle samengestelde getallen ‘gereduceerd’ tot hun grondgetal. 84 = 8+4 = 12 = 1+2 = 3 Numerologen noteren: 84/3. (Voor de geïnteresseerden: net als astrologen horoscopen maken, maken numerologen numeroscopen. En zo’n numeroscoop ‘vertelt (…) ons duidelijk wat iemands levensles is, wat iemands proces en bestemming is, zelfs de bedoeling van iemands leven op Aarde’ (www.ditrianum.org). Nou moe.) Een laatste betekenis van ‘reduceren’ is, volgens Van Dale, ‘zuurstof onttrekken aan’, voornamelijk in de context van chemische reacties. Deze betekenis is intrigerend in deze context, want bij een intertheoretische reductie (de reductie van een theorie tot een andere theorie) wordt onvermijdelijk ook zuurstof onttrokken aan de gereduceerde theorie. De bestaansreden van de ene theorie lijkt weg te vallen. Nogmaals: waarom zouden we Freud nog lezen als we een fysische beschrijving en verklaringen kunnen geven van smetvrees?

Waarom zouden we niet-fysische wetenschappen reduceren tot de fysica? Waarom zouden we Freuds werk aan de kant schuiven en ons op de fysica storten? Het antwoord van Smart (en reductionisten in het algemeen): omdat het zuiniger is. Waarom blijven werken met een arsenaal aan wetenschappelijke beschrijvingen en verklaringen als je ook rond komt met een enkele wetenschap, namelijk de fysica?

Niet-reductionisme

Niet-reductionistische fysicalisten zijn het eens met reductionistische collega’s dat de fysica alles exhaustief (of volledig) kan beschrijven en verklaren en dat niettemin zowel de psychologie als de fysica bestaansrecht hebben.

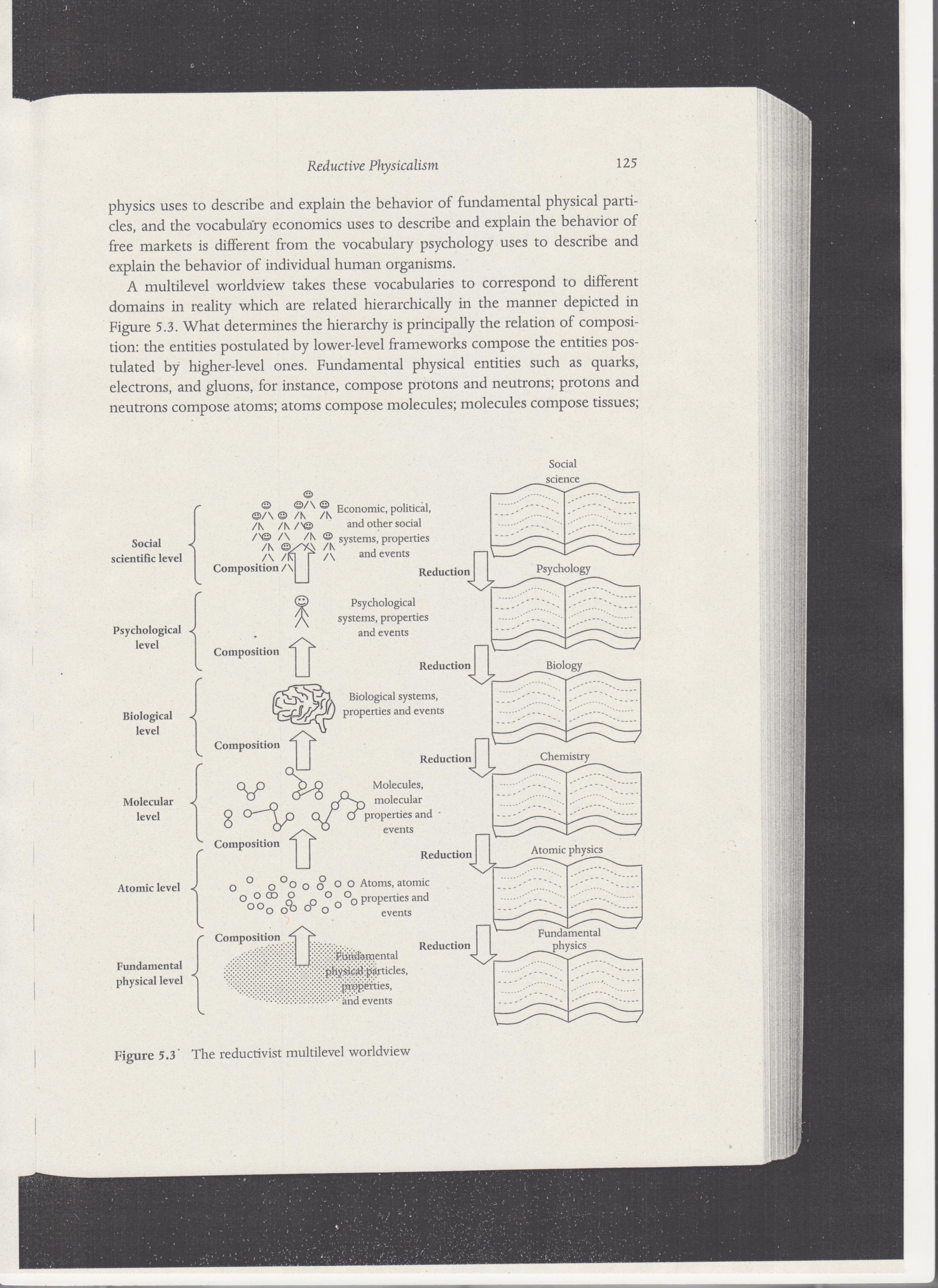
Het meningsverschil tussen R fysicalisten en niet-R fysicalisten gaat over de vraag of je niet-fysische wetenschappen kan reduceren (zie reductionisme supra) tot de fysica. De fysica beschrijft en verklaart de meest elementaire entiteiten, eigenschappen en wetten van de werkelijkheid, en die komen niet altijd overeen met de entiteiten, eigenschappen en wetten van de werkelijkheid zoals die beschreven wordt door bijvoorbeeld de sociologie, de biologie of de psychologie. De categorieën van ‘macht’ (in de sociologie), ‘leven’ (in de biologie) en ‘pijn’ (in de psychologie) maken geen deel uit van de woordenschat van de fysicus, en komen niet netjes overeen met de categorieën van de fysica. De sociologie, biologie en psychologie zijn daarom autonome wetenschappen – ze kunnen niet onderling gereduceerd worden, en ook niet tot de fysica. ‘The categories of fundamental physics enable us to describe and explain the behavior of the compositionally most basic entities there are, and to formulate the most general laws, but we are not always interested in describing and explaining the most basic entities and formulating the most general laws. Sometimes we have other interests in mind, and according to nonreductive physicalists the categories of the special sciences enable us to satisfy them’ (J 147).

De vraag die zich dan stelt is als alles fysisch is (zoals fysicalisten, ook de niet-reductionistische fysicalisten, beweren) en als de psychologie de werkelijkheid beschrijft met categorieën die niet overeenkomen met de categorieën van de fysica, hoe kan het dan dat onze beschrijvingen overeenkomen met de fysische werkelijkheid? Kunnen de niet-fysische wetenschappen ware kennis opleveren, gegeven het feit dat ze niet reduceerbaar zijn tot de fysica?

Er zijn in het algemeen twee antwoorden op die vraag, en die komen overeen met de twee varianten van niet-reductionistisch fysicalisme die we in het vervolg van deze les bespreken. Eerste antwoord is het antwoord van het functionalisme. Functionalisten beweren dat de niet-fysische wetenschappen niet iets beschrijven of verklaren dat de fysica niet kan beschrijven of verklaren. Ze beschrijven of verklaren niets nieuws; ze doen het gewoon op een andere manier. Meer bepaald op een abstracte manier. De beschrijving van ‘pijn’, bijvoorbeeld, is een abstracte beschrijving van hersentoestanden, die op hun beurt abstracte beschrijvingen zijn van chemische reacties, etc In deze optiek is alleen de beschrijving van fysische entiteiten, eigenschappen en wetten een concrete beschrijving. Mooie omschrijving van Jaworski: ‘The distinctions introduced by the predicates and terms of the special sciences are not distinctions written in the book of nature as it comes off the press, but instead notes we jot in the margins – our commentary on a text that is written in the language of physics’ (J 148). Anders gezegd: de categorieën van de niet-fysische wetenschappen komen misschien niet netjes overeen met de categorieën van de fysica, maar ze zijn evenmin totaal verschillend: de categorieën van de niet-fysische wetenschappen *worden gerealiseerd door* de categorieën van de fysica. Net zoals een computerprogramma (het mentale) zich realiseert (en zich ook enkele kan realiseren) met behulp van computerhardware (de hersenen) => hierom compatibel met fysicalisme.

Tweede antwoord: superveniëntie-fysicalisme (*supervenience physicalism*). De niet-fysische wetenschappen slagen erin om toch de fysische werkelijkheid te beschrijven omdat hun categorieën superveniëren op de categorieën die beschreven worden door de fysica. Wat is het verschil tussen een realisatie-relatie (het mentale realiseert zich in het fysische) en een superveniëntie-relatie (het mentale supervenieert op het fysische)? Die vraag is niet zo makkelijk te beantwoorden. Mijn prioriteit is om u bekend te maken met de basisbegrippen en argumenten van zowel het functionalisme als het superveniëntie-denken

Reductionistsche kijk op de wereld.



Wereld bestaat uit verschillende lagen, die elk overeenkomen met een bepaalde wetenschappelijke discipline. Elke discipline heeft z’n eigen onderzoeksmethoden en protocollen, zijn eigen woordenschat, zijn eigen principes etc

Er is niettemin een hiërarchie in de verhouding van de wetenschappen, en die wordt bepaald door een relatie van ‘samenstelling’ (*composition*). Moleculen zijn bijvoorbeeld samengesteld uit atomen, en atomen zijn samengesteld uit de meest fundamentele bouwstenen van het universum, zoals die beschreven worden door de fysica. En net zoals moleculen gereduceerd kunnen worden tot atomen, kan de chemie, als wetenschappelijke discipline, gereduceerd worden tot de fysica.

* + 1. **Varianten van niet-reductionistisch fysicalisme**
       1. Functionalisme

Functionalisme is de claim dat mentale toestanden geen fysische toestanden zijn, maar functionele toestanden. Het zijn toestanden die ons in staat stellen om inputs op een gepaste manier te verbinden met outputs. Minder abstract: mentale toestanden zijn toestanden die ons in staat stellen om adequaat te reageren (output) op zintuiglijke informatie (input), en die ook altijd verbonden zijn met specifieke andere mentale toestanden. Pijn, bijvoorbeeld, is een mentale toestand die veroorzaakt wordt door schade aan het lichaam (in een glasscherf stappen, met je hoofd tegen een balk stoten, je vingers verbranden aan een strijkijzer, enzovoort), en die resulteert in typische vormen van gedrag (allerlei verwensingen roepen, je vingers onder stromend water houden, pleisters plakken, enzovoort). Bovendien is pijn als mentale toestand verbonden met allerlei andere mentale toestanden (verlangen dat de pijn verdwijnt, overtuigd zijn dat het helpt om je verbrande vingers onder koud water te houden, verdriet, enzovoort). Mentale toestanden maken dus altijd deel uit van een ingewikkeld web van inputs, outputs en andere mentale toestanden.

In hun hoedanigheid van functionele toestanden zijn mentale toestanden vergelijkbaar met nieren, flessenopeners, en rekenmachines. Ze vervullen een bepaalde functie, en hoe ze die vervullen doet er eigenlijk niet toe. Een kunstnier kan de functie van een echte nier vervullen; flesjes kan je ook met aanstekers openen, of met je tanden; en rekensommetjes kan je met een computerchip oplossen of met een ouderwetse ‘rekenaar’ (met tandwielen etc). Hilary Putnam zei ooit, in 1975: ‘We could be made of Swiss cheese and it wouldn’t matter’. Een andere vergelijking: het organigram van een goed geolied bedrijf bestaat allereerst uit functies. Er is de functie van CEO, de functie van management assistants, secretaresses, enzovoort. Hoe die functies ingevuld worden (duobaan? Freelancers?) en door wie (man? vrouw? computer?) doet er eigenlijk niet toe. Het zal u niet verwonderen dat het functionalisme gestoeld is op de stelling dat mentale toestanden door verschillende ‘dragers’ kunnen gerealiseerd worden, maar daarover dadelijk meer.

De meest bekende metafoor waarvan het functionalisme zich bedient om haar visie op de aard van mentale toestanden te beschrijven, is de metafoor van de software. Het mentale, zo zeggen functionalisten, vormt eigenlijk de software van het brein. Computernerds weten dat computersoftware programma’s bevatten, en zulke programma’s zijn eigenlijk gewoon ellenlange scripts van omschrijvingen van taken. Een voorbeeld van zo’n omschrijving: ‘als input x, dan output y’, waarbij ‘x’ bijvoorbeeld staat voor ‘het indrukken van de spatiebar’, en ‘y’ voor ‘voeg een spatie in de tekst’ of ‘wissel van geweer’ of ‘zet het filmpje op pauze’). Een computerprogramma kan bedacht worden los van een computer, maar uiteindelijk heeft het de hardware van een computer nodig om zich te realiseren. Idem dito voor de menselijke geest. Je kan de geest op een abstracte manier beschrijven, maar het heeft wel een drager nodig om zich te realiseren. Functionalisme is dan ook een variant van fysicalisme.

Alan Turing: turing machine

Een wat complexere definitie van functionalisme brengt ons bij het werk van Alan Turing. Putnam zei ooit dat mensen Turing-machines zijn. Wat bedoelde hij met die vergelijking?

Alan Turing was een Britse wiskundige die in de jaren 1930 furore maakte met een theorie over ‘rekenaars’, tijdens de tweede wereldoorlog een centrale rol speelde in het kraken van de Duitse Enigma-code, en enkele jaren later veroordeeld werd voor homoseksualiteit.

Wat is een Turing-machine? Voor Turing was het niet zozeer een fysieke machine (zoals een computer of een lasapparaat), maar eerder een wiskundig concept. Turing droomde van een soort universele machine die elke mogelijke reeks van instructies zou kunnen uitvoeren (anders gezegd: ‘een machine die alles zou kunnen berekenen’). Later heeft men allerlei pogingen ondernomen om Turings concept concrete gestalte te geven (zie bijvoorbeeld volgende slide). Een Turing-machine bestaat uit drie onderdelen: a) een lange tape waarop allerlei symbolen geschreven en gewist kan worden (het geheugen van de machine), b) een bewegende kop, die de symbolen op de tape kan lezen en bewerken, en c) een instructieboek (of machine-tabel) waarin vermeld staat welke actie moet ondernomen worden bij welk symbool (het programma van de machine). De machine bevindt zich ook altijd een bepaalde interne toestand.

Waarin lijken mensen precies op Turing-machines? Net als zulke machines kan je mensen omschrijven door middel van inputs, outputs, interne toestanden, en de regels om al die elementen met elkaar te verbinden. Anders gezegd: mentale toestanden zijn functionele toestanden, meer bepaald computationele toestanden van een Turing-machine. De claim dat de menselijke geest moet begrepen worden naar analogie met een Turing-machine is misschien nog wat vaag, maar zal wellicht duidelijker worden in de les over modulariteit. Een aantal elementen van onze geest, zegt Fodor, zijn modulair. Dat wil zeggen: het zijn informatieverwerkingsmechanismen die er op gericht zijn een enkele functie uit te voeren. De manier waarop die informatie verwerkt wordt, doet inderdaad denken aan de manier waarop de Turing-machine te werk gaat. Maar daarover dus meer in de les over modulariteit.

Turing test

Waarom is de Turing-test relevant in deze context? Het functionalisme beweert, zoals we dadelijk zullen zien, dat mentale toestanden zich kunnen realiseren in verschillende materialen. Het hebben van mentale toestanden is met andere woorden niet het voorrecht van wezens met hersenen, zoals wij. Ook computers zijn in staat om mentale toestanden te hebben. Ook computers, zegt de functionalist, kunnen denken. De Turing-test geeft ons een criterium om te bepalen of computers inderdaad kunnen denken.

De setting: een proefpersoon staat via een chatbox in verbinding met enerzijds een menselijk wezen, en anderzijds een computer. De proefpersoon stelt een aantal vragen aan beide ‘conversatiepartners’, en zowel de mens als de computer beantwoorden de vragen. Een computer slaagt voor de Turing Test wanneer de proefpersoon na het vragenuurtje niet in staat is om de mens te onderscheiden van de computer, dat wil zeggen wanneer de computer even intelligent overkomt als de proefpersoon. Een computer die, in zekere zin, de schijn kan hooghouden dat hij (of zij?) een mens is van vlees en bloed, slaagt voor de Turing-test.

Sinds 1991 wordt er een jaarlijkse wedstrijd georganiseerd, de Loebner Prize Competition, waarin computerprogramma’s het beste van zichzelf mogen geven, en waarin gepoogd wordt om hen door de Turing-test te loodsen. Turing voorspelde zelf dat computers in staat zouden zijn om te denken voor het jaar 2000, maar er is geen eensgezindheid over een winnaar. Critici van de Turing-test beweren dat er programma’s zijn die de test al lang doorstaan hebben. Het satirische computerprogramma Eliza, bijvoorbeeld, kwam al in de jaren 1970 als winnaar uit de bus. Het was een programma dat ontworpen was om het gebrek aan intelligentie van een bepaalde soort psychotherapeuten aan te tonen (voor de liefhebbers: de Rogeriaanse psychotherapeuten). Het programma herhaalde eenvoudigweg wat de patiënt zei, en vroeg af en toe naar wat meer details.

Argumenten pro

Argument van meervoudige realiseerbaarheid

Mentale toestanden zijn meervoudig realiseerbaar, zo zeggen niet-reductionistische fysicalisten. Wat bedoelen ze daarmee? Hersenen zijn volgens hen niet het enige soort fysisch ‘spul’ (*stuff*) waarin mentale toestanden zich kunnen realiseren/verwezenlijken/verwerkelijken. Het is best mogelijk dat andere planeten in het universum levende wezens bevatten met mentale toestanden die erg lijken op onze mentale toestanden, en die niettemin geen hersenen hebben. Dat wil zeggen: het zou best kunnen dat hun mentale toestanden gerealiseerd worden met behulp van silicium-circuits, zoals onze computers, of met behulp van een materiaal dat ons niet bekend is, maar overvloedig aanwezig is op hun planeet. Functionalisten verwijzen in deze context vaak naar de geschiedenis van de computer. De eerste computers deden hun rekenwerk (*computation*) met behulp van metalen tandraderen, met kleppen, en later pas met silicium-circuits.

Merk op: er is geen strikt verband tussen functionalisme en fysicalisme, ook al zijn de meeste functionalisten de facto ook fysicalisten. Mentale toestanden zijn interne toestanden die inputs koppelen aan outputs – het zijn toestanden die zich kenmerken door een specifieke functie. Die functie kan in principe door verschillende dragers gerealiseerd worden: hersenen, silicium-chips, een alsnog onbekend chemisch element, maar ook een Cartesiaans ego. ‘Unlike the identity theory, then, functionalism is not committed to physicalism. It is compatible with physicalism, and in fact most functionalists are also physicalists, but functionalism does not imply physicalism. A functionalist could be a substance dualist, a dual-attribute theorist, or even an idealist’ (J 142).

Het argument van meervoudige realiseerbaarheid is een argument pro functionalisme, maar tegelijkertijd ook een argument tegen de identiteitstheorie. De identiteitstheorie stelt dat mentale toestanden identiek zijn aan fysische toestanden. Deze claim is een claim over types van mentale toestanden en types van fysische toestanden. Alle instantiaties van het type X zijn identiek aan hersenprocessen van het type Y. Een voorbeeld: elke toestand van pijn is identiek aan een fysisch proces in een specifiek gebied van de hersenen. Die claim kan je herformuleren als: iedereen die pijn heeft, heeft ook een hersenproces van het type Y. Het probleem met die claim is dat pijn meervoudig realiseerbaar is. Dat wil zeggen: pijn komt niet alleen overeen met een hersenproces van het type Y (in dit specifieke deel van de hersenen), maar ook met een hersenproces van het type Z (in een ander deel van de hersenen). Met andere woorden: pijn kan zich in principe realiseren in verschillende gebieden in de hersenen, of met verschillende soorten van hersenprocessen.

Er zijn twee soorten argumenten voor deze claim van meervoudige realiseerbaarheid. Enerzijds zijn er a priori argumenten, anderzijds a posteriori argumenten.

A PRIORI: opnieuw een argument van denkbaarheid. Niet-reductionistische fysicalisten zijn ervan overtuigd dat het denkbaar is dat voorbeelden van één bepaald type van mentale toestanden (‘types’ komt niet helemaal overeen met de soorten of categorieën van mentale toestanden uit de eerste les, met name ervaringen, propositionele attitudes, en handelingen; elke soort van mentale toestanden kan je nog in ‘types’ opdelen: de soort ‘ervaringen’ kan je bijvoorbeeld opdelen in types als ‘pijn’ en ‘angst’) overeenkomen met voorbeelden van verschillende types van fysische toestanden. Science-fiction verhalen hangen aan elkaar met allerlei vreemde wezens die een heel andere fysiologie hebben dan de onze, en toch in staat zijn om honger te hebben, pijn te voelen, en gelukkig te zijn – en dat zijn uiteraard mentale toestanden.

A POSTERIORI: we zagen eerder al dat denkbaarheidsargumenten vatbaar zijn voor allerlei tegenwerpingen, en dus hebben niet-reductionistische fysicalisten ook een aantal empirische argumenten bedacht om hun claim van meervoudige realiseerbaarheid kracht bij te zetten. Empirische argumenten komen uit verschillende wetenschappelijke disciplines. A) Biologie, B) neurowetenschappen (voorbeeld van plasticiteit van het menselijk brein), C) artificial intelligence (J 132-3).

Interessante bedenking: argument van meervoudige realiseerbaarheid is ook een argument tegen dualistische argumenten van denkbaarheid, zoals Descartes’ argument van het Malin Génie, of Harts argument van de lichaamloze geest. Waarom? Dualisten claimen dat men zich kan voorstellen dat een geest bestaat zonder lichaam, en dat het daarom ook (metafysisch) mogelijk is dat een geest bestaat zonder lichaam. Functionalisten kunnen die claim onderuit halen door te zeggen dat het misschien mogelijk is om zich voor te stellen dat een geest bestaat zonder hersenen, maar daarom niet *zonder eender welk fysisch systeem*. Het is inderdaad mogelijk om zich voor te stellen dat ook aliens mentale toestanden hebben, ook al hebben ze geen hersenen. Maar hun mentale toestanden worden dan gewoon door een ander materiaal gerealiseerd – een of andere slijmerige substantie die enkel terug te vinden is op hun planeet.

Reactie aanhangers identiteitstheorie

Verdediger van de identiteitstheorie zeggen dat het argument van meervoudige realiseerbaarheid gebaseerd is op de veronderstelling dat onze classificatie van mentale types (pijn, verlangen,…) en fysische types niet zal veranderen in de toekomst. Die veronderstelling is echter onjuist. Merk op: verdedigers van de identiteitstheorie doen geregeld beroep op toekomstvoorstellingen om hun zaak te verdedigen – denk maar aan het verhaal over de ‘volledige of ideale fysica’. De huidige stand van zaken is misschien problematisch, beamen ze. Bepaalde types van mentale toestanden lijken zich inderdaad te realiseren in verschillende types van fysische toestanden, zodat er geen één-op-één relatie bestaat tussen mentale en fysische toestanden. Maar de oorzaak van deze verwarring is dat onze classificatie van mentale toestanden nog niet op punt staat. Psychologen zullen in de toekomst ONZE toestand M opdelen in verschillende ‘subtypes’ van M (M¹, M², M³,…), die elk mooi overeenkomen met een specifiek type van fysische toestand P (P¹, P², P³,…).

Vergelijking met nood aan endofenotypes in de psychiatrie. Endofenotypes = verdere opdelingen van ziektecategorieën, zoals depressie of schizofrenie, om gerichter onderzoek te kunnen doen. Sommige filosofen van de psychiatrie claimen dat onze ziektecategorieën veel te breed zijn, zodat eenzelfde categorie eigenlijk gebruikt wordt om naar verschillende achterliggende aandoeningen te verwijzen.

Argument contra functionalisme

We zullen hier drie argumenten tegen de funtionalistische theorie van het reductionistisch fysicalisme poneren (mind probeer te schetsen in welke zin ze al dan niet voor dualisme zouden zijn of voor een andere theorie van reductionistisch fysicalisme of misschien zelfs algemener een andere vorm van fysicalisme)

Eerste argument contra: liberalisme

Een van de voordelen van het functionalisme is dat het een theorie is die ons toestaat om mentale toestanden toe te schrijven aan niet-menselijke wezens. Volgens de identiteitstheorie is het onmogelijk dat marsmannetjes en robotten pijn hebben, of emoties, zelfs als ze dezelfde gedragingen vertonen als mensen. Het functionalisme heeft een meer liberale (in de zin van: vrijgeviger, minder kleingeestig) opvatting over mentale toestanden. Wanneer een marsmannetje of een robot in staat is om op een gepaste manier inputs (vb. duimspijker in vinger) te verbinden met outputs (vb. uitroepen, vinger in mond steken,…), dan is hij of zij ook in staat om mentale toestanden te ervaren.

Maar is deze liberale houding van de functionalisten echt een voordeel? Of eerder een beperking? Anders gezegd: is het een argument voor of tegen het functionalisme? Het functionalisme stelt dat mentale toestanden in principe door eender welk materiaal gerealiseerd kunnen worden: menselijke hersenen, silicium-circuits, lasterstralen – you name it. Laten we even nagaan hoe plausibel die gedachte precies is. Een van de grote verwezenlijkingen van de geneeskunde is dat ze defecte onderdelen van het menselijk lichaam probleemloos kan vervangen – denk maar aan knieprothesen en valse tanden. Je kan je inbeelden dat de neurologie vroeg of laat ook een manier bedenkt om defecte onderdelen van het brein te vervangen. Individuele neuronen zouden bijvoorbeeld kunnen vervangen worden door silicone-neuronen, die precies dezelfde *functionele* kwaliteiten hebben als de ‘echte’ neuronen. Silicone-neuronen kunnen onderling communiceren zoals ‘echte’ neuronen, en ze kunnen communiceren met de verschillende onderdelen van het lichaam, net zoals ‘echte’ neuronen. Nu gaan we een stap verder. Je kan je een scenario inbeelden waarbij de silicone-neuronen vervangen worden door Chinezen, dat wil zeggen door inwoners van China (het voorbeeld komt van de filosoof Ned Block). Wanneer we er in slagen om de inwoners van China dezelfde functionele organisatie mee te geven als het oorspronkelijke brein, dan moeten we toegeven dat iemands mentale toestanden ook gerealiseerd kunnen worden door de inwoners van China. Of door bijen, of watermoleculen.

En dat lijkt een anomalie te zijn. Het houdt volgens Block geen steek te zeggen dat een samenleving in staat is om pijn te ervaren, of emoties te tonen. ‘Matter doesn’t matter’, zeggen de functionalisten. Het maakt niet uit waarin of waarmee mentale toestanden gerealiseerd worden. Het hoeven helemaal geen hersenen te zijn. Maar waar ligt dan de grens? Het lijkt absurd om te claimen dat iemands mentale toestanden gerealiseerd kunnen worden met behulp van Chinezen, en die claim volgt niettemin uit het functionalisme.

Tweede argument contra: chinese kamer

Opnieuw: wat zegt het functionalisme? Mentale toestanden zijn identiek aan functionele toestanden. En wat zijn functionele toestanden? Het zijn toestanden die een rol spelen in de verbinding van inputs en outputs. Uit deze centrale stelling van het functionalisme volgt de volgende claim: Wanneer twee organismen functioneel identiek zijn, dat wil zeggen wanneer ze inputs verbinden met outputs op exact dezelfde manier, dan moeten ze ook identieke mentale toestanden vertonen. Searles gedachte-experiment van de chinese kamer haalt deze claim onderuit, en dus ook de basisclaim van het functionalisme. Het argument ‘toont immers aan’ (het is en blijft een gedachte-experiment natuurlijk) dat twee organismen functioneel identiek kunnen zijn en toch mentaal verschillend. Hoezo dan?

Setting en bespreking van het gedachte-experiment

De scène is als volgt: in een afgesloten ruimte zit Freddy Vandemoortele, een postbode uit Aalst die geen enkele kennis heeft van de Chinese taal. Aan de ene kant van de kamer is er een doorgeefluik waardoor in het Chinees opgestelde boodschappen in de kamer belanden. Die boodschappen worden geschreven door mensen buiten de kamer. Het is Freddy’s taak om deze boodschappen, die hij dus absoluut niet begrijpt, te ‘beantwoorden’ met een passende (eveneens) Chinese boodschap. Maar hoe weet Freddy hoe hij de inkomende boodschappen moet beantwoorden? In de kamer bevindt zich ook een immens instructieboek, waarin uitgelegd staat, in het Aalsters, welke antwoorden bij welke boodschappen horen. Het boek staat dus vol met instructies van de vorm: ‘als je boodschap X binnenkrijgt, dan moet je antwoord Y geven’, waarbij Y eveneens een in het Chinees opgestelde boodschap is. Freddy schrijft het antwoord over op een briefje, en post het briefje door een tweede doorgeefluik aan de andere kant van de kamer.

Een nietsvermoedende Chinese bezoeker die Freddy een schriftelijke vraag stelt, en een passend antwoord krijgt, zal denken dat Freddy vloeiend Chinees spreekt. In werkelijkheid voert Freddy gewoon instructies uit – hij begrijpt geen jota van wat er op de briefjes staat, en weet zelfs niet dat de boodschappen opgesteld zijn in het Chinees. In het gedachte-experiment is Freddy functioneel identiek aan iemand die perfect Chinees spreekt. Waarom? Omdat hij vragen (input) op een juiste manier met antwoorden (output) verbindt. Ondanks deze functionele gelijkenis, zegt Searle, is er toch een duidelijk mentaal verschil. In tegenstelling tot de Chinees verstaat Freddy namelijk helemaal niets van de Chinese taal. Het gedachte-experiment levert met andere woorden een situatie op waarin twee mensen functioneel identiek zijn, en toch mentaal verschillend. En dat is precies een situatie die volgens het functionalisme onmogelijk is.

Het argument van de Chinese kamer is niet alleen een argument tegen het functionalisme, maar ook tegen één van de steunpilaren van het functionalisme, met name de Turing-test. Na enige training zou Freddy probleemloos de Turing-test doorstaan. Zijn antwoorden suggereren dat hij de inkomende boodschappen begrepen heeft, maar niets is minder waar. In werkelijkheid begrijp Freddy helemaal niet wat hij aan het doen is. Een gelijkaardig verhaal vind je in het boek van Dooremalen et al. ‘In 2007 werd er gewaarschuwd voor in Rusland ontwikkelde malware met de naam CyberLover. Dit programma zou in staat zijn om op grote schaal flirts aan te gaan met mannen achter de PC en zo persoonlijke gegevens te ontfutselen. Aangezien de heren niet doorhadden dat het hier geen vrouw, maar een programma betrof, lijkt aan de Turing-test voldaan. Zouden we nu werkelijk moeten zeggen dat CyberLover dezelfde mentale eigenschappen heeft als een aantrekkelijke vrouw?’

Derde en laatste argument: qualia

Leerde Mary ons iets bij? Volledige bespreking argument: zie paragraaf eliminativistisch fysicalisme

* + - 1. Superveniëntie-theorie

Even recapituleren. Functionalisme is een niet-reductionistische variant van fysicalisme. Wat betekent dat precies? Grijpen we even terug naar de computer-metafoor van het functionalisme. Software heeft hardware nodig om zich te ‘realiseren’. Een computerprogramma is volstrekt onnuttig wanneer er geen computer voorhanden is om het programma op te draaien. Daarnaast is het duidelijk dat de software van een computer niet gereduceerd kan worden tot de hardware. De abstracte beschrijvingen en instructies van een software-programma hebben silicium-chips nodig om zich te realiseren, maar ze vallen niet samen met de werking van die chips. Idem dito voor de relatie tussen mentale toestanden en de hersenen. Mentale toestanden hebben de hersenen nodig om zich te realiseren, maar dat betekent niet dat ze identiek zijn aan de hersenen. En ten slotte: idem dito voor de relatie tussen folk psychology (en bij uitbreiding: de niet-fysische wetenschappen) en fysica. Niet-fysische wetenschappen kunnen ware kennis opleveren, ook al zijn ze niet reduceerbaar tot de fysica.

Maar hoe kan dat? Het antwoord van de functionalist: wat we beschrijven en verklaren met behulp van de niet-fysische wetenschappen is niet iets dat niet beschreven en verklaard kan worden met behulp van de fysica. Er wordt eigenlijk niets nieuws gezegd; het wordt gewoon op een andere manier gezegd, meer bepaald op een abstracte manier. De beschrijving van ‘pijn’, bijvoorbeeld, is een abstracte beschrijving van hersentoestanden, die op hun beurt abstracte beschrijvingen zijn van chemische reacties, etc In deze optiek is alleen de beschrijving van fysische entiteiten, eigenschappen en wetten een concrete beschrijving. Anders gezegd: de categorieën van de niet-fysische wetenschappen komen misschien niet netjes overeen met de categorieën van de fysica, maar ze zijn evenmin totaal verschillend: de categorieën van de niet-fysische wetenschappen *worden gerealiseerd door* de categorieën van de fysica. Net zoals mentale toestanden gerealiseerd worden door hersentoestanden, en computerprogramma’s door allerlei hardware-onderdelen van de computer.

Er is ook nog een tweede antwoord op de vraag hoe we de relatie tussen geest en hersenen op een niet-reductionistische en toch fysicalistische manier kunnen denken. Tweede antwoord: superveniëntie-fysicalisme (*supervenience physicalism*). De relatie tussen functionalisme en superveniëntie-denken is erg onduidelijk. Sommigen beweren dat functionalisme gebaseerd is op superveniëntie-denken (of zulk denken impliceert), terwijl anderen beweren dat het om twee logisch onderscheiden theorieën gaat.

Wat bedoelen we met superveniëntie? Een superveniëntie-relatie is, net als de realisatie-relatie, een bepalingsrelatie. Superveniëntie-theorie bepaalt de relatie tussen het mentale en het fysische. Mentale eigenschappen, zo zegt men dan, superveniëren op fysische eigenschappen. De hamvraag is natuurlijk: wat betekent dat precies? Superveniëntie impliceert asymmetrie: mentale eigenschappen zijn enerzijds afhankelijk van fysische eigenschappen, maar anderzijds kunnen ze er niet toe gereduceerd worden. Je kan het ook zo formuleren: het is onmogelijk dat twee fysisch identieke exemplaren van een individu (twee klonen, zeg maar) toch verschillende mentale toestanden zouden hebben. Het omgekeerde kan wel: het kan best dat twee individuen dezelfde mentale toestanden hebben en toch niet fysisch identiek zijn. Ter illustratie: stel dat de hersenen van X zich in neurofysiologische toestand F bevinden op het moment dat X denkt dat water dorstlessend is. Wanneer de hersenen van Y zich nu eveneens in toestand F bevinden, kan het niet anders dan dat Y op dat moment ook denkt dat water dorstlessend is. Maar indien X en Y tegelijkertijd denken dat yoghurt gezond is, dan is het niet nodig dat hun hersenen zich in identiek dezelfde neurofysiologische toestand F\* bevinden.

Mentale toestanden zijn afhankelijk van fysische toestanden, zoals alle rechtgeaarde fysicalisten beweren. Dat wil zeggen dat mentale toestanden niet kunnen veranderen zonder verandering in de bijhorende fysische toestanden. Denk maar aan het argument van teletransportatie (het denkbaarheidsargument van de fysicalisten): de fysiek identieke kloon die op Mars ‘geprint’ wordt zal identiek dezelfde mentale toestanden hebben als het origineel.

Tegelijkertijd menen superveniëntie-voorstanders, in tegenstelling tot aanhangers van IT, dat mentale toestanden niet zomaar gereduceerd kunnen worden tot hersentoestanden.

Voorbeeldje van de Lego-brandweerauto (zie extra PDF bij deze les Niet ontvangen).

Superveniëntie theorie is compatibel (wilt niet zeggen: hetzelfde, maar overeenkomstig) met emergentisme, epifenomenalisme en hylomorfisme.

* + 1. **Tekst what mary didn’t know van Peter Jackson**

Achtergrond

Jackson maakt gebruik van het fenomeen van qualia om het fysicalisme te ondermijnen. De redenering is dat er aspecten zijn van ons mentale leven die zich niet laten inpassen in een fysicalistisch discours. Herinnert u zich dat er ruwweg drie soorten van mentale toestanden zijn: ervaringen, propositionele attitudes, en mentale handelingen. Qualia zijn subjectieve aspecten van ervaringen, meer bepaald van een deelverzameling van ervaringen: de zintuiglijke ervaringen. Jackson had het ook over een heel andere boeg kunnen gooien. Hij had zich bijvoorbeeld kunnen richten op specifieke aspecten van propositionele attitudes om het fysicalisme te ondermijnen. Een van de eigenaardigheden van PA is dat ze intentioneel zijn – ze zijn altijd ergens op gericht, op een ‘gedachte-inhoud’. Volgens Franz Brentano was intentionaliteit het watermerk (‘hallmark’) van het mentale. Hij formuleerde het fenomeen als volgt: ‘In de voorstelling wordt [altijd] iets voorgesteld, in het oordeel iets aanvaard of verworpen, in de liefde iets bemind, in de haat iets gehaat, in het begeren iets begeerd enzovoort. Deze intentionaliteit is uniek voor de psychische verschijnselen. Bij geen enkel fysisch verschijnsel doet zich iets dergelijks voor’ (geciteerd in D 42). Ruw samengevat: mijn verlangen naar een grote zak chips gaat over die zak chips, maar die zak chips zelf gaat helemaal nergens over. Een zak chips heeft geen intentionaliteit. Net als qualia lijkt intentionaliteit zich te onttrekken aan een fysicalistische woordenschat.

Het bestaan en de aard van qualia worden vaak ingeroepen als argumenten tegen het fysicalisme. En wat verstaat bijvoorbeeld Jackson onder ‘fysicalisme’? Hij definieert ‘fysicalisme’ in de eerste voetnoot van ‘What Mary didn’t know’: ‘If physicalism is true, then if you know everything expressed or expressible in explicitly physical language, you know everything’ (291). Mary is een super-wetenschapper die alles weet wat kan gezegd worden met behulp van de fysica, en toch lijkt ze iets bij te leren wanneer ze een appel ziet. Er was dus blijkbaar toch iets wat ze niet wist. En daarom faalt het fysicalisme. Maar welke soort van fysicalisme heeft Jackson eigenlijk voor ogen in zijn gedachte-experiment? In het laatste luik van de tekst gaat hij een debat aan met Churchland, de meest hevige aanhanger van het eliminativistische fysicalisme. De rest van de tekst lijkt in de eerste plaats te argumenteren tegen het reductionistisch fysicalisme. Waarom? Hij definieert fysicalisme als de stelling dat ‘volledige kennis van het fysische is volledige kennis tout court’ (‘complete physical knowledge is complete knowledge simpliciter’, 291). Niet-reductionistische fysicalisten zouden dan kunnen tegenwerpen dat hun opvattingen niet ondermijnd worden door het kennis-argument van Jackson. Zij beweren immers niet dat fysische kennis de enige vorm van kennis is. Niet-fysische wetenschappen leveren ook een vorm van kennis op die niet zomaar herleid kan worden tot fysische kennis. Maar Jackson lijkt het daar niet mee eens te zijn. Hij richt zijn pijlen in ieder geval ook op de functionalisten. Mary, zo zegt hij, weet niet alleen alles over hersentoestanden etc, maar ook over de functies die hersentoestanden vervullen, en de causale relaties tussen hersentoestanden (zie bijvoorbeeld p. 291, eerste paragraaf, en p. 292, laatste regels). Waarom beweert Jackson dat ook het functionalisme geen blijf weet met qualia? Qualia kan je niet analyseren met behulp van inputs en outputs, omdat qualia geen relationele fenomenen zijn (gaan geen relaties aan met andere mentale fenomenen), en dat het fenomenen zijn die je niet verder kan analyseren. Maar waarom zouden qualia niet-relationeel en niet-analyseerbaar zijn?

Het begrip qualia (eindelijk goed uitgelegd!)

Het is niet altijd duidelijk wat men bedoelt wanneer men over ‘qualia’ spreekt. Het begrip ‘quale’ (enkelvoud) of ‘qualia’ (meervoud) is een erg vaag begrip. Het zal u dan ook niet verwonderen dat er filosofen zijn die het spreken over qualia willen verbieden, omdat het een verward spreken is. Daniel Dennett, bijvoorbeeld, vindt het qualia-debat overroepen. Volgens hem bestaan qualia eenvoudigweg niet.

Brede betekenis: verwijzen naar subjectieve kwaliteit van ervaringen. Subjectieve kwaliteit of anders geformuleerd, fenomenale kwaliteit, hoe ze zich aan mij presenteren. Ervaringen zijn in deze zintuigelijk zoals emoties en gevoelens. Voorbeelden: Hoe het is om Jeuk te hebben, hoe het is om afgewezen te worden,…

Enge technische betekenis: mentale eigenschappen waar subject geprivilegieerde manier mee vertrouwd is.

Beide betekenissen overlappen elkaar. De conclusie is in beide gevallen dat een andere wetenschap nodig dan de fysische. Een nieuwe wetenschap die niet in derde persoon spreekt. Wanneer we spreken over hoe het voelt op gepitst te worden kunnen we dat niet anders dan in een subjectieve beschrijven.

Jackson geeft in zijn tekst aan dat Mary, die van alle wetenschappelijke kennis alles weet alsook weet wat de functies kunnen zijn van de hersenen, toch iets bij leert wanneer ze een appel ziet met de kleur rood.

De vraag waarom qualia zich ‘schijnen te onttrekken’ aan een fysische beschrijving, wordt meestal niet beantwoord. Een gangbare intuïtie onder filosofen van de geest, zeg maar. De uitdaging voor fysicalisten is dan om aan te tonen dat qualia toch kunnen beschreven en verklaard worden met behulp van fysische categorieën.

Een ander voorbeeld is van Thomas Nagel *what it is like to be a bat.*

Hoe is het eigenlijk om een vleermuis te zijn?

Vleermuizen nemen de wereld op een heel andere manier waar dan mensen. Mensen winnen informatie over de buitenwereld met behulp van hun gezichtsvermogen en hun gehoorvermogen, terwijl vleermuizen gebruik maken van een sonar-systeem. Ze stootten voortdurend klanken uit, geluidsgolven die tegen allerlei voorwerpen in de omgeving aanbotsen en weerkaatst worden – voorwerpen als huizen, lantaarnpalen, en natuurlijk ook roofdieren en prooidieren, zoals op de foto hierboven. Op basis van deze weerkaatste geluidsgolven reconstrueert de vleermuis een kaart van zijn omgeving. Het is duidelijk dat de manier waarop vleermuizen hun omgeving ervaren fundamenteel verschilt van de manier waarop wij onze omgeving ervaren. (Of niet? Er is een geval bekend van een blinde jongen die met zijn skateboard naar school gaat, en daarbij navigeert met behulp van een eenvoudig solar systeem. Hij gebruikt een ‘klikker’, en meer bepaald de weerkaatsing van de geluiden van die ‘klikker’ om obstakels te ontwijken. Zulke voorbeelden tonen aan dat de ervaring van vleermuizen misschien toch niet fundamenteel verschilt van de onze.)

U zou kunnen tegenwerpen dat de wetenschap ons alles vertelt wat we moeten weten over de ervaringen van vleermuizen: de hersenprocessen die met hun ervaringen gepaard gaan, de fysische eigenschappen van de geluidsgolven, enzovoort. Maar, zegt Nagel, de wetenschap leert ons alleen objectieve feiten over de ervaringen van vleermuizen. Ze neemt het perspectief in van een derde persoon – Nagel spreekt zelfs over ‘a God’s eye view’, het perspectief van een of andere godheid die het universum van op afstand gadeslaat – maar leert ons niets over het perspectief van de eerste persoon. De wetenschap leert ons niets over de subjectieve beleving van de vleermuis – hoe het is om vleermuis te zijn.

Overzicht tekst van Jackson

In zijn tekst bespreekt Jackson eerst het geval Mary en wat haar zo specifiek maakt. Daarna geeft hij zijn bekende kennis-argument en treed hij in debat met Churchland wie een bekend aanhanger is van eliminiativisch fysicalisme. In de laatste plaats geeft Jackson replieken op de argumenten die Churchland aanhaalt.

Het geval Mary

Mary, superwetenschapper in de verre toekomst. Fysica heeft op dat moment een eindtoestand bereikt: Mary weet werkelijk alles wat de wetenschappen te vertellen hebben over kleurwaarneming. Zelf is ze echter opgegroeid, en leeft ze nog steeds in een kleurloze omgeving. Haar huis is volledig ingericht met zwarte en witte meubels, haar kleren zijn zwart of wit, ze heeft zwart haar en een vaalwitte teint, en ze kijkt naar het nieuws op een oude zwart-wit televisie. Haar kennis ontleent ze aan boeken over de fysiologie, neurologie en biochemie van kleurwaarneming, en alle prentjes in die boeken zijn in zwart-wit.

Op een mooie dag opent Mary een geheime deur naar de buitenwereld, en het eerste wat ze ziet is een helrode appel (of tomaat, of aardbei, al naargelang de versie van het gedachte-experiment). De hamvraag is: leert Mary iets bij wanneer ze die appel ziet? Leert ze iets nieuws – iets dat ze niet al wist op basis van haar studie? Ze wist namelijk perfect wat er zich afspeelt in het oog, het zenuwstelsel en de hersenen wanneer iemand de kleur ‘rood’ ziet; ze wist alle fysische feiten over de kleur in kwestie. Maar ze had zelf nog nooit een rood voorwerp gezien. En als ze iets bijleert, wat leert ze dan precies? Het antwoord van Jackson: ze leert hoe het is om de kleur ‘rood’ te zien.

Het kennisargument

De redenering gaat als volgt.

Redenering anders geformuleerd: Als fysicalisme waar is, dan weet Mary alles. Mary weet niet alles dus fysicalisme is niet waar.

Hoe ziet Jackson het: Jackson meent dat Mary wel degelijk iets bijleert wanneer ze de appel ziet. Maar als ze werkelijk iets bijleert, dan was haar kennis toch niet volledig. Ze wist met andere woorden niet alles over het waarnemen van rood, terwijl ze wel op de hoogte was van alle fysische feiten terzake. Ergo: wat Mary bijleert gaat niet over de fysica van kleurwaarneming. Mary leert iets bij over een niet-fysisch feit: qualia.

Note: Jacksons kennisargument dat voortkomt uit zijn gedachte-experiment is de grond voor het epifenomenalisme. Zoals reeds gezien is dit een vorm van eigenschapsdualisme of de Dual attribute theories (DAT). De conclusie luidt als volgt: qualia zijn geen fysische fenomenen, maar zijn neveneffecten van fysische processen.

Problemen met het kennisargument

Er zijn verschillende manieren om het kennis-argument te ondermijnen.

Een eerste strategie is de minst gesofisticeerde: Mary leert helemaal niets bij. Nagenoeg alle formuleringen van het kennis-argument doen beroep op uitdrukkingen als ‘het lijkt erop dat’, ‘het heeft er alle schijn van dat (Mary iets bijleert)’. Aanhangers van het kennis-argument doen met andere woorden beroep op een intuïtie, en het eerste tegenargument stelt dat die intuïtie vroeg of laat verkeerd zal blijken te zijn. Waarom zou die toekomst-fysica niet in staat zijn om ook de kwalitatieve aspecten van een ervaring te beschrijven en te verklaren met behulp van een fysicalistische woordenschat? Aanhangers van het kennis-argument verdedigen zich tegen deze objectie met een waarschijnlijkheidsargument: het lijkt niet echt waarschijnlijk dat neurowetenschappers ons iets zullen kunnen vertellen over de subjectiviteit van ervaringen zolang ze onderzoek blijven doen over de objectieve aspecten van ervaringen.

Andere critici geven toe dat Mary iets bijleert. Maar wat leert ze precies bij? Jackson zegt: ze leert een nieuw feit, en dat kan geen fysisch feit zijn (want die feiten kende ze allemaal). Wat ze bijleert is een niet-fysisch feit.

Volgens Churchland leert Mary helemaal geen nieuw feit (een niet-fysisch feit) wanneer ze een tomaat ziet. Wat ze bijleert is daarentegen een nieuwe manier van kijken naar een feit dat ze al kende – een fysisch feit.

Volgens Lewis leert Mary helemaal geen nieuw feit wanneer ze een tomaat ziet. Wat ze bijleert is daarentegen een nieuw talent, een nieuw vermogen (ability). Ze leert het talent om rode dingen te herkennen.

* 1. **Eliminativistisch fysicalisme**

Inleiding

*‘Ordinary language embodies the metaphysics of our remote ancestors, and it is quite possible for it to change’ (Bertrand Russell)*

*‘We must no longer speak of feeling pain or being wracked by jealousy, but talk instead of C-fibres firing or of circuits in the triune cortex discharging with spiking frequencies of various orders of hertz’ (Keith Maslin)*

*“Paul, don’t speak to me, my serotonin levels have hit bottom, my brain is awash in glucosteroids, my blood vessels are full of adrenaline, and if it weren’t for my endogenous opiates I’d have driven the car in to a tree on the way home. My dopamine levels need lifting. Pour me a chardonnay and I’ll be down in a minute.” (What pat said to paul, quoted in the new Yorker, February 12, 2007)*

Het citaat van Russell: ‘Ons taalgebruik verwijst nog steeds naar de metafysica van onze verre voorouders, dus het is niet onwaarschijnlijk dat het nog zal veranderen’.

Het citaat van Maslin is exemplarisch voor een uitgebreide literatuur waarin het E fysicalisme genadeloos geridiculiseerd wordt. *No way* dat onze manier van denken en spreken over mentale toestanden – onze *folk psychology* – ooit plaats zal ruimen voor die koude taal van de fysica of de neurofysiologie.

Een ander voorbeeld is het laatste citaat. In een dubbelinterview in The New Yorker leggen de Churchlands uit hoe steeds grotere onderdelen van het dagelijks taalgebruik zullen vervangen worden door wetenschappelijke uitspraken. En voor alle duidelijkheid: het gaat niet alleen om een nieuwe manier van communiceren op de werkvloer tussen pakweg ingenieurs of computerwetenschappers of astrofysici – het gaat om een nieuwe manier van denken en spreken die zich uitstrekt tot in de meest intieme hoekjes en kantjes van ons bestaan. Inclusief de communicatie tussen geliefden. In casu Paul & Patricia Churchland.

Het citaat is voor alle duidelijkheid geen grap. Het hele interview vindt u op Toledo, en het citaat wordt voorafgegaan door de volgende passage: ‘Paul and Pat, realizing that the revolutionary neuroscience they dream of is still in its infancy, are nonetheless already preparing themselves for this future, making the appropriate adjustments in their everyday conversation. One afternoon recently, Paul says, he was home making dinner when Pat burst in the door, having come straight from a frustrating faculty meeting. She said’…

* + 1. **Aanvullingen op tekst what mary didn’t know**

Het kennis-argument van Frank Jackson heeft een enorme impact gehad in de filosofische gemeenschap, waar het debat over Mary nog steeds aan de gang is. Vanwaar die populariteit?

Achtergrond tekst zie deel over niet-reductionistisch fysicalisme. We zullen ons hier beperken tot enkele aanvullingen in verband met de voorbeelden die onder de paragraaf van Jacksons tekst gegeven zijn.

Een aanvulling op de vleermuis

Nagel zou wellicht toegeven dat Ben Underwood, de blinde jongen die beroep doet op een vorm van echolocatie, in zekere zin WEL weet hoe het is om een vleermuis te zijn, althans wat ruimtelijke waarneming betreft. Maar dat ondermijnt zijn argument niet. Het probleem wordt gewoon verplaatst: Nagels claim is dat de wetenschappen ons niet kunnen vertellen hoe het is om een vleermuis te zijn, maar ook niet hoe het is om Ben Underwood te zijn. Of beter: hoe het is om de ruimte waar te nemen zoals Ben Underwood die waarneemt. Het punt is dat er aan elke wetenschappelijke analyse van mentale toestanden iets ontsnapt, namelijk de subjectieve, kwalitatieve aspecten van zulke toestanden. En dat is ook de conclusie van een gelijkaardig gedachte-experiment – het argument van Mary.

Extra uitleg bij het kennis argument

Jacksons argument heeft de logische vorm van een modus tollens. Dat wil zeggen: een redenering waarin de conclusie volgt uit twee premissen. De eerste premisse is een voorwaardelijke uitspraak (‘als p dan q’), terwijl de tweede premisse de consequens (de tweede helft dus) van de eerste premisse ontkent (‘niet q’). De conclusie is onvermijdelijk: ‘niet p’.

a) Hernemen we even Jacksons definitie van fysicalisme, met enkele kleine aanpassingen: ‘’If physicalism is true then someone who knows everything about the physical knows everything simpliciter’ (291). ‘Simpliciter’ vertaalde ik eerder al als ‘tout court’. Er lijkt weinig af te dingen op deze uitspraak – het is een definitie van fysicalisme waar alle fysicalisten het eens mee zouden zijn.

b) De tweede premisse luidt: ‘het is niet het geval dat iemand die alles weet over het fysische ook werkelijk alles weet’. Deze uitspraak volgt uit het gedachte-experiment van Jackson. Jackson meent dat Mary wel degelijk iets bijleert wanneer ze de appel ziet. Maar als ze werkelijk iets bijleert, dan was haar kennis toch niet volledig. Ze wist met andere woorden niet alles over het waarnemen van rood, terwijl ze wel op de hoogte was van alle fysische feiten terzake.

c) De conclusie is onvermijdelijk. Jackson redeneert nog even verder om uit te komen bij zijn verdediging van het epifenomenalisme (een vorm van eigenschapsdualisme die we in de tweede les besproken hebben). Als Mary alle fysische feiten kent, en toch iets bijleert wanneer ze de kleur ‘rood’ waarneemt, dan is wat ze bijleert een niet-fysisch feit. Qualia zijn dus geen fysische fenomenen. Maar Jackson beseft ook dat het universum een causaal gesloten systeem is, en dus moet hij besluiten dat qualia epi-fenomenen zijn, dat wil zeggen ‘bij-producten’ van fysische processen, die zelf geen fysische effecten kunnen hebben.

Aanvulling bij de problemen van het kennisargument

Mary leert niets bij(zie supra)

Andere critici geven toe dat Mary iets bijleert. Maar wat leert ze precies bij? Jackson zegt: ze leert een nieuw feit, en dat kan geen fysisch feit zijn (want die feiten kende ze allemaal). Wat ze bijleert is een niet-fysisch feit.

Zie Jacksons tweede ‘opheldering’ (*clarification*, p. 292) en zijn antwoord op eerste tegenargument van Churchland (p. 293-4). De filosoof David Lewis argumenteert dat Mary helemaal geen nieuw *feit* leert wanneer ze een tomaat ziet. Wat ze bijleert is daarentegen een nieuw talent, een nieuw vermogen (*ability* of *skill*). Ze leert het talent om rode dingen te herkennen. In de epistemologie maakt men een onderscheid tussen verschillende soorten van kennis. Zo onderscheidt men bijvoorbeeld propositionele kennis van procedurele kennis. De eerste vorm van kennis is een kennis van feiten (*knowledge that*). De tweede vorm gaat meer over vaardigheden – weten hoe je iets moet aanpakken (*knowledge how*). De twee soorten kennis kunnen onafhankelijk van elkaar bestaan. Mijn vader weet bijvoorbeeld heel erg veel over muziektheorie en muziekgeschiedenis (*knowledge that*), maar kan geen instrument bespelen (*knowledge how*). Wat Mary leert, zegt Lewis, is geen propositionele kennis, maar procedurele kennis. Ze leert geen nieuwe feiten over de kleur ‘rood’, ze leert de kleur te herkennen. Ze leert zich voor te stellen hoe het is voor andere mensen om rood te zien. Jackson stemt gedeeltelijk in met deze kritiek in ‘What Mary didn’t know’: ‘Now it is certainly true that Mary will acquire abilities of various kinds after her release (…). But is it plausible that that is *all* she will acquire?’ Mary verwerft toch ook nieuwe propositionele kennis, of niet? Twijfels van Jackson aan eigen argument op p. 295 bovenaan…

Het is niet helemaal duidelijk of Churchland een gelijkaardig tegenargument verdedigt (zoals Jackson zelf opmerkt). Volgens Churchland leert Mary helemaal geen nieuw feit (een niet-fysisch feit) wanneer ze een tomaat ziet. Wat ze bijleert is daarentegen een nieuwe manier van kijken naar een feit dat ze al kende – een fysisch feit. Jacksons tweede ‘opheldering’ gaat in zekere zin in tegen deze objectie van Churchland. Wat bedoelt Jackson met ‘inten*s*ionaliteit van kennis’ (p. 292)? Denk even terug aan het onderscheid tussen betekenis en verwijzing (in de les over R fysicalisme en in de tekst van Smart). Twee uitspraken kunnen verschillen in betekenis, en toch dezelfde verwijzing hebben. ‘Tom Boonen’ en ‘de bom van Balen’ hebben duidelijk een andere betekenis, maar ze verwijzen wel naar dezelfde figuur. Zou het niet kunnen, zegt Churchland, dat Mary helemaal geen nieuwe feiten leert, maar wel een nieuwe manier om over eenzelfde feit te spreken – net zoals ‘avondster’ en ‘ochtendsterk verschillende manieren van spreken zijn over eenzelfde planeet, namelijk Venus? Churchlands herformulering van Jacksons argument (op p. 293) suggereert dat Jackson twee soorten kennis verwart. Ditmaal gaat het niet over het verschil tussen ‘knowledge that’ en ‘knowledge how’, maar wel over het verschil tussen ‘knowledge by description’ (vrij vertaald als ‘iets weten’) en ‘knowledge by acquaintance’ (‘vertrouwd zijn met iets’) – een onderscheid dat afkomstig is van Bertrand Russell. ‘Knowledge by description’ komt min of meer overeen met propositionele kennis. Maar wat betekent ‘knowledge by acquaintance’? Het is een ervaringsgebaseerde vorm van kennis – een kennis die je enkele kan opdoen door onmiddellijke interactie met iemand of iets. Onze eigen geest, bijvoorbeeld, kennen we door middel van ‘acquaintance’ – we zijn er immers op een directe manier mee vertrouwd. En dat is precies Churchlands punt: we kennen onze eigen mentale toestanden niet op dezelfde manier als we onze hersentoestanden kennen. Kennis van onze mentale toestanden is ‘knowledge by acquaintance’, terwijl kennis van onze hersentoestanden onvermijdelijk ‘knowledge by description’ is. Jacksons argument verwart beide vormen van kennis, en is dus ongeldig.

Zie Jacksons antwoord op tweede tegenargument van Churchland. Volgens Churchland bewijst het kennis-argument niet alleen dat het fysicalisme onwaar is, maar ook dat het dualisme onwaar is. Jacksons repliek luidt dat dit tegenargument niet relevant is. Jackson is niet geïnteresseerd in een verdediging van het dualisme. Hij is eerder, zoals ik al zei, een voorstander van epifenomenalisme, en epifenomenalisme is natuurlijk een variant van dualisme, maar van alle dualistische varianten is het wel de minst dualistische.

* + 1. **Wat is eliminativistisch fysicalisme**

Inleiding

E fysicalisme is een controversiële, en lichtjes marginale positie in het debat over geest en lichaam, maar kan niettemin op een aantal beroemde ‘aanhangers’ rekenen – ook al profileerden ze zich niet allemaal als E fysicalisten, en verschilden ze nog van mening over bepaalde details. Volgens de Amerikaanse filosoof Willard van Orman Quine, bijvoorbeeld, was het onvermijdelijk dat we in het dagelijks leven zullen blijven gebruik maken van concepten als ‘overtuiging’, ‘verlangen’ en ‘angst’, ook al beantwoordt er strikt genomen helemaal niets aan zulke concepten in de werkelijkheid. In het Grote Boek van de Natuur staat er niets over overtuigingen en verlangens, niets over angst en liefde. De Churchlands zijn wellicht de meest radicale voorstanders van E fysicalisme. In hun ogen moeten we er zo snel mogelijk werk van maken om folk psychology te bannen uit ons leven, om open te staan voor nieuwe categorieën die ons zullen aangeleverd worden door de neurowetenschappen. Tijd voor een radicale heroriëntatie van ons denken.

*‘Eliminative materialism is the thesis that our common-sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience’ (Churchland 1981, 214)*

Een eerste definitie

Zie vorig citaat van Churchland. Over welke theorie gaat het? Over folk psychology. En wat verstaat Churchland onder ‘folk psychology’? Churchlands definitie van folk psychology: ‘”Folk psychology”’, zegt hij, ‘verwijst naar het pre-wetenschappelijke, common-sense conceptueel raamwerk dat alle normale mensen gebruiken om het gedrag van mensen en bepaalde diersoorten te begrijpen, te voorspellen, te verklaren en te manipuleren. Het raamwerk bevat een groot aantal concepten, zoals de concepten ‘geloof’, ‘verlangen’, ‘pijn’, ‘lust’, ‘liefde’, ‘haat’, ‘vreugde’, ‘angst’, ‘wantrouwen’, ‘geheugen’, ‘herkenning’, ‘woede’, ‘sympathie’, ‘intentie’, enzovoort’ (geciteerd in W 86).

Een van de basisprincipes van de wetenschapsfilosofie is dat elke wetenschappelijke theorie falsifieerbaar is. Dat wil zeggen dat de theorie tegengesproken kan worden door nieuwe experimenten of observaties. Wanneer een theorie gefalsifieerd wordt, zijn er twee mogelijkheden. Soms kan de theorie gered worden mits een paar kleine aanpassingen. Maar soms is een theorie zo fundamenteel problematisch dat bijvijlen onbegonnen werk is. In dat geval is het onvermijdelijk dat de theorie verworpen wordt. Denk bijvoorbeeld aan de flogiston-theorie over ontvlambaarheid. Of aan de geocentrische kosmologie van Ptolemaeus en Aristoteles. Of aan de hekserij-verhalen uit het verleden. De centrale stelling van E fysicalism is dat folk psychology, net als de kosmologie van Aristoteles etc, een waardeloze wetenschappelijke theorie is, en dat deze theorie daarom vroeg of laat zal verworpen worden – geëlimineerd worden. Waarom is FP een waardeloze wetenschappelijke theorie? Om allerlei redenen, zoals we dadelijk zullen zien, maar de meest fundamentele reden is wellicht dat haar ontologie niet deugt. Een ontologie is de verzameling van entiteiten waarvan een theorie claimt dat ze bestaan. De ontologie van FP bevat onder andere propositionele attitudes – mentale toestanden zoals overtuigingen, verlangens, verwachtingen, intenties, enzovoort – en zulke mentale toestanden bestaan gewoon niet, zeggen eliminativisten.

E fysicalisme is een soort profetie: het is de voorspelling dat we vroeg of laat op een radicaal andere manier over onszelf zullen denken. Wanneer? Zodra de neurowetenschappen hun eindtoestand bereikt hebben. Waarom is E fysicalisme controversieel? Het eliminativisme is een controversiële positie in philosophy of mind omdat ze als enige theorie de legitimiteit van FP in twijfel trekt. Alle andere varianten van fysicalisme en dualisme zijn gebaseerd op de gedachte dat minstens een deel van FP compatibel is (of moet zijn) met de uiteindelijke theorie over de verhouding tussen lichaam en geest. Waarom is het controversieel om de legitimiteit van FP in twijfel te trekken? In de eerste plaats natuurlijk omdat FP ons nauw aan het hart ligt – het is een raamwerk dat ons dagelijks leven zin geeft en coherent maakt. Maar daarnaast stuit het eliminativisme ook op weerstand omdat FP in het algemeen ‘goed lijkt te werken’. Zoals eerder gezegd zijn we eigenlijk behoorlijk succesvol in het toekennen van mentale toestanden aan anderen, ook al hebben we geen rechtstreekse toegang tot hun geest. We zijn in staat om emoties te herkennen, zelfs overtuigingen en verlangens, en we kunnen in beperkte mate zelfs voorspellen wat iemand in een bepaalde situatie zal denken! Daarover dadelijk meer.

Intermezzo (wel kennen!)

De term ‘folk psychology’ verwijst naar het geheel van intuïties van gewone mensen over hun eigen gedrag en psychologie, en het gedrag en de psychologie van andere mensen. De term weerspiegelt gelijkaardig (en eerder) onderzoek naar hoe gewone mensen denken over fysische systemen (*folk physics*), over organismen (*folk biology*), en over psychiatrische stoornissen (*folk psychiatry*).

*Folk physics* wordt ook soms *naive physics* genoemd – een naiëve, voorwetenschappelijke kijk op de werking van fysische systemen die we op een spontane en natuurlijke manier verwerven terwijl we opgroeien. De naïeve en de wetenschappelijke versie van de fysica spreken elkaar vaak tegen. We zeggen bijvoorbeeld dat de zon opkomt en ondergaat, maar wat er in werkelijkheid gebeurt is dat de aarde rond de zon draait. Een ander voorbeeld: Galileo zou ooit een experimentje hebben uitgevoerd vanop de scheve toren van Pisa (het was wellicht eerder een gedachte-experiment). Zijn plan was om een zware kanonskogel van 300 kilogram en een knikker gelijktijdig te laten vallen vanop de top van de toren. Wat zegt uw intuïtie? De meeste mensen voorspellen dat de kanonskogel eerder de grond zal raken dan de knikker. Waarom? Omdat onze intuïties, onze *folk physics*, eerder overeenkomen met een Aristotelische mechanica dan met een moderne Newtoniaanse of Einsteiniaanse mechanica. Een van de basisprincipes van de hedendaagse fysica is immers dat de zwaartekracht alle objecten op dezelfde manier doet versnellen, ongeacht hun massa of hun samenstelling. Het lijkt er dus op dat Russell gelijk had in het citaat dat ik aanhaalde in het begin van de les: onze intuïties lijken soms wel atavismen – relicten uit de naïeve wereld van onze voorouders, in dit geval de antieke oudheid.

*Folk biology* verwijst naar onze common sense intuïties over hoe organismen (levende wezens) functioneren. Onderzoek wijst uit dat gewone mensen overal ter wereld de neiging hebben om een unieke en natuurlijke essentie toe te kennen aan elke soort (species) van planten en dieren – een fenomeen dat bekend staat als ‘psychologisch essentialisme’. Het lijkt erop dat onze intuïties over de levende natuur nog doordrenkt zijn van het typologische denken van Plato – een manier van denken die in de biologische wetenschappen al lang vervangen is door het populatiedenken van Darwin. (Het onderscheid tussen typologisch en populatiedenken is afkomstig van filosoof Ernst Mayr.) Volgens Plato gaan er een vast aantal onveranderlijke types of ideeën schuil onder de wilde veelheid van fenomenen in de werkelijkheid. Enkel deze ideeën zijn reëel; de rest zijn louter afschaduwingen van de ideeën. Darwins populatiedenken maakte definitief komaf met deze essentialistische denkwijze, door de radicale uniciteit van ieder levend wezen te benadrukken. Er is zoveel variatie in de levende wereld, zegt Darwin, dat geen twee organismen identiek zijn. Populaties zijn verzamelingen van individuen, en enkel deze individuen zijn reëel. Mayr concludeert: ‘the ultimate conclusions of the population thinker and of the typologist are precisely the opposite. For the typologist, the type (eidos) is real and the variation an illusion, while for the populationist the type (average) is an abstraction and only the variation is real. No two ways of looking at nature could be more different’ (Mayr 1975, 27). Hoe dan ook, onze intuïties sluiten eerder aan bij het typologische denken dan bij het populatiedenken.

Idem dito voor onze intuïties over de aard van geestesziekten (*folk psychiatry*). Nick Haslam heeft heel wat onderzoek gedaan naar zulke intuïties, en merkte tot zijn verrassing dat gewone mensen ook de neiging hebben om geestesziekten te essentialiseren, dat wil zeggen om een natuurlijke en unieke essentie toe te kennen aan alle psychiatrische patiënten met eenzelfde diagnose. In een experimentje liet hij proefpersonen een fictief artikel lezen over een vermeende doorbraak in het biologisch onderzoek naar een of andere geestesziekte (‘een gen voor stoornis X’). Daaruit bleek dat de manipulatie van een enkele (essentialistische) overtuiging vaak allerlei andere (essentialistische) overtuigingen activeerde: ‘Instead of passively assimilating information about the biological basis of mental disorders as an accretion of facts and opinions, the layperson’s mind is equipped to fit it into an implicit ontology that resonates deeply with the disease model [i.e. the natural kind view]. The mind is readied, that is, to construe mental disorders as natural kinds if this essentialist ontology is suitably triggered’ (Haslam 2000, 1045). Ook *folk psychiatry* vloekt met de huidige wetenschappelijke stand van zaken in de psychiatrie, in die zin dat een essentialistische benadering van geestesziekten radicaal achterhaald is (al moet ik er eerlijkheidshalve aan toevoegen dat er in de psychiatrie niet zo’n grote wetenschappelijke consensus bestaat over het anti-essentialisme als in de biologie).

Deze drie vormen van *folk knowledge* suggereren dat zulke kennis altijd minstens beperkt is, en soms zelfs ronduit verkeerd. Volgens E fysicalisten past *folk psychology* mooi in dit rijtje van achterhaalde theorieën, en net zoals fysica op geen enkele manier beroep doet op *folk physics*, of psychiatrie op *folk psychiatry*, zullen de neurowetenschappen van de toekomst (*completed neuroscience*) geen beroep doen op *folk psychology*.

Het failliet van Folk psychology

E fysicalisme is gebaseerd op de gedachte dat folk psychology een achterhaalde theorie is. Die claim is tweeledig. Enerzijds wordt er beweerd dat FP een theorie is. Anderzijds dat die theorie een achterhaald theorie is. First things first: waarom is FP een theorie? Omdat het de structuur heeft van een theorie. Er worden hypothesen geconstrueerd (vb. dat andere mensen ook een geest hebben), verklaringen gegeven, voorspellingen gedaan; er is een ontologie (met propositionele attitudes als elementaire bouwstenen); er zijn wetten (die wetmatigheden tussen bouwstenen duiden);… Welke wetmatigheden? Churchland spreekt van ‘common sense laws’, en geeft het volgende voorbeeld. Als iemand verlangt om benoemd te worden, en weet dat publiceren tot een benoeming leidt, dan zal die iemand publiceren. Of een beetje formeler: als A [attitude] [propositie] p, en weet dat x tot p leidt, dan zal A x doen. Wanneer we iemands gedrag willen begrijpen, dan kunnen we beroep doen op zulke ‘common sense wetten’.

Churchland vergelijkt ten slotte de (huidige) verhouding tussen FP en het fysicalisme, met de (historische) verhouding tussen alchemie en chemie, en tussen Aristotelische kosmologie (gebaseerd op het onderscheid tussen het ondermaanse en het bovenmaanse, waarbij het bovenmaanse aan heel andere wetten zou gehoorzamen, en uit een heel andere substantie zou bestaan dan het ondermaanse) en de klassieke mechanica: ‘One is reminded of how alchemy must have looked as elemental chemistry was taking form, how Aristotelean cosmology must have looked as classical mechanics was being articulated or how the vitalist conception of life must have looked as organic chemistry marched forward (222).

Waarom is FP een achterhaalde theorie? Churchland gebruikt een heel arsenaal aan adjectieven om FP te karakteriseren: ‘inadequate’, ‘false’, ‘confused’, ‘fundamentally defective’, ‘highly superficial’, ‘positively misleading’, ‘partial’, ‘a stagnant & degenerating research program’,… Om deze claim te evalueren moeten we naar Churchlands argumenten kijken.

Argumenten van Churchland tegen Folk psychology (omtrent de theoretische deugden)

Ten eerste is een goede theorie een theorie met een zekere **verklaringskracht**. Volgens Churchland faalt FP jammerlijk in dit opzicht: FP leert ons niets over, onder andere, het geheugen, over slaap (waarom brengen wij een derde van ons leven door in bed?), en over de aard van geestesziekten. MAAR: Churchland geeft weinig of geen uitleg bij deze kritiek. Concentreren we ons even op de kritiek dat FP geen verklaring biedt voor geestesziekten. Het eerste punt is: welk aspect van geestesziekten? Hun oorzaken? Hun verloop? Hun behandeling? De verschillen tussen ziekten onderling? Churchland laat ons een beetje in het ongewisse. Nemen we even aan dat Churchland bedoelt dat FP geen verklaring biedt voor het ontstaan van geestesziekten – hoe komt het dat sommige mensen bijvoorbeeld depressief worden? Een tweede objectie tegen Churchland is dat FP wel een antwoord biedt op deze vraag. FP is namelijk doordrongen van het psychoanalytische denken, en maakt gewag van verlangens die verdrongen worden en terugkeren als symptomen etc. Churchland kan riposteren dat zulke verklaringen psychoanalytische verklaringen zijn, en dat FP en psychoanalyse twee verschillende theorieën zijn, die elk op hun merites moeten worden beoordeeld. De psychoanalyse biedt misschien een verklaring voor het ontstaan van geestesziekten, maar FP niet, of enkel in die mate dat FP zich heeft laten beïnvloeden door de psychoanalyse. Maar zo’n antwoord is niet overtuigend: het punt is immers dat de psychoanalyse enkel zo’n immense invloed op FP heeft kunnen uitoefenen omdat de psychoanalyse gebouwd is op FP. Een derde bedenking bij Churchlands kritiek is dat zijn eisen met betrekking tot een goede theorie wel erg hoog zijn. Waarom zou een eenvoudige psychologische theorie als FP een veelheid van fenomenen moeten kunnen verklaren? Niet-reductionistische fysicalisten (zie vorige les) waren realistischer op dit punt: volgens hen komt de veelheid van theorieën over menselijk gedrag tegemoet aan onze veelheid van beschrijvende en verklarende interesses. Als FP al een theorie is, dan is het een theorie die ons toelaat om individuele keuzes, gevoelens, persoonlijkheden etc te beschrijven en te verklaren, eerder dan de fysiologische correlaten van zulke keuzes etc.

Ten tweede, zegt Churchland, is het eigen aan een goede theorie dat ze **vruchtbaar** is. Een vruchtbare theorie is een theorie die toelaat om nieuwe voorspellingen te doen, die nieuwe onderzoekspistes suggereert, en die op een of andere manier vooruitgang maakt – zich ontwikkelt, verbetert. Ook daar wringt een schoentje in het geval van FP: ‘The FP of the Greeks is essentially the FP we use today, and we are negligibly better at explaining human behavior in its terms than was Sophocles’ (Churchland 1981, 221). Churchlands claim is nogal radicaal. Er lijken een aantal prima facie tegenargumenten te bestaan tegen zijn veronderstelling dat FP niet veranderd is sinds de Antieke Oudheid. We zagen zoëven dat de psychoanalyse in de twintigste eeuw de FP minstens aangevuld heeft, en misschien zelfs grondig veranderd. In de zestiende eeuw deed niemand beroep op verklaringen in de trant van ‘hij heeft die herinnering verdrongen, en daarom is hij ziek geworden’. Zulke verklaringen, die duidelijk deel uitmaken van onze FP, zijn nieuw, en dus houdt Churchlands claim geen steek.

Ten derde en ten laatste is een goede theorie een **consistente** theorie – men spreekt dan van consilience (een toestand waarbij meerdere theorieën elkaar ondersteunen en versterken). Een goede theorie is een theorie die niet alleen staat, maar op een vruchtbare manier aansluiting vindt bij andere (succesvolle) theorieën. Is FP een consistente theorie? Neen, zegt Churchland. Evolutiebiologie, neurowetenschappen, fysiology, organische chemie etc zijn vormen een coherent geheel, die ten minste de indruk geven dat ze vroeg of laat zullen versmelten in een grootse theoretische synthese. In haar huidige vorm zal FP echter nooit deel kunnen uitmaken van deze synthese. Waarom niet? Omdat ze gebruik maakt van categorieën die a priori niet te verzoenen zijn met specifieke ‘zekerheden’ uit de synthese. Een voorbeeld: intentionaliteit. De FP claimt dat propositionele attitudes altijd ergens over gaan, ergens op gericht zijn, en dat geldt hoegenaamd niet voor fysische toestanden. Mijn verlangen naar een zak chips is gericht op de zak chips, maar de zak chips zelf is helemaal nergens op gericht. Hersentoestanden zijn nergens op gericht, dus lijkt het onmogelijk om mentale toestanden te reduceren tot hersentoestanden. En gegeven de twee voorgaande argumenten tegen FP, is er volgens Churchland maar één oplossing: FP elimineren. MAAR: is het echt zo dat FP op geen enkele manier aansluit bij wetenschappen die enig succes geboekt hebben bij het beschrijven en verklaren van menselijk gedrag? Die claim lijkt opnieuw te radicaal. Neem bijvoorbeeld de evolutiepsychologie, waar we later nog een hele les aan zullen wijden. Er is een hypothese in de evolutiepsychologie die bekend staat als de *social brain hypothesis*. De hypothese luidt dat de complexiteit van ons brein ten dele te wijten is aan het feit dat wij, net als onze voorouders, deel uitmaken van grote sociale groepen, waarin het cruciaal is om over een aantal cognitieve sociale vaardigheden te beschikken (*social cognition).* Voorbeelden van zulke vaardigheden zijn: herkenning van gezichten, regulering van (eigen) emoties, herkennen van emoties bij anderen, en last but not least: ‘a theory of mind’. ‘Theory of mind’ is een technische term uit de evolutiepsychologie en de psychiatrie die verwijst naar wat wij verstaan onder ‘folk psychology’. De term ‘Machiavellistische intelligentie’ is een term die min of meer dezelfde inhoud dekt. Al deze termen verwijzen naar het feit dat we er in het dagelijks leven goed in slagen om elkaars gedrag te duiden door elkaar allerlei mentale toestanden toe te schrijven. (Extreme gevallen van autisme zijn in deze zienswijze het gevolg van een disfunctie in ‘theory of mind’.) Kortom: FP maakt deel uit van de sociale cognitie van onze voorouders, en die sociale cognitie stelde hen in staat om te overleven en zich voort te planten, en uiteindelijk het wereldtoneel te domineren. Als de *social brain hypothesis* steekt houdt, dan is FP wel compatibel met andere wetenschappen, zoals de evolutiepsychologie. Of de evolutiepsychologie zelf een deugdelijke wetenschappelijke theorie is, is voer voor een latere les

* + 1. **Argumenten tegen eliminativistisch fysicalisme**

We zullen hier drie argumenten poneren die ingaan tegen de eliminativistisch positie van fysicalisme.

Eerste argumenten: E fysicalisme spreekt zichzelf tegen

Een eerste argument tegen E fysicalisme is dat het een theorie is die zichzelf tegenspreekt. In het Engels spreekt men van een ‘self-refuting theory’. Waarom is E fysicalisme self-refuting?

Omdat E fysicalisten enerzijds beweren dat er geen mentale toestanden bestaan. U herinnert zich dat er verschillende soorten van mentale toestanden zijn: ervaringen, propositionele attitudes, en handelingen. ‘Geloven’ en ‘overtuigd zijn’ zijn voorbeelden van propositionele attitudes. E fysicalisten beweren met andere woorden dat ‘geloven’ en ‘overtuigd zijn’ zinloze categorieën zijn – ze bestaan immers niet. Anderzijds, zo gaat het tegenargument, *geloven* E fysicalisten dat hun theorie een betere theorie is dan het R fysicalisme of het niet-R fysicalisme. Ze zijn ervan *overtuigd* dat mentale toestanden niet bestaan. E fysicalisten beweren dus dat overtuigingen bestaan en niet bestaan. En dus spreken ze zichzelf tegen.

E fysicalisten kunnen deze objectie gemakkelijk pareren door te stellen dat ze niet geloven (noch niet geloven) in hun eigen theorie. Het tegenargument is wel interessant in die zin dat het aantoont hoe pervasief FP is, en hoe moeilijk het is om te ontkomen aan dat psychologiserend discours.

Tweede argument: FP is geen wetenschap

FP laat ons niet alleen toe om onze verlangens en overtuigingen te beschrijven, te verklaren en te voorspellen, maar bevat ook een normatief ideaal over hoe rationele wezens zouden moeten omgaan met zulke verlangens en overtuigingen. Volgens Dennett gaan we er bij het beschrijven, verklaren en voorspellen van het gedrag van anderen vanuit dat die anderen rationele wezens zijn – de uitdrukking ‘the intentional stance’ (de intentionele houding/instelling) verwijst precies naar deze veronderstelling.

Hoe dan ook, omdat FP niet alleen beroep doet op declaratieve of descriptieve uitspraken (uitspraken over feiten), maar ook op normatieve uitspraken (uitspraken over waarden en normen), verschilt ze, als theorie, van de meeste wetenschappelijke theorieën.

Derde argument: Het succes van FP

Wanneer ik wil voorspellen wat iemand zal doen in bepaalde omstandigheden, dan kan ik beroep doen op mijn kennis over wat hij of zij gelooft of verlangt, wat hij of zij graag heeft, of net een hekel aan heeft. Idem dito wanneer ik wil verklaren of beschrijven wat iemand doet of gedaan heeft. Zoals ik in een eerdere les al opmerkte zijn zulke voorspellingen, verklaringen en beschrijvingen niet altijd accuraat, en soms zijn ze zelfs totaal verkeerd. Maar denkt u even na wat uw buikgevoel (*gut reaction*) u vertelt wanneer u de bal mis slaat, dat wil zeggen wanneer u voorspelt dat iemand iets zal doen, terwijl hij of zij het niet doet. Of wanneer u er niet in slaagt om iemands gedrag te verklaren met de kennis die u hebt over die persoon. Uw eerste reactie is niet: die persoon heeft helemaal geen overtuigingen of verlangens (de keuze van de eliminativist). Maar wel: ik moet dringend even praten met die persoon, want ik weet duidelijk niet wat hij of zij denkt of verlangt, en hoe hij of zij zich voelt. Wanneer FP niet werkt denken we er met andere woorden niet aan om FP meteen overboord te gooien, maar wel om onze kennis over de mentale toestanden van de ander bij te spijkeren.

Moraal van het verhaal: om te begrijpen hoe het komt dat FP zo succesvol is, moeten we aannemen dat er toch iets beantwoordt aan de concepten en categorieën van FP. Om te begrijpen waarom FP zo succesvol is, moeten we aannemen dat mentale toestanden WEL bestaan.

1. Persoonsidentiteit

Op basis waarvan zeggen we dat persoon A identiek is aan persoon B? Anders geformuleerd: welke redenen hebben we om aan te nemen dat persoon A identiek is aan persoon B, gegeven het feit dat er duidelijke, soms zelfs drastische verschillen bestaan tussen hen? En met verschillen bedoel ik zowel lichamelijke als geestelijke verschillen.

Tekst die gelezen werd bij dit hoofdstuk is die van Anton Quinton, namelijk The soul.

* 1. **Het probleem**

Achtergrond

De hamvraag is: wat maakt iemand tot dezelfde persoon op verschillende tijdstippen? Waarin bestaat de identiteit van een persoon? Een antwoord op deze vraag bestaat uit een lijst van voldoende en noodzakelijke voorwaarden waaraan een persoon moet voldoen om dezelfde persoon te zijn op twee verschillende tijdstippen. Voor vele mensen is het debat over persoonsidentiteit niet louter een academisch tijdverdrijf. Het is een debat over hun toekomst. Sinds jaar en dag prediken tal van religies dat er leven is na de dood, dat er zoiets bestaat als ‘het hiernamaals’, dat er mogelijk een verrijzenis komt van alle doden, enzovoort. Maar wie is het dan die zal verrijzen? Wie overleeft er de dood? Wie zal er rondhangen in het hiernamaals? Gelovige mensen verlangen ernaar om als dezelfde persoon voort te bestaan, dat wil zeggen: ze verlangen ernaar hun identiteit te behouden over de dood heen. Ze nemen geen genoegen met het *iemands* voortbestaan – ze willen zelf eeuwig leven. Ook in de context van liefde en seksualiteit speelt het concept van persoonsidentiteit soms een belangrijke rol. Beeld u zich even in dat u getrouwd bent met de ene helft van een identieke tweeling. Uw schoonzus is met andere woorden een perfecte kopie van uw vrouw (of uw schoonbroer is een perfecte kopie van uw man). Stel dat uw vrouw/man een bijzonder gevoel voor humor heeft (of gewoon hoofdpijn), en zonder uw medeweten haar/zijn zus/broer inroept om ‘s avonds met u de liefde te bedrijven. De kans is groot dat u zich diep gekwetst zal voelen. Waarom? Omdat u niet geïnteresseerd bent in een replica van uw geliefde, maar enkel in *the real stuff*. Uw schoonbroer of –zus mag dan al quasi-identiek zijn aan uw geliefde, maar u wil enkel het bed delen met een en dezelfde persoon – uw geliefde.

Het concept ‘identiteit’ heeft verschillende betekenissen. Enerzijds is er kwalitatieve identiteit. Twee personen zijn kwalitatief identiek wanneer ze een perfecte kopie zijn van elkaar. Denk aan identieke tweelingen, bijvoorbeeld. Anderzijds is er ook zoiets als numerieke identiteit. Identieke tweelingen zijn niet numeriek identiek, want hoezeer ze ook op elkaar lijken, toch zijn het twee verschillende personen. Omgekeerd ben ik niet kwalitatief identiek aan de kleine jongen die ik was, maar wel numeriek identiek. De kleine jongen en ik zijn namelijk dezelfde persoon. Een ander voorbeeld van numerieke identiteit: Iocaste is numeriek identiek aan Oedipus’ moeder, ook al dacht Oedipus zelf dat ‘ie met numeriek verschillende personen te maken had. Het analytische debat over persoonsidentiteit gaat over numerieke identiteit, en niet over kwalitatieve identiteit.

De vraag of twee personen, zoals ik en de kleine jongen op de foto, numeriek identiek zijn , is niet louter een epistemologisch probleem. Het is niet louter de vraag hoe we *weten* of twee personen identiek zijn. Neem het voorbeeld van de geadopteerde Amerikaanse skieër Toby Dawson. Een man uit Zuid-Korea beweerde zijn zoon te herkennen toen deze laatste een bronzen medaille won op de winterspelen van Turijn in 2006. Toby was inderdaad geadopteerd door een Amerikaans koppel nadat hij als peuter zijn moeder was kwijtgespeeld op een drukke markt in Zuid-Korea (zie slide verderop). Hoe weten we of Toby Dawson en de verloren zoon dezelfde persoon zijn? Er zijn allerlei niet-wetenschappelijke en wetenschappelijke technieken om hun identiteit te bewijzen, zoals medische gegevens, vingerafdrukken of een DNA-test. Zelfs de meest gesofisticeerde technieken zijn echter feilbaar, dus het zou kunnen dat zowel Toby als zijn vader zich vergissen wanneer ze claimen dat Toby en de verloren zoon dezelfde persoon zijn. Zelfs de meest gesofisticeerde technieken om de numerieke identiteit van twee personen te achterhalen zeggen ons dus niets over wat de essentie is van de identiteit van een persoon. Persoonsidentiteit is met andere woorden niet louter een epistemologische kwestie, maar ook een metafysische kwestie. Zoals Maslin opmerkt: ‘The fallible evidence to which we appeal in trying to settle issues of personal identity does not tell us what is logically constitutive of the identity of persons over time. In answering that question, we need to spell out the necessary and sufficient conditions for a person at T1 to be one and the same as a person at T0. In other words, we need to complete the formula: “Person X at T1 is one and the same person as person X at T0 *if, and only if*,…’ (Maslin 244).

Kort samengevat: epistemologisch: hoe weten we dat 2 personen identiek zijn.

Metafysisch: wat maakt 2 personen identiek met andere welke zijn de voldoende en noodzakelijke voorwaarden om te spreken van 2 identieke personen.

Twee soorten theorieën over persoonsidentiteit

Er zijn twee verschillende soorten van theorieën over wat persoonsidentiteit is, en elke soort van theorie kent een aantal varianten.

Enerzijds zijn er *essentialistische* theorieën over persoonsidentiteit. Zulke theorieën, die vaak aansluiten bij een dualistische zienswijze, verdedigen de stelling dat er zoiets bestaat als een immateriële essentie van een persoon – een essentie die ongevoelig is voor de toevalligheden en wisselvalligheden van het dagelijks leven, en die alle elementen bevat die een persoon tot *deze* persoon maken. In de antieke oudheid, in het christendom en in het Cartesiaanse denken is deze immateriële subtantie uiteraard de ziel. Quinton spreekt van ‘the [spiritual] substance theory of personal identity’ (393). Zo’n ziel kan probleemloos ontkoppeld worden van het lichaam, zoals we gezien hebben in eerdere lessen, en een lang en gelukkig lichaamloos bestaan genieten. Essentialistische theorieën over PI zijn niet alleen dualistisch, maar ook niet-reductionistisch. Er zijn filosofen, zoals Richard Swinburne, die beweren dat het concept van persoonsidentiteit niet verder analyseerbaar is; dat het niet kan gereduceerd worden tot andere, eenvoudiger concepten; dat het een ‘bruut feit’ is, zoals een axioma in de wiskunde. Een laatste kenmerk van essentialistische theorieën over PI is dat ze de identiteit van een persoon beschouwen als een kwestie van alles of niets. Ofwel gaat het om dezelfde persoon, ofwel om een andere persoon. Er is geen grijze zone, geen tussenweg. Men kan niet een beetje identiek zijn.

Deze theorie is echter om allerlei redenen problematisch. Ik bespreek twee redenen. Om te beginnen (en dit is de eerste ‘weakness’ die Quinton bespreekt op p. 393-4) is het zo dat essentialistische theorieën het probleem verschuiven. De spirituele substantie of immateriële essentie bepalen de identiteit van een persoon, maar wie of wat bepaalt de identiteit van de betreffende substantie? Ten tweede is de spirituele substantie (of het Ego, zoals het in deze context ook vaak genoemd wordt) principieel onobserveerbaar, en daarom onbruikbaar als criterium om een persoon te identificeren (zie Quintons derde ‘weakness’, p. 394-5). In Descartes’ zienswijze hebben we bijgevolg geen zicht op wat de identiteit van een persoon bepaalt. Het zou daarom onmogelijk zijn om te weten of iemand op twee verschillende tijdstippen dezelfde persoon is. Het zou best kunnen dat ieder levend wezen bezield wordt door 274 verschillende personen. Bovendien laat een essentialistische theorie over PI niet toe om personen van elkaar te onderscheiden. Zulke claims vloeken echter met onze intuïtie, die ons influistert dat we personen wel van elkaar kunnen onderscheiden, en dat elk menselijk wezen een enkele persoon huisvest. Essentialistische theorieën over PI komen in deze les verder NIET aan bod.

Anderzijds bestaan er ook *conventionalistische* (of empirische, zoals Quinton ze noemt) theorieën over PI (waaraan deze les dus gewijd is). Zulke theorieën zijn in de regel –verrassing!- fysicalistisch, reductionistisch en gradualistisch. Beginnen we bij het reductionistische aspect. Sommige filosofen beweren dat het concept van persoonsidentiteit probleemloos geanalyseerd kan worden met behulp van andere concepten – concepten die op geen enkele manier gebruik maken van het concept van persoonsidentiteit. De meeste varianten van deze reductionistische soort van theorieën over PI maken gebruik van het concept van continuïteit. Twee personen zijn numeriek identiek dan en slechts dan als er sprake is van een continuïteit tussen beide personen. Er is iets dat hen bindt – iets dat een virtuele lijn trekt tussen beide, of een rode draad vormt in hun leven. De meningen verschillen over waarin die rode draad precies bestaat. Enerzijds zijn er filosofen die een groot belang toekennen aan fysische continuïteit – de continuïteit van het lichaam, of van een of ander onderdeel van het lichaam, meestal het brein. Twee personen zijn numeriek identiek omdat ze hetzelfde lichaam (of lichaamsdeel) delen. Anderzijds zijn er filosofen die de continuïteit van het lichaam ondergeschikt achten aan de continuïteit van het mentale leven. In deze zienswijze zijn twee personen identiek omdat ze dezelfde herinneringen, verlangens, overtuigingen, en verwachtingen delen, niet omdat ze hetzelfde brein hebben.

Waarom Convetionalistisch.

Quinton noemt deze theorieën ook empirische theorieën. In de meeste gevallen zijn ze fysicalistisch, reductionistisch en gradueel zoals we al zagen.

In onderstaande tekst zullen we uitleggen waarom ze gradueel zijn en conventioneel genoemd worden.

De twee reductionistische theorieën over persoonsidentiteit worden in de literatuur vaak ‘conventionalistische’ theorieën genoemd. Wat bedoelt men daarmee? Continuïteit doet zich onvermijdelijk voor in gradaties. Het is duidelijk dat er bijvoorbeeld meer continuïteit bestaat tussen mijn huidige lichaam en het lichaam dat ik had als 18-jarige student, dan tussen mijn huidige lichaam en het lichaam dat ik zal hebben als 80-jarige grijsaard. Ook psychologische continuïteit is een graduele kwestie. De mentale toestanden die mijn geest vandaag bevolken haken mooi in op de mentale toestanden die ik gisteren had, maar veel minder met de mentale toestanden die ik had als peuter. Als (fysische of psychologische) continuïteit inderdaad gradueel is, en als continuïteit inderdaad bepalend is voor PI, dan is PI ook gradueel. Om het met een cliché te zeggen: het is geen verhaal van of-of (ofwel is het dezelfde persoon, ofwel is het een andere persoon), maar een verhaal van en-en (het is gedeeltelijk dezelfde persoon, en gedeeltelijk een andere persoon).

Personen zijn in deze zienswijze vergelijkbaar met objecten (en nu kom ik tot een uitleg over waarom deze zienswijze ‘conventionalistisch’ genoemd wordt – zie ook Quinton p. 398). Met ‘objecten’ verwijs ik dan naar fenomenen als een land, een instituut, een machine, enzovoort. Neem het voorbeeld van het Hoger Instituut voor Wijsbegeerte. Het HIW bestaat onafgebroken sinds 1889 – er is dus sprake van enige continuïteit. Bestaat er zoiets als de identiteit van het HIW? Alle studenten en docenten die toentertijd rondwaarden in deze gebouwen zijn verdwenen, en het Thomistische gedachtegoed is niet meer zo nadrukkelijk aanwezig als in de begindagen (toen je nog een ‘Baccalaureaat in de Thomistische Wijsbegeerte kon behalen – een diploma waar Jean-Luc Dehaene graag mee uitpakt). Beeld je nu even in dat ook de filosofen vroeg of laat moeten verkassen naar Gasthuisberg, en dat ze zich daar uitsluitend zouden toeleggen op filosofische kwesties in de kernfysica. Zou het HIW ook zulke ingrijpende veranderingen overleven? Uw eerste intuïtie zegt u wellicht dat die vraag moeilijk te beantwoorden is met ‘ja’ of ‘neen’. In ieder geval ligt het voor de hand te zeggen dat het HIW van vandaag misschien *identieker* is aan het HIW in zijn begindagen dan aan het HIW in de toekomst. Of neem een ander voorbeeld van een object: een auto. Ik heb bijvoorbeeld jarenlang rondgereden met een oude Saab 900. Toen ik de auto wilde verkopen maakte ik een lijst met alle onderdelen die in de loop der jaren vervangen waren – een indrukwekkende lijst, meteen ook de reden waarom ik de auto van de hand wilde doen. Er zaten nauwelijks nog originele onderdelen in. Ondanks deze ingrijpende veranderingen beschouwde ikzelf (en andere mensen die me zagen rondrijden) mijn Saab als *dezelfde* auto als diegene die in de Saab-fabriek van de band rolde in januari 1991. Zulk spreken is nuchter bekeken onwaar (de Saab die ik verkocht was helemaal niet identiek aan de Saab die de vorige eigenaar in de winkel kocht), maar het is wel handig. Het toeschrijven van een identiteit aan objecten is conventioneel, in die zin dat we uit gemakzucht over ‘eenzelfde object’ spreken, ook al is het al talloze keren ingrijpend veranderd. Je zou zelfs een stap verder kunnen gaan en stellen dat de identiteit van objecten in zekere zin een fictie. En als persoonsidentiteit inderdaad analoog is aan objectidentiteit, dan is PI ook een fictie. In het dagelijks leven spreken we voortdurend over ‘dezelfde persoon’, ook al weten we eigenlijk wel dat het niet echt identieke dezelfde persoon is.

De common sense theorie

De slotsom van conventionalistische theorieën is dat persoonsidentiteit eigenlijk een louter theoretische constructie is. Het is geen onveranderlijke essentie die losgekoppeld kan worden van de psychologie en het lichaam van het individu. Deze conclusie vloekt echter met het *common sense*-denken over PI. In het dagelijks leven lijken we vaak beroep te doen op een essentialistische invulling van het concept van persoonsidentiteit. Een voorbeeld: het berechten van oorlogsmisdadigers. Neem het geval van John Demjanjuk – de Oekraïens-Amerikaanse oorlogsmisdadiger die terecht stond in Duitsland voor zijn vermeende bijdrage aan het vermoorden van 27900 Joden in het vernietigingskamp Sobibor, tijdens de tweede wereldoorlog. De zaak-Demjanjuk is in de eerste plaats een mooie illustratie van de epistemologische problemen rondom het concept van persoonsidentiteit. Hoe weten we dat de ziekelijke oude man die in 2009 berecht werd voor oorlogsmisdaden dezelfde persoon was als de beruchte ‘Ivan De Verschrikkelijke’, de beul van Sobibor? Zijn eerste veroordeling als oorlogsmisdadiger, in Israel in 1988, werd later herroepen omwille van vervalst bewijsmateriaal. Daarnaast zijn er ook interessante metafysische problemen in de zaak-Demjanjuk. Wat bedoelen tegenstanders van Demjanjuk eigenlijk wanneer ze beweren dat de ziekelijke oude John Demjanjuk dezelfde persoon is als de kampmedewerker Ivan Mykolaiovych Demianiuk (zoals hij toen officieel heette)? Het feit dat de man meer dan een halve eeuw later nog steeds aansprakelijk gesteld wordt voor gruweldaden uit een ver verleden, suggereert dat zijn tegenstanders een strikte identiteit zien tussen beide personen. De vraag of Demjanjuk dezelfde persoon is als de kampmedewerker in Sobibor kan enkel beantwoord worden met ‘ja’ of ‘neen’ – het gezond verstand laat geen tussenvormen toe. Het doet er niet toe dat Demjanjuk in de tussentijd een ziekelijke oude man geworden is, die misschien aan een vergevorderde vorm van geheugenverlies lijdt. Het doet er niet toe dat er weinig of geen fysische of psychologische continuïteit is tussen beide personen – de persoon Demjanjuk die terecht staat voor oorlogsmisdaden staat los van diens lichaam en psychologie, en precies daarom is hij nog steeds verantwoordelijk voor zijn toenmalige gruweldaden. Zoals Madell zegt: ‘our ordinary view of human rights and responsibilities, indeed a whole range of our ordinary moral attitudes, is based on a conception of personal identity totally at variance with the conventionalist view. If identity is a “fiction”, then I cannot suppose that I should feel any shame for my wrongdoing, or regard myself as responsible for it’. De common sense-opvatting over PI sluit met andere woorden aan bij het essentialistische denken over PI.

Toen Demjanjuk een eerste keer vrijgesproken werd in 1993 door een Israëlische rechtbank, verantwoordden de rechters hun oordeel als volgt: ‘*Ivan Demjanjuk has been acquitted by us, because of doubt, of the terrible charges attributed to Ivan the Terrible of Treblinka. This was the proper course for judges who cannot examine the heart and mind, but have only what their eyes see and read’. ‘*’The heart and mind’ van Demjanjuk is met andere woorden niet toegankelijk voor de rechters.

Haalt u even de metafoor van het levensverhaal voor ogen – de metafoor die ik daarnet gebruikte om het concept van psychologische continuïteit te illustreren. Literatuurcritici spreken wel eens over ‘een strak gecomponeerde roman’. Een strak gecomponeerd levensverhaal is een verhaal waarvan de verschillende delen op een complexe manier samenhangen; een verhaal waarin motieven en thema’s de continuïteit voorzien, net als in een klassiek muziekwerk. Ook popmuziek waarderen we vaak omwille van de continuïteit van een bepaalde song – een refrein dat steeds opnieuw, soms in licht gewijzigde vorm, herhaald wordt; een pompende baslijn die plots terugkeert; enzovoort. Sommige romans en muziekstukken vinden we minder goed dan andere, en dat verschil in waardering heeft, denk ik, vaak te maken met hoeveel continuïteit ze bieden. Punt is: net zoals een roman in mindere of meerdere mate ‘gecomponeerd’ kan zijn, zo kunnen twee personen in mindere of meerdere mate identiek zijn.

* 1. **PI als fysische continuïteit**

Basisclaims

Wat maakt iemand tot eenzelfde persoon op verschillende tijdstippen? Een eerste conventionalistische theorie, die intuïtief erg aantrekkelijk is, luidt dat persoonsidentiteit niet meer en niet minder is dan de continuïteit van ons lichaam doorheen de tijd. Men spreekt ook wel eens van de ‘flesh and blood theory of PI’. Uiteindelijk, zeggen aanhangers, kan niemand ontkennen dat we ons (en andere mensen) in grote mate identificeren met ons lichaam. Wij zijn ons lichaam. We zeggen bijvoorbeeld niet: ik ben blij dat hij enkel mijn arm aanraakte, en niet mezelf (mijn persoon).

Twee personen zijn identiek wanneer je, op basis van eenvoudig observatiewerk of eender welke andere techniek, een *onafgebroken* virtuele lijn kan trekken doorheen de tijd en de ruimte, tussen hun beider lichamen. De basale bouwstenen van mijn lichaam bestaan sinds 14 oktober 1979 (en eigenlijk al negen maanden eerder), en sindsdien heeft mijn lichaam op elk moment een plaats ingenomen in de ruimte. Mijn lichaam slaat als het ware een brug tussen de persoon die ik nu ben en de kleine jongen op de foto. Voor alle duidelijkheid: het lichaam van die kleine jongen is helemaal niet identiek aan mijn lichaam vandaag, maar dat wordt ook niet beweerd. In het debat over persoonsidentiteit gaat het immers niet om *kwalitatieve* identiteit van lichamen (twee identieke lichamen, zoals de lichamen van identieke tweelingen), maar om *numerieke* identiteit van lichamen.

Maar wat verstaan we dan onder continuïteit? In de wiskunde spreekt men wel eens over een continue lijn tussen twee punten, d.i. een lijn die zonder onderbreking twee punten met elkaar verbindt. Hoe die lijn er precies uitziet doet er eigenlijk niet toe – het zou best kunnen dat de lijn erg kronkelig is, en gaandeweg dikker of dunner, gladder of rafeliger wordt. En zo is het ook met het lichaam. Het menselijk lichaam ondergaat extreme veranderingen tussen het moment van conceptie en de dood. Elke woeste viking was ooit een boreling met een baby-velletje. Tanden, haren en nagels komen en gaan. Organen worden verwijderd of vervangen. Bovendien vertellen wetenschappers ons dat de cellen van ons lichaam zich met een rotvaart vernieuwen – in die mate zelfs dat we om de x aantal weken quasi-nieuwe nieren (en andere organen) krijgen. Maar doorheen al deze min of meer ingrijpende lichamelijke veranderingen is er een belangrijke constante: een verzameling levende cellen is onafgebroken aanwezig in de ruimte en in de tijd.

Je kan de vergelijking maken met het mythische schip van Theseus. Theseus was niet tevreden over de houtkeuze voor zijn schip, en besloot om alle planken *een voor een* te vervangen door planken van een andere houtsoort. Het schip ging met andere woorden niet onder de sloophamer, maar werd gaandeweg aangepast, plank voor plank. De vraag is dan: is het eindproduct van deze ingrijpende renovatie nog steeds hetzelfde schip? Je kan dezelfde vraag stellen over allerlei andere fenomenen die zo lang meegaan dat ze tot een stadium komen waarbij elk van hun onderdelen al eens vernieuwd werd. Een voorbeeld: er is een nieuwe trend op de tweedehandsmarkt voor piano’s. Oude piano’s worden voor de verkoop teruggestuurd naar hun originele fabrikant (Bösendorfer, Yamaha, Steinway & Sons,…), waar alle onderdelen van de piano vernieuwd worden: hamertjes, veertjes, het volledige klavier, de snaren uiteraard,… Je kan zo’n piano bezwaarlijk een nieuwe piano noemen – het wordt in ieder geval nog steeds verkocht als een tweedehands artikel. Maar is het dezelfde piano? Aanhangers van deze theorie over PI zouden zeggen van wel, omdat de piano gedurende het gehele proces een plaats innam in tijd en ruimte.

Drie argumenten pro

Quinton bespreekt drie voordelen van de theorie dat PI bestaat in fysische continuïteit (p. 400): a) theoretische eenvoud (de identiteit van personen en objecten wordt op eenzelfde manier bepaald); b) praktisch succes (in het dagelijks leven slagen we er bijna altijd in om een persoon te identificeren aan de hand van zijn lichaam); c) realisme (de theorie is een goede afspiegeling van de praktijk van identificeren). Een vierde argument (waar Quinton duidelijk bedenkingen bij heeft) luidt dat de theorie komaf maakt met het geloof in een lichaamloos bestaan. Of deze claim een argument voor of tegen de theorie is, hangt enigszins af van wie het argument hanteert: theïsten of atheïsten…

Een laatste argument, dat niet behandeld wordt door Graham, is gebaseerd op de gedachte dat lichamen (in hun geheel genomen) uniek zijn. Een van de basisprincipes van de hedendaagse evolutiebiologie is dat geen twee levende wezens identiek zijn. De variatie in de natuur is niet minder dan overweldigend. Psychologische eigenschappen zijn daarentegen minder uniek. Noem mij een verlangen waarmee u te kampen hebt en ik noem u iemand anders die precies hetzelfde verlangen heeft (u bent niet de enige die een Ipad verlangt). Je kan je zelfs voorstellen dat twee personen, zoals identieke tweelingen, ook psychologisch volledig identiek zijn.

Drie argumenten contra

De vraag is echter of fysische identiteit voldoende is om de identiteit van een persoon te kunnen garanderen. We zullen in het vervolg van dit puntje drie argumenten contra PI als fysische continuïteit poneren met name: argument van de bevruchte eicel en het lijk, de schoenlapper en de prins en als laatste argument breintransplantatie.

De bevruchte eicel en het lijk.

Ten eerste is het zo dat zelfs overtuigde katholieken (inclusief Thomas van Aquino) een onderscheid maken tussen een bevruchte eicel en een persoon – een onderscheid dat uiteraard een grote rol speelt in allerlei discussies, waaronder het debat over abortus. Een bevruchte eicel is niet meer of minder dan een hoop cellen, en wordt niet als een persoon beschouwd – zelfs niet door de paus. Pas na enkele weken (er is heel wat onenigheid over het aantal weken) spreekt men van een persoon. Nochtans is er een duidelijke fysische continuïteit tussen een hoop cellen in de buikholte van mijn moeder in de donkere maand januari van het jaar 1979, en de persoon die ik nu ben. Fysische continuïteit en identiteit lijken met andere woorden niet helemaal samen te vallen. Idem dito voor mijn lijk, trouwens. Er is een duidelijke fysische continuïteit tussen de persoon die ik nu ben en het lijk dat ik zal zijn wanneer ik net overleden ben. Net als een bevruchte eicel wordt een lijk echter niet als een persoon beschouwd. (Tussen haakjes: het lijkt zelfs paradoxaal om te spreken over ‘mijn lijk’, zoals in de uitroep ‘over mijn lijk!’ Het woord ‘mijn’ lijkt immers de aanwezigheid van een persoon te veronderstellen.) Ik kan dus niet identiek zijn aan ‘mijn lijk’, ondanks de duidelijke fysische continuïteit. Een eenvoudige repliek bij deze objectie is dat persoonsidentiteit bestaat in de fysische continuïteit van een *levend* organisme, maar dat lost het probleem van de bevruchte eicel niet op. Het concept van fysische continuïteit is duidelijk breder dan het concept van persoonsidentiteit.

De schoenlapper en de prins

Een tweede kritiek luidt dat fysische continuïteit niet altijd een voldoende voorwaarde is voor persoonsidentiteit. De Britse filosoof John Locke illustreerde deze kritiek met een gedachte-experiment. Stel je eens voor, zei Locke, dat we twee individuen van lichaam doen wisselen: een prins en een schoenlapper. De ziel van de prins verhuist naar het lichaam van de schoenlapper, en omgekeerd. En de ziel van beide mannen bevat hun herinneringen, hun karaktertrekken, verlangens, ambities, enzovoort. De vraag is dan: wie is, na deze transformaties, de prins, en wie is de schoenlapper? Volgens Locke is het evident dat we de man met het lichaam van een schoenlapper nu aanspreken als de prins, terwijl we de man met het prinselijke lichaam aanspreken als de schoenlapper. En ook de prins beschouwt zichzelf nog steeds als prins, en niet als schoenlapper, ook al zit hij plots opgezadeld met het lichaam van de schoenlapper. Volgens Locke speelt het lichaam dan ook helemaal geen rol in het bepalen van iemands identiteit – persoonsidentiteit wordt integendeel gegarandeerd door de continuïteit van het bewustzijn (psychologische continuïteit), waarover dadelijk meer.

Breintransplantaties

De kritiek van Locke weerklinkt nog in een recenter gedachte-experiment van de Amerikaanse filosoof Sydney Shoemaker (of all people!). Shoemaker vraagt ons om twee individuen voor ogen te houden die tegelijkertijd, in twee verschillende operatie-zalen, een ingrijpende hersenoperatie moeten ondergaan. We noemen hen gemakshalve mevrouw Herssens en mevrouw Breinaert. Tijdens de operatie verwijderen de chirurgen het volledige brein tijdelijk uit de schedel, om het na de operatie weer terug te plaatsen. De operatie loop echter in het honderd, en door een administratieve vergissing (‘shit, wiens brein is dit?’) wordt het brein van mevrouw Herssens in de schedel van mevrouw Breinaert geplaatst (en omgekeerd).

Wat gebeurt er wanneer beide ‘monsters’ wakker worden? Mevrouw Herssens, wiens brein in het lichaam van mevrouw Breinaert beland is, zal verwonderd zijn om het gezicht van mevrouw Breinaert in de spiegel te zien (en omgekeerd, maar dat blijf ik niet herhalen). Het lijkt vanzelfsprekend dat deze ervaring enigszins bevreemdend is – het gezicht dat mevrouw Herssens in de spiegel ziet zal ze niet ervaren als haar gezicht (een situatie die min of meer vergelijkbaar is met …-patiënten die bepaalde ledematen als lichaamsvreemd ervaren, als ledematen die niet tot hun eigen lichaam behoren, en die soms met geweld afgestoten worden). En stel dat je de kans kreeg om te spreken met deze ongelukkige dames. Wat zouden ze je vertellen wanneer je hen uitvroeg over hun werk, hun huis, hun vriendenkring? Het lijkt opnieuw vanzelfsprekend dat de persoon met het brein van mevrouw Herssens informatie zou geven over het leven van mevrouw Herssens, ook al heb je te maken met het lichaam van mevrouw Breinaert. Onze intuïtie zegt ons, net als in het gedachte-experiment van Locke, dat de identiteit van mevrouw Herssens voornamelijk, en misschien zelfs uitsluitend, te vinden is in haar herinneringen, haar vriendenkring, haar karakter, verlangens en overtuigingen – ook al liggen die herinneringen etc tijdelijk vervat in een ander lichaam. Quinton vult deze analyse mooi aan door het perspectief van de geliefden, familie en vrienden van beide dames in te nemen. ‘In our general relations with other human beings their bodies are for the most part intrinsically unimportant. We use them as convenient recognition devices enabling us to locate without difficulty the persisting character and memory complexes in which we are interested, which we love or like. It would be upsetting if a complex with which we were emotionally involved came to have a monstrous or repulsive physical appearance (…), but it is surely clear that our concern and affection would follow the character and memory complex and not its original bodily associate’ (402-3).

Kortom: de continuïteit van het hele lichaam is op zich niet voldoende om de identiteit van een persoon vast te leggen. (Zoals Quinton zegt: ‘Bodily identity is not a logically complete [= sufficient] criterion of personal identity; at best it could be a necessary condition of personal identity’ (p. 402)). De continuïteit van een van de belangrijkste onderdelen van het lichaam, namelijk het brein, lijkt wel een voldoende voorwaarde te zijn. Merk op dat onze intuïties op dit punt mooi overeenkomen met een fysicalistische zienswijze. Onze intuïtie zegt ons dat persoonsidentiteit bepaald wordt door psychologische continuïteit. Het fysicalisme stelt dat onze psychologie (onze geest, zeg maar) ons brein *is*, of tenminste op ons brein supervenieert. Shoemakers gedachte-experiment brengt onze intuïtie en het fysicalisme samen: mijn identiteit als persoon hangt samen met (de continuïteit van) mijn brein. Houdt de theorie van fysische continuïteit dan toch stand? Het volgende en laatste tegenargument suggereert van niet.

* 1. **PI als psychologische continuïteit**

Basisclaims

Het verhaal van Sacks (zie tekst op toledo) is een bevestiging van een vermoeden dat ik al eerder uitsprak, in de context van de eerste theorie over persoonsidentiteit, namelijk de fysische continuïteitstheorie. Het vermoeden is dat de continuïteit van het lichaam misschien wel een noodzakelijke voorwaarde vormt voor persoonsidentiteit, maar geen voldoende voorwaarde. De mijnheer Thompson die in het ziekenhuis met Oliver Sacks spreekt heeft nog steeds hetzelfde lichaam als mijnheer Thompson van de delicatessenzaak, maar ik vermoed dat weinig mensen zullen beweren dat het in beide gevallen om dezelfde persoon gaat. Maar wat verklaart dan het verlies van identiteit in het geval van mijnheer Thompson? Sacks geeft zelf het antwoord: de levensgeschiedenis van mijnheer Thompson is volledig verhakseld. ‘Allemaal hebben we een levensgeschiedenis, een innerlijk verhaal waarvan de continuïteit en de zin ons leven is. Je zou kunnen zeggen dat ieder van ons een “verhaal” opbouwt en leeft, en dat wij dit verhaal zijn, dat dit onze identiteit is’ (128). Het levensverhaal van mijnheer Thompson is een boek waaruit hele hoofdstukken verdwenen zijn; in de overgebleven hoofdstukken ontbreken talloze pagina’s, en zijn allerlei woorden en namen geschrapt en van plaats verwisseld. Het verhaal is onleesbaar geworden, zowel voor de argeloze lezer als voor de auteur, mijnheer Thompson zelf. Mijnheer Thompson leeft ‘in die onveranderlijke tegenwoordige tijd van hem’ (131). De continuïteit tussen zijn huidige mentale toestanden en zijn vroegere mentale toestanden is onderbroken. En deze onderbreking verklaart waarom we geneigd zijn om te spreken over ‘een andere persoon’. In deze zienswijze hangt persoonsidentiteit niet zozeer vast aan de continuïteit van het lichaam, maar wel aan de continuïteit van de geest.

‘Continuïteit’ betekent in deze context zoveel als ‘de verbondenheid van een reeks mentale toestanden’. Deze definitie vereist een beetje uitleg. Aanhangers van deze theorie over persoonsidentiteit beschouwen de identiteit van een persoon als de eenheid van een geest. ‘Eenheid’ kan je in deze context op twee manieren interpreteren. Enerzijds is er het probleem van de *synchrone* eenheid van de geest. Wat maakt dat de verschillende mentale toestanden die ik op eenzelfde ogenblik ervaar (voelen, horen, denken,…), toestanden van eenzelfde geest zijn? Een moment van introspectie leert ons dat we ons op eenzelfde tijdstip bewust kunnen zijn van verschillende mentale toestanden, maar introspectie leert ons niets over een centrale dirigent – een ik die eigenaar of auteur is van al deze toestanden. Anderzijds is er het probleem van de *diachrone* eenheid van de geest. Wat maakt dat de verschillende mentale toestanden die ik op verschillende, opeenvolgende tijdstippen ervaar, toestanden van eenzelfde geest zijn? Eenvoudiger geformuleerd: hoe komt het dat ik mijn levensverhaal als een enkel verhaal ervaar, en niet als een veelheid van verhalen? Aanhangers van deze theorie over persoonsidentiteit (als psychologische continuïteit) beantwoorden deze vraag als volgt: mijn levensverhaal is een enkel verhaal in de mate dat de verschillende fasen van het verhaal op een of andere manier op elkaar inhaken. Persoonsidentiteit is mentale verbondenheid. Voorbeelden van zulke verbondenheid zijn legio: we herinneren ons mentale ervaringen uit het verleden (de pijn toen we als kind van de trap vielen); ons temperament verandert niet of nauwelijks doorheen ons leven; we voeren plannen uit die we beraamden en koesterden toen we kind waren; enzovoort.

Maar hoe sterk moet de continuïteit of verbondenheid dan zijn? Ons geheugen laat ons toch voortdurend in de steek? Of neem een ernstiger probleem: retrograde amnesie. Amnesie=geheugenverlies. Na een ongeval (of bij dementie, of in het syndroom van Korsakov, of in het geval van overmatig alcoholgebruik) kunnen mensen zich soms niets meer herinneren van wat er gebeurde *onmiddellijk voor* het ongeval. Soms gaat het over enkele minuten, soms over enkele uren of dagen. Hun geheugen is als het ware gedeeltelijk gewist, en de continuïteit van hun levensverhaal wordt bijgevolg onderbroken. Verliezen ze daarom ook hun identiteit als persoon? En is het niet zo dat mensen voortdurend van mening veranderen in hun leven? Sommigen veranderen zelfs van karakter, ten goede of ten kwade. En al klagen we wel eens dat iemand ‘niet meer dezelfde is als vroeger’ (Philippe Gilbert!), toch spreken we niet gauw over ‘een nieuwe persoon’. Hoe is deze vaststelling te rijmen met de theorie over PI als psychologische continuïteit? De theorie is gebaseerd op de gedachte dat er verschillende loci (meervoud van ‘locus’, Latijn voor ‘plaats’) zijn van continuïteit. In de eerste les heb ik uitgelegd dat de geest ons toelaat om allerlei dingen te doen: we kunnen ervaren, herinneren, redeneren, fantaseren, plannen beramen, enzovoort. Elk van deze mentale activiteiten en toestanden heeft haar eigen continuïteit. En wanneer ons geheugen het even laat afweten, zijn er nog tal van andere activiteiten en toestanden die de continuïteit verzorgen. Anders gezegd: er zijn verschillende rode draden in ons levensverhaal, en wanneer de ene draad eindigt zijn er nog steeds tal van andere draden die verderlopen.

De geheugentheorie over PI

Een van de belangrijkste theorieën die PI analyseren in termen van psychologische continuïteit, is de zogenaamde ‘geheugen-theorie’ over PI. De theorie werd voor het eerst verdedigd door de Britse filosoof John Locke, en speelt een duidelijke rol in het verhaal van Sacks dat we zonet bespraken. In wat volgt vat ik Lockes theorie samen, en formuleer ik enkele pijnpunten.

Het geheugen is volgens Locke de enige voldoende en noodzakelijke voorwaarde voor persoonsidentiteit. Het geheugen maakt met andere woorden dat een persoon op tijdstip 1 identiek is aan een persoon op tijdstip 2; het is de *ratio essendi* van persoonsidentiteit. Locke zelf definieert het geheugen als ‘the consciousness [that] can be extended backwards to any past action or thought’ – een bewustzijn van ervaringen, handelingen en gebeurtenissen uit het verleden.

Er zijn echter verschillende soorten of vormen van geheugen. Ik herinner me bijvoorbeeld uit de colleges Nederlandse literatuur dat de dichter Herman Hendrik Ter Balkt (Harry voor de vrienden) tussen 1969 en 1977 het eigenaardige pseudoniem ‘Habakuk II De Balker’ gebruikte. En ik herinner me dat de docent van het vak Nederlandse literatuur sigaretten rookte van het merk Groene Michel. Zulke herinneringen zijn feitelijke herinneringen, en behoren bijgevolg tot het feitelijk geheugen (*factual memory*). Daarnaast herinner ik me ook hoe je een lekkere risotto met boschampignons moet bereiden, en wat je precies moet doen om je moeder te laten geloven dat je ziek bent en niet naar school kan. Zulke herinneringen zijn herinneringen van gewoonten, en behoren bijgevolg tot het habituele geheugen (*habitual memory*). Het onderscheid tussen feitelijke en habituele herinneringen weerspiegelt trouwens het onderscheid tussen ‘weten dat’ en ‘weten hoe’ in de epistemologie – een onderscheid dat eerder aan bod kwam in de tekst van Smart. Deze twee vormen van geheugen maken mij niet bepaald uniek. De meeste andere studenten in de les Nederlandse literatuur kenden het pseudoniem van Ter Balkt ook. En mijn broer was even doortrapt als ik wanneer het op spijbelen aankwam. Zulke herinneringen maken mij met andere woorden niet tot een unieke persoon. Onze identiteit ontlenen we eerder aan een derde vorm van geheugen: het ervaringsgeheugen (*experiential memory*). Feitelijke en habituele herinneringen zijn onpersoonlijke herinneringen; ik deel ze met andere personen. Herinneringen aan persoonlijke ervaringen uit het verleden zijn daarentegen hoogst persoonlijk. Ik alleen heb rechtstreekse toegang tot zulke herinneringen. Andere personen kunnen zich mijn ervaringen herinneren (wanneer ik die ervaringen met hen gedeeld heb, bijvoorbeeld door erover te spreken), maar ze kunnen zich niet het hebben van mijn ervaringen herinneren. In het Engels klinkt het beter: ‘I can remember other people’s experiences, but I cannot remember their having those experiences’, zegt Cuypers.

Twee argumenten contra

De pro argumenten kan men afleiden uit het verhaal van Sacks (inderdaad, doe zelf ook eens iets).

De twee contra argumenten die we zullen bespreken zijn de problemen van circulariteit en absurditeit.

Probleem van circulariteit

Een eerste probleem met Lockes analyse van het concept van persoonsidentiteit is dat ze (de analyse) circulair is. Het gaat met andere woorden om een cirkelredenering – een van de drogredenen uit de logica. U herinnert zich nog dat de geheugentheorie een variant is van een reductionistische theorie over PI. PI-reductionisten beweren dat het concept van PI ‘gereduceerd’ (of ‘verder geanalyseerd’) kan worden tot andere concepten, zoals het concept van continuïteit. Locke beweert dat het concept van PI gereduceerd kan worden tot het concept van geheugen, of continuïteit van ervaringsherinneringen. Het probleem met deze analyse is dat het concept van ervaringsgeheugen eigenlijk het concept van PI *vooronderstelt*. Mijn herinnering aan een bepaalde ervaring in het verleden is altijd een herinnering aan MIJN toenmalige ervaring. Ik kan me enkel mijn eigen *persoonlijke* ervaringen herinneren – ervaringen die ik had in het verleden.

b) Je kan de cirkelredenering als volgt samenvatten: het geheugen vormt de bepalende voorwaarde van iemands identiteit als persoon (ervaringsgeheugen 🡪 PI), maar veronderstelt tegelijkertijd deze persoonsidentiteit (PI 🡪 ervaringsgeheugen).

c) Volgens aanhangers van de geheugentheorie is deze kritiek gebaseerd op taalverwarring. Volgens hen is het mogelijk om concept van geheugen te analyseren zonder gebruik te maken van het concept van persoonsidentiteit. Wat bedoelen we eigenlijk wanneer we zeggen dat iemand zich een bepaalde ervaring herinnert?

*I remember having experience E’ if and only if a) I apparently remember having experience E; b)* ***I*** *had experience E; and c) my apparent memory of having experience E is caused in an appropriate way by* ***my*** *having experience E in the past*

*‘I remember having experience E’ if and only if a) I apparently remember having experience E; b) someone had experience E; and c) my apparent memory of having experience E is caused in an appropriate way by someone having experience E in the past*

Het eerste citaat (bovenaan) illustreert de cirkelredenering in de geheugentheorie over PI. Het geheugen bepaalt zogezegd de identiteit van een persoon, maar doet tegelijkertijd beroep op die identiteit (‘I had…’; ‘My having experience…’). Misschien kort wat uitleg. Voorwaarde b) wijst erop dat we ons enkel een ervaring kunnen herinneren wanneer we die ervaring ook daadwerkelijk gehad hebben. Deze voorwaarde lijkt misschien redundant (overbodig), maar ze bepaalt eenvoudigweg dat bijvoorbeeld herinneringen die gebaseerd zijn op hallucinaties of waanbeelden geen echte herinneringen zijn. Voorwaarde c) is in zekere zin een uitbreiding van voorwaarde b) en stipuleert dat een herinnering op een gepaste manier moet veroorzaakt worden door een ervaring uit het verleden. Het gebeurt soms dat iemand zich iets denkt te herinneren, terwijl de herinnering in kwestie eigenlijk gebaseerd is op de herinnering van iemand anders aan een bepaalde ervaring. Mijn moeder vertelt vaak dat ik haar als kind voortdurend en overal volgde. Wanneer ze naar de badkamer ging wachtte ik bijvoorbeeld als een trouwe hond aan de deur. Mijn moeder heeft dit verhaal al zo vaak verteld dat ik er voor mezelf een mentaal beeld van gevormd heb, en op basis van dat beeld denk ik me mijn gedrag te herinneren. In werkelijkheid was ik te klein om echte herinneringen te vormen – mijn herinneringen zijn geen echte herinneringen, maar reconstructies van een voorval dat door iemand anders (mijn moeder) herinnerd werd.

Hoe dan ook, de drie voorwaarden in het eerste citaat geven aan wat echte herinneringen zijn, en dus ook wat het geheugen eigenlijk is. Het tweede citaat (onderaan) is een herformulering van het eerste citaat, met enkele kleine doch belangrijke aanpassingen. Je merkt dat woorden als ‘ik’ en ‘mijn’ in deze variant verdwenen zijn, en vervangen door het onpersoonlijke ‘iemand’ (someone). Wanneer ik zeg dat ik me een bepaalde ervaring herinner hoeft dat niet te betekenen dat ik die ervaring had – het is voldoende dat iemand die ervaring had, en dat die iemand mij de ervaring doorspeelde op een causaal gepaste manier. Het tweede citaat toont met andere woorden aan dat je het concept van geheugen kan analyseren zonder beroep te doen op het concept van persoonsidentiteit.

Probleem van absurditeit

Het is bekend dat sommige mensen helemaal geen herinneringen hebben aan bepaalde delen van hun leven. De meeste mensen herinneren zich weinig of niets van de ervaringen die ze hadden tijdens hun eerste drie levensjaren. Niemand kan zich zijn of haar geboorte herinneren. In de logica van de geheugentheorie leidt deze vaststelling tot allerlei absurditeiten. Een van de critici van John Locke, Thomas Reid (1710-1796), gaf het volgende voorbeeld (dat door Quinton aangehaald wordt op p. 397 onderaan). Stel je voor dat iemand als kind gestraft wordt voor het stelen van appels uit de tuin van de buurman. Diezelfde persoon herinnert zich het voorval (de straf) wanneer hij jaren later, als jonge officier, een medaille krijgt voor heldenmoed tijdens de oorlog. En aan het eind van zijn leven herinnert deze persoon, die ondertussen generaal op rust is, zich de medaille-ceremonie, maar NIET het voorval met de appels. De geheugentheorie leert ons dat het kind dezelfde persoon is als de jonge officier, en dat de oude generaal dezelfde persoon is als de jonge officier, maar ook dat de oude generaal NIET dezelfde persoon is als het kind. Die conclusie is echter absurd, zegt Reid. Waarom? Omdat persoonsidentiteit per definitie transitief is. Als X=Y en Y=Z dan volgt daar onvermijdelijk uit dat X=Z. Volgens de regels van de logica moet de oude generaal identiek zijn aan de kleine dief – een conclusie die haaks staat op de conclusie die volgt uit de geheugentheorie.

Aanhangers van de geheugentheorie beantwoorden deze kritiek door een onderscheid te maken tussen een directe (rechtstreekse) en een indirecte (onrechtstreekse) continuïteit van het geheugen, en door te stellen dat enkel de indirecte vorm van geheugencontinuïteit transitief is. Indirecte continuïteit van het geheugen houdt in dat er sprake is van een ononderbroken keten van overlappende episodes van directe continuïteit.

Cuypers formuleert het als volgt: ‘Memory continuity, then, consists of an uninterrupted chain of overlapping links of (direct) memory connectedness’. Continuïteit is transitief, connectiviteit niet.

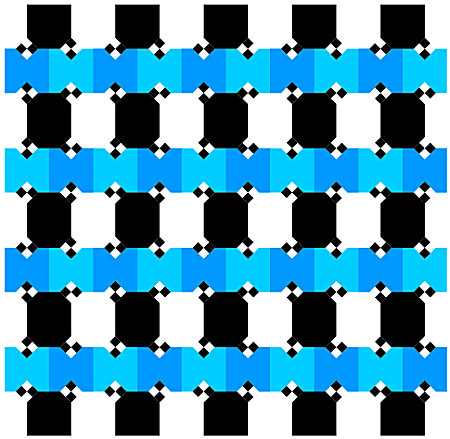
Vergeet echter niet dat het geheugen maar één van de vele elementen is in een omvattende theorie die PI beschrijft als psychologische continuïteit. Persoonsidentiteit wordt niet alleen bepaald door de continuïteit van ervaringsherinneringen. Onze identiteit hangt ook samen met de continuïteit van andere mentale toestanden, zoals karakter, intenties, verlangens, overtuigingen, enzovoort. Zoals Quinton zegt: ‘a more plausible criterion can be offered in terms of continuity of character *and* memory’ (p. 398, mijn cursivering). Wel is het zo dat herinneringen uiteindelijk een groter gewicht hebben in het bepalen van iemands identiteit als persoon, omdat bijvoorbeeld overtuigingen en verlangens vaak minder uniek zijn dan herinneringen.

1. Modulariteit



De waar-is-wally afbeelding

U weet niet waar Wally zich precies bevindt in de afbeelding. U zoekt naarstig, en u vindt hem, precies in het midden. Tenzij u ernstig dement bent of een of andere neurologische afwijking hebt, kan u Wally, nadat u hem hebt gevonden, niet meer NIET vinden. Het lijkt erop dat uw geheugen uw visuele waarneming min of meer blijvend kan beïnvloeden. Tip: deze mogelijkheid wordt beschreven in de tekst van Samuels op p. 52, waarin hij aangeeft dat er verschillende vormen van cognitieve beïnvloeding bestaan op de waarneming (‘many sorts of external cognitive influence on perception’), met als eerste voorbeeld: ‘shifts in loci of focal attention brought about by clues, instructions, or preferences about where to look’. Wally zoeken en vinden is een goed voorbeeld van hoe instructies (die u moet begrijpen, en begrijpen is een cognitief proces) de waarneming kunnen beïnvloeden.



Hoofdpijn veroorzakende optische illusie

Visuele illusies zijn fenomenen die in zekere zin het tegendeel lijken aan te tonen van het vorige voorbeeld. Zelfs wanneer ik mijn meetlat bovenhaal en u er met harde feiten op wijs dat de lijnen perfect parallel lopen, dan kan u niet anders dan kromme lijnen te zien. U kan de kromme lijnen niet NIET zien, ook al kent u de ware toedracht. Uw kennis en redeneringsvermogen lijken in dit voorbeeld geen toegang te hebben tot uw visuele waarneming. Cognitiewetenschappers noemen de visuele waarneming dan ook een ‘ingekapseld’ systeem, en ‘ingekapseld’ is, zoals we zullen zien, een bepalend kenmerk van modulariteit.

In de vorige les bespraken we een thema, met name persoonsidentiteit, dat nog steeds nauw aanleunde bij het hoofdthema van de eerste lessen, met name de verhouding tussen lichaam (of hersenen) en geest. In deze les concentreren we ons vooral op de vraag naar de architectuur van de menselijke geest. Waaruit is de menselijke geest opgebouwd? Een van de mogelijke antwoorden op deze vraag is dat de geest een modulaire architectuur heeft, dat wil zeggen opgebouwd is uit een reeks van mentale modules. Deze modules worden vaak op een functionalistische manier begrepen – het zijn elementen die ons toestaan om inputs om te zetten in outputs. Modules stellen ons in staat om bepaalde dingen te doen. In die zin wordt de menselijke geest soms wel eens vergeleken met een zwitsers zakmes, waarvan elk van de onderdelen een specifieke functie vervult: een tandenstoker, een tang, een nagelknipper, een vergrootglas, een schroevendraaier, enzovoort.

* 1. **Wat is modulariteit?**

Een eerste definitie

Er zijn verschillende varianten van de theorie dat de menselijke geest modulair is, zoals ik vandaag zal aantonen. ‘Modularisten’ verschillen vooral van mening over de vraag in welke mate de menselijke geest modulair is. Maar voor we deze varianten tegen het licht houden, moeten we eerst iets zeggen over wat modules precies zijn. Een eerste definitie vind je in de tekst van Richard Samuels, ‘Is the human mind massively modular?’ Samuels maakt een onderscheid tussen twee definities van de term of het concept ‘module’:

De term ‘module’ wordt soms op een minimale manier ingevuld, als een synoniem voor de term ‘functioneel cognitief mechanisme’ of ‘mentaal vermogen’. Denk daarbij aan de manier waarop het functionalisme de geest beschrijft, namelijk als een black box die ons in staat stelt om ‘inputs’ (bijvoorbeeld: zintuiglijke informatie) om te zetten in ‘outputs’ (bijvoorbeeld: gedrag). Volgens Samuels hanteren evolutiepsychologen, zoals John Tooby, Leda Cosmides en Steven Pinker, soms een minimale definitie van het concept ‘module’: ‘complex structures that are functionally organized for processing information’ (p. 39). Even verderop in de tekst (p. 40) haalt hij Tooby & Cosmides nogmaal aan, in een citaat waarin zij modules eenvoudigweg gelijkstellen met computers.

Andere definities van de term ‘module’ zijn ‘robuuster’. Men spreekt dan van ‘Fodor-modules’, omdat deze zienswijze op wat modules zijn afkomstig is van een van de grote filosofen van de voorbije decennia: Jerry Fodor.

Jerry Fodor en de fodor-modules

Fodor zette zijn ideeën over modulariteit voor het eerst uiteen in een boek uit 1983, met de bedrieglijke titel *The Modularity of Mind*. De titel is bedrieglijk omdat de geest volgens Fodor niet louter bestaat uit (robuuste) modules. De architectuur van de menselijke geest is tweevoudig, meent hij, en enkel ‘perifere’ mentale vermogens, zoals taal en waarneming, zijn modulair. ‘Centrale’ mentale vermogens, zoals redeneren en beslissen, zijn niet-modulair. Maar daarover straks meer. Waarom is Fodors invulling van de term ‘module’ robuust? Omdat een module niet zomaar een cognitief mechanisme is, maar een zeer specifiek cognitief mechanisme, met specifieke kenmerken.

Volgens Fodor bestaat er niet zoiets als de essentie van modulariteit (zie ook voetnoot 3 in de tekst van Samuels). We noemen een mentaal vermogen modulair wanneer het een aantal van de bovengenoemde eigenschappen vertoont. Hoeveel? Die vraag laat Fodor onbeantwoord. Wel is het zo dat sommige eigenschappen meer gewicht hebben. ‘Ingekapseld’ (*encapsulated*) en ‘domein-specifiek’ zijn bijvoorbeeld erg belangrijke eigenschappen van modulariteit, zoals Samuels ook aangeeft. Je kan deze aanpak vergelijken met de manier waarop psychiaters tot een diagnose komen voor een bepaalde patiënt. Ik denk dan vooral aan de diagnostische werkwijze van het Amerikaanse *Diagnostic and Statistical Manual of Mental Disorders* (DSM) – een van de dominante psychiatrische handboeken van vandaag. De DSM stipuleert dat een patiënt een aantal symptomen moeten vertonen uit een (uitgebreide) lijst om in aanmerking te komen voor een bepaalde diagnose. Daarbij is het wel zo dat sommige symptomen een groter gewicht hebben dan andere. In het geval van schizofrenie, bijvoorbeeld, moet er sprake zijn van … symptomen uit een lijst van …, behalve wanneer er sprake is van hallucinaties of waanvoorstellingen. Deze symptomen zijn min of meer definiërend voor de ziekte in kwestie, en dus leiden ze ook tot een diagnose wanneer er nauwelijks sprake is van andere symptomen uit de lijst.

De eigenschappen van de Fodor-modules kan je in vijf verschillende clusters samenbrengen:

Ingekapseld en ontoegankelijk

Elk cognitief systeem verwerkt informatie uit een specifiek domein. Het gezichtsvermogen, bijvoorbeeld, verwerkt enkel visuele prikkels. Een module is ingekapseld in de mate dat ze geen toegang heeft tot ‘andere’ informatie terwijl ze ‘haar eigen’ informatie verwerkt. Wat bedoel ik dan precies met ‘andere informatie’? Daarmee bedoel ik: informatie die elders opgeslagen of verwerkt wordt, bijvoorbeeld in het geheugen. Gegeven Fodors dualiteit kan het dus zowel gaan om informatie uit andere modules als om informatie uit centrale cognitie-processen. Een voorbeeld kan veel verduidelijken. Op de volgende slide ziet u een visuele illusie, de zogenaamde Zöllner-illusie. De lange lijnen in de tekening lijken niet mooi parallel te lopen – ze lijken op het eerste gezicht schots en scheef. Wanneer we de meetlat erbij halen, merken we echter dat de lange lijnen wel mooi evenwijdig lopen. Vreemd genoeg heeft deze (centrale) informatie weinig of geen effect op mijn gezichtsvermogen. Ook al weet ik dat de lijnen in werkelijkheid mooi evenwijdig lopen, toch lijkt het nog steeds alsof ze schuin lopen. Ergo: ons gezichtsvermogen is een ingekapselde module. Ze kan enkel beroep doen op interne informatie en inputs om haar informatie te verwerken. Informatie uit andere systemen (in het voorbeeld: de informatie dat de lijnen wel evenwijdig zijn) heeft geen invloed op de manier waarop de ‘gezichtsmodule’ haar informatie verwerkt.

‘Ontoegankelijkheid’ is eigenlijk de keerzijde van ‘ingekapseldheid’. Modules zijn ingekapseld in de mate dat ze ongevoelig zijn voor externe informatie (dat wil zeggen informatie die niet tot de module zelf behoort); ze zijn ontoegankelijk in de mate dat hun informatieverwerking zich onttrekt aan introspectie. We hebben met andere woorden geen toegang tot de informatie die verwerkt wordt in modules (terwijl omgekeerd de module ook geen toegang heeft tot informatie van buitenaf). Wanneer we luisteren naar wat iemand ons vertelt, dan is het wellicht zo dat ons brein naarstig verschillende opties afloopt, om ons uiteindelijk één betekenis voor te houden (uit verschillende mogelijkheden). Het zoekproces is echter niet toegankelijk voor ons. We moeten genoegen nemen met het resultaat – de output van de taalmodule.

Automatisch snel en oppervlakkig

Een module is ‘automatisch’ (*mandatory*) in de mate dat ze zich onttrekt aan bewuste controle. Je kan met andere woorden niet NIET willen wat een module doet. De module start op wanneer een relevante stimulus zich aandient, en stopt pas met werken wanneer de stimulus verwerkt is. Een voorbeeld: wanneer iemand in jouw moedertaal tegen je spreekt, kan je geen ruis horen. Je hoort betekenisvolle woorden en zinnen, hoe hard je ook je best doet om gewoon klanken te horen – zoals wanneer je naar de Russische radio zou luisteren.

Een module is ‘snel’ wanneer ze minder dan een halve seconde nodig heeft om informatie te verwerken. Het gezichtsvermogen is een goed voorbeeld van een snelle module. Wanneer je proefpersonen op een scherm een reeks van beelden toont die elk gedurende 125 milliseconden (dus 1 tiende van een seconde) zichtbaar zijn, dan kunnen ze zeven op de tien beelden herkennen (negen op de tien wanneer je elk van de beelden 167 milliseconden toont). Fodor legt de lat op 0,5 seconde (dus 500 milliseconden). Snelheid hangt uiteraard samen met ingekapseldheid – een module die geen rekening houdt met wat er in andere modules en in het centrale systeem gebeurt, kan sneller werken. Vergelijk het met een eenmanszaak: iemand die ‘zijn eigen baas’ is, hoeft geen rekening te houden met de meningen van medewerkers of bazen. Hij of zij zal sneller tot beslissingen komen. (Noot: de snelheid van modulaire systemen wordt vaak uitgebuit door handige reclame-jongens.)

Een module is ‘oppervlakkig’ (*shallow*) wanneer ze weinig ‘energie’ (of rekenkracht) behoeft om tot een resultaat te komen. Je zou kunnen spreken van ‘goedkope’ systemen – systemen die weinig vergen van het organisme.

Dissocieerbaar en lokaliseerbaar

Een module is dissocieerbaar in de mate dat haar uitvallen (of disfunctioneren) de werking van andere modules en de werking van centrale systemen niet belemmert. De neuropsychologie leert ons dat sommige patiënten specifieke mentale vermogens volledig kunnen verliezen, zonder dat ze daarom ophouden te functioneren. Mensen met dyslexie, bijvoorbeeld, kunnen best heel erg sociaal of intelligent zijn. Mensen met totale kleurenblindheid (achromatopsia) kunnen niettemin perfect horen en spreken, enzovoort.

De neurologie leert ons dat de problemen van zulke mensen vaak veroorzaakt wordt door hersenletsels (tumor, trauma, ontsteking,…), en dat betekent dat de modules waarvan het uitvallen de problemen veroorzaakt, lokaliseerbaar zijn in het brein. Het zien van kleuren, het detecteren van beweging, en het herkennen van gezichten zijn vermogens die sterk anatomisch verankerd zijn.

Domein-specifiek

Het domein is de verzameling van relevante stimuli. Fodor verwoordt het op een andere manier met name de verzameling van vragen waarop een mentaal vermogen antwoord biedt. Des te minder vragen men dan heeft des te kleiner het domein is.

Aangeboren (tegenover verworven)

In de zin van biologisch geprogrammeerd. Een module is aangeboren in de mate dat haar ontstaan op een of andere manier biologisch geprogrammeerd is. Meestal is er enkel een kleine trigger nodig om de constructie van de module in gang te zetten. Fodor zelf spreekt over ‘developing according to specific endogenously determined patters under the impact of environmental releasers’ (1983). ‘Aangeboren’ staat hier tegenover ‘verworven’. Wanneer een module aangeboren is dan zal ze zich ontwikkelen ongeacht de natuurlijke of sociale omstandigheden van het individu. Taal is een goed voorbeeld. Overal ter wereld leren kinderen min of meer op eenzelfde tempo een taal. Rond hun eerste levensjaar kunnen ze enkele woorden uitspreken; rond anderhalf jaar spreken ze reeds in telegram-stijl; en rond hun tweede levensjaar kunnen ze complexe zinnen vormen. ‘Aangeboren’ en ‘universeel’ zijn echter niet logisch verbonden. Het is best mogelijk dat een module enkel aangeboren is bij een welomlijnde groep van mensen (net zoals sommige genetische afwijkingen enkel voorkomen in bepaalde populaties); en omgekeerd is het ook mogelijk dat een universeel cognitief systeem niet aangeboren is.

* 1. **Bescheiden modulariteit**

Jerry Fodor verdedigt een theorie die je zou kunnen omschrijven als de ‘bescheiden modulariteitstheorie’. De titel van zijn eerste boek over het onderwerp, *The Modularity of Mind*, is eigenlijk bedrieglijk. Fodor meent namelijk niet dat alle mentale vermogens (of systemen, of hoe je de onderdelen van de menselijke geest ook wil noemen) modulair zijn. Enkele systemen, zoals taal en perceptie, zijn overduidelijk modulair, beweert Fodor. Maar daarnaast zijn er nog een heleboel systemen die allesbehalve modulair zijn. Fodor doet nogal geheimzinnig over de aard van deze systemen, maar hij lijkt genoegen te nemen met een negatieve typering: ‘centrale systemen’ zijn alles wat modules niet zijn.

Richard Samuels is het trouwens helemaal eens met Fodor op dit punt: ‘I thus advocate a middle way between those who endorse a thoroughgoing massive modularity [zoals Peter Carruthers & de evolutiepsychologen] and those who reject modularity altogether [zoals Jesse Prinz]’ (p. 37). En dat is precies ook zijn conclusie: na meer dan twintig jaar onderzoek over modulariteit blijkt de zienswijze van Fodor nog steeds de meest plausibele te zijn. ‘The situation is, in other words, much as Fodor advocated over two decades ago’ (p. 52).

****

Wat zijn ‘input systems’? Dat zijn mentale systemen of vermogens, zoals het gezichtsvermogen of de tastzin, die inkomende informatie omzetten in hypothesen over voorwerpen in de wereld. Deze hypothesen (‘daar loopt mijn vriend Barrie’; ‘het regent (want ik voel regendruppels)’) vormen de basis waarop de ‘central systems’ verderwerken. Eenvoudig geformuleerd: zintuigelijke informatie speelt een cruciale rol in het vormen van oordelen, verwachtingen, herinneringen, enzovoort. Ik vermoed dat Fodor het eens zou zijn wanneer we zouden zeggen dat de centrale systemen zich voornamelijk bezighouden met het produceren van propositionele attitudes.

Hoe dan ook: Fodors theorie steunt dus eigenlijk op twee verschillende claims:

a) sommige mentale vermogens zijn modulair;

b) andere mentale vermogen zijn niet-modulair.

Laten we beide claims even tegen het licht houden.

Tegenargumenten tegen stelling a)

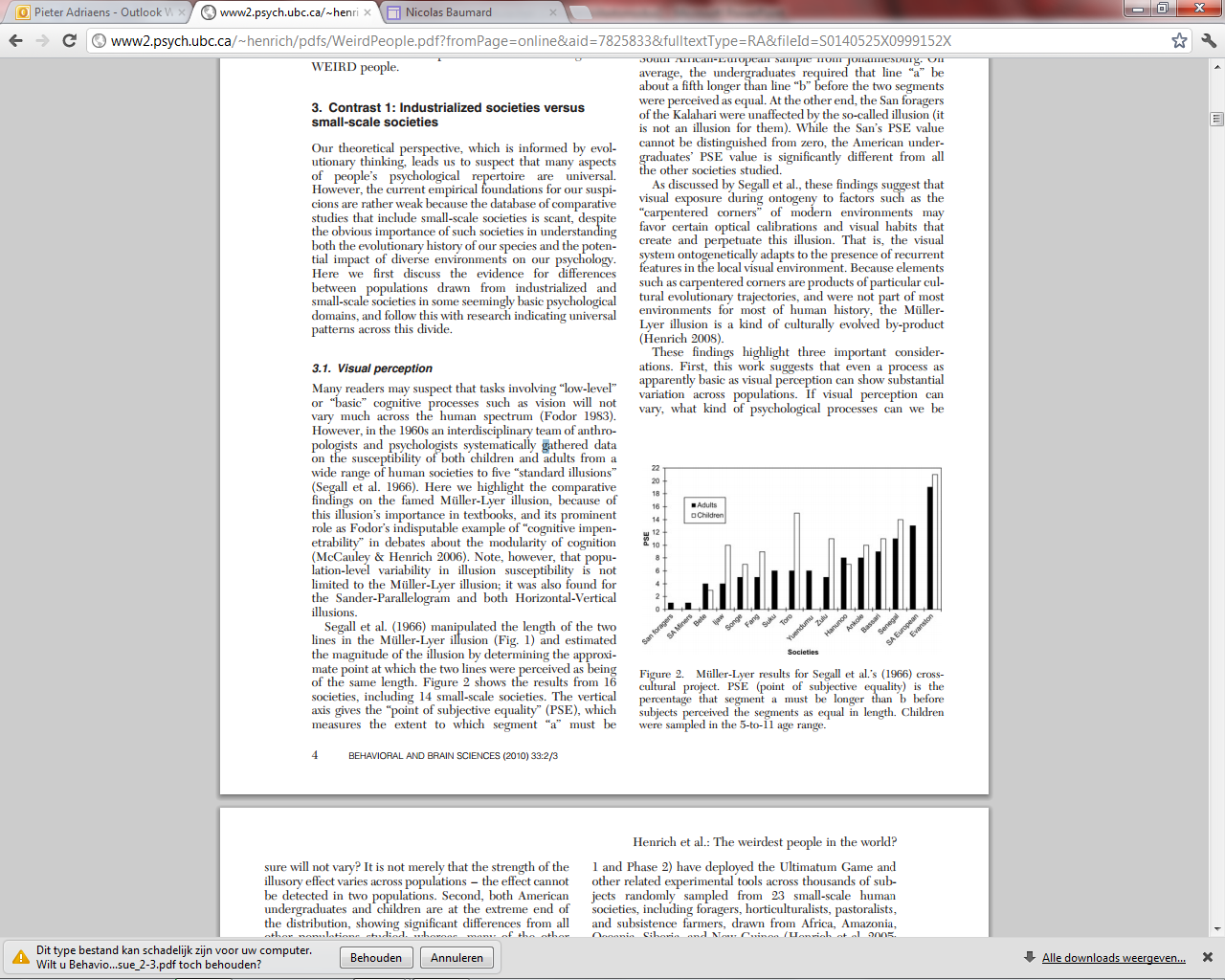
We zullen in de komende alinea’s twee argumenten poneren die tegen de eerste claim – taal en waarneming zijn modulair – die Fodor legt, ingaan. Eerste: Taal en waarneming zijn niet modulair want niet ingekapseld, tweede: taal en waarneming zijn niet modulair want niet aangeboren.

Eerste kritiek: Ingekapseld?

Een eerste argument tegen de stelling dat taal en waarneming modulair zijn, werd aangedragen door de jonge Amerikaanse filosoof Jesse Prinz. Hij verwijst naar een merkwaardig effect uit de psycho-akoestiek, waarbij ontbrekende elementen in een gesproken boodschap automatisch aangevuld worden door onze hersenen. In het dagelijks leven komt het vaak voor dat omgevingsgeluiden een conversatie verstoren; in zulke gevallen slaagt het brein er niettemin in om lacunes op te vullen. Volgens Prinz worden ontbrekende fonemen (‘taalklanken’) aangevuld met behulp van allerlei overtuigingen van de luisteraar over wat er precies gezegd wordt. Met andere woorden: taalwaarneming (het horen van gesproken taal) wordt blijkbaar WEL beïnvloed door informatie uit een centraal systeem, met name onze overtuigingen. Taalwaarneming is dus niet ingekapseld, en als we ‘ingekapseldheid’ als een definiërend criterium voor modulariteit beschouwen, is taalwaarneming niet modulair.

MAAR: het *phoneme restoration effect* is een reeël effect, maar het kan op verschillende manieren geïnterpreteerd worden. De interpretatie van Prinz is dat het een effect is waarbij centrale systemen zogenaamd modulaire systemen beïnvloeden. Maar je kan evengoed beweren dat het verwerken van informatie over zinnen met ontbrekende fonemen niet gebeurt met behulp van informatie van buitenaf (zoals overtuigingen), maar dat het geheel en al gebeurt met behulp van informatie die aanwezig is in de taalwaarnemingsmodule. Het zou bijvoorbeeld kunnen dat we de ontbrekende klanken aanvullen met behulp van een soort linguïstische databank met allerlei types van woorden.

Ingekapseld en aangeboren?



Een tweede argument tegen de stelling dat taal en waarneming modulair zijn, werd aangedragen door de Amerikaanse psycholoog Joseph Henrich, die zich daarbij beriep op data uit de jaren 1960, over de invloed van cultuur op de visuele waarneming (de kritiek van Prinz ging vooral over de waarneming van taal – de akoestische waarneming). Onderzoekers konden toen bijvoorbeeld aantonen dat de vatbaarheid voor visuele illusies erg variabel is over verschillende culturen en leeftijdsgroepen heen. Op de volgende slide vind je een grafiek die deze variabiliteit illustreert. Blijkbaar is het zo dat Westerlingen veel gemakkelijker ‘in de val trappen’ dan individuen uit andere culturen. Sommige populaties, zoals de San in de Kalahari-woestijn, zijn totaal ongevoelig voor de illusie. Zij zien de twee lijnen met andere woorden meteen als precies even lang. Daarnaast is het ook zo dat kinderen uit nagenoeg all culturen gevoeliger zijn voor de illusie dan volwassenen, zoals je op de grafiek kan zien. De Müller-Lyer illusie lijkt dus cultuur- en leeftijdsgebonden te zijn. Daaruit concludeert Henrich dat de visuele waarneming beïnvloed kan worden door externe informatie, bijvoorbeeld culturele informatie of persoonlijke ervaring – een conclusie die ingaat tegen Fodors stelling dat de waarneming ingekapseld is. Daarnaast toont de grafiek volgens Henrich ook aan dat de visuele waarneming niet eenvoudigweg aangeboren kan zijn, zoals Fodor beweert. Zoals ik daarnet al aanstipte meent Fodor dat modules zich ontwikkelen zonder invloed van de omgeving (op de trigger na, dan). Ook deze claim wordt weerlegd door de vaststelling dat de Müller-Lyer illusie zich vaker voordoet bij kinderen dan bij volwassenen.

MAAR: Henrich heeft het uitsluitend over de invloed van externe informatie op de waarneming van het individu *doorheen de tijd* (diachroon), niet op een bepaald ogenblik (synchroon): ‘Nothing about any of the findings we have discussed’, zegt Henrich, ‘establishes the synchronic cognitive penetrability of the Müller-Lyer stimuli’. ‘Cognitive impenetrability’ is min of meer een synoniem voor ‘ingekapseldheid’.

Contra argumenten tegen b)

Een andere claim die fodor maakt is het feit dat sommige systemen of mentale vermogens – zoals denken – niet modulair zijn. Omtrent deze claim geven wij opnieuw een tegenargument.

Fodors eerste claim is dat taal en waarneming modulaire mentale systemen zijn. Ik besprak zonet twee argumenten tegen deze stelling. Geen van deze tegenargumenten waren bijzonder overtuigend. Laten we even kijken naar Fodors tweede claim: denken, fantaseren, problemen oplossen, etc zijn niet-modulaire mentale systemen. De inzet van deze claim is behoorlijk groot. Als denken inderdaad niet-modulair is, zegt Fodor, dan kan het niet bestudeerd worden door de psychologie, noch door de neurologie. Een wetenschap van het denken is dan eenvoudigweg onmogelijk – niet meteen een opbeurende gedachte voor hersenwetenschappers, cognitieve wetenschappers en filosofen. Fodor spreekt zelf over een ‘mysterie’ (zie eerder slide met vergelijkende tabel).

Waarom meent Fodor dat het denken niet-modulair is? In een eerdere les probeerde ik al een antwoord te geven op de vraag of dieren kunnen denken. Davidson meende van niet, omdat een overtuiging altijd een groot web van andere overtuigingen veronderstelt. Eén overtuiging is geen overtuiging, redeneert Davidson, en het lijkt weinig plausibel dat dieren in staat zijn om zo’n web van overtuigingen te onderhouden. Fodor doet hier beroep op een gelijkaardige redenering om aan te tonen dat denken niet modulair kan zijn. Modules zijn niet in staat om informatie uit te wisselen – noch met elkaar, noch met niet-modulaire (centrale) systemen. Denken daarentegen bestaat maar bij gratie van een *web of beliefs* – een web waarin voortdurend informatie uitgewisseld wordt, zowel tussen overtuigingen onderling als tussen overtuigingen en zintuiglijke systemen. Men spreekt in deze context vaak van ‘holisme’ – een term die echter zo uitgehold is, onder andere door allerlei new-age adepten, dat hij nagenoeg betekenisloos geworden is. De ‘centrale systemen’ van onze geest zijn holistisch omdat ze gebruik maken van allerlei informatie die in het web van overtuigingen te vinden is. Zoals Steven Mithen zegt: ‘The fundamental character of cognition is that it is holistic, the exact opposite of the input systems which are all dedicated to dealing with one specific type of information alone. And this is what Fodor sees as the most puzzling feature of cognition: “its non encapsulation, its creativity, its holism and its passion for the analogical”. Fodor feels defeated by the central systems, declaring they are impossible to study. For him, “thought”, “problem solving”, “imagination” and “intelligence” are unresolvable’ (1996, p. 39).

De enige uitweg uit Fodors pessimisme is door te ontkennen dat modules ingekapseld zijn, of ten minste dat alle modules even sterk ingekapseld zijn. Je zou kunnen argumenteren dat sommige modules ‘zacht’ ingekapseld zijn, dat wil zeggen dat ze tot op zekere hoogte wel beïnvloed kunnen worden door informatie van buitenaf. Het denken zou dan wel volledig modulair kunnen zijn, waardoor het opnieuw een vruchtbaar voorwerp van wetenschappelijke studie kan worden.

* 1. **Massieve modulariteit**

Fodor is de bekendste vertegenwoordiger dat de geest gedeeltelijk modulair is – ook al suggereert de titel van zijn boek uit 1983 dat de geest volledig modulair is. Na Fodor zijn er een aantal wetenschappers en filosofen op de voorgrond getreden die Fodors bescheidenheid niet deelden, en die geprobeerd hebben om aan te tonen dat de hele geest modulair is of, anders gezegd: dat de menselijke geest massief modulair is. ‘Massief’ is dan min of meer een synoniem voor ‘volledig’. Deze stelling werd in de jaren 1990 met veel lawaai verdedigd door de zogenaamde evolutiepsychologen, zoals John Tooby, Leda Cosmides, en Steven Pinker; en vervolgens in het begin van de 21ste eeuw door de filosoof Peter Carruthers.

Samuels neemt de tijd om na te gaan of zulke claims gebaseerd zijn op een minimale of een robuste invulling van de term ‘module’ (zie eerdere slide) (vooral p. 41 bovenaan). Volgens hem is de MM-theorie enkel een interessante (‘interesting’) of nieuwe (‘distinctive’) theorie wanneer ze de stelling verdedigd dat de geest volledig opgebouwd is uit robuuste modules, dat wil zeggen modules met de kenmerken die Fodor oplijstte. Carruthers’ claim is dat de geest volledig opgebouwd is uit minimale modules, en wordt door Samuels door de mangel gehaald: ‘Call this a version of massive modularity if you like. But it surely isn’t an interesting or distinctive one’ (p. 41). Zie volgende slide voor citaat.

De evolutiepsychologen blazen daarentegen warm en koud, maar Samuels beschouwt hen niettemin als vertegenwoordigers van de theorie dat de geest volledig is opgebouwd uit robuuste modules. Hij verwijst bijvoorbeeld naar Pinkers theorie over de manier waarop we de levende wereld interpreteren – de zogenaamde *folk biology* die in een eerdere les al aan bod kwam. Net als redeneren en beslissen wordt classificeren (de levende natuur onderverdelen in groepen en subgroepen) vaak beschouwd als een voorbeeld van een ‘centraal proces’. En volgens Pinker is folk *biology* modulair. We hebben van onze voorouders een module geërfd om de levende natuur te classificeren.

In *The Architecture of the Mind* legt Carruthers uit dat hij de notie ‘modulariteit’ niet op dezelfde manier begrijp als Fodor. Wat betekent ‘modulair’ dan wanneer we spreken over ‘massieve modulariteit’? Het zijn nog steeds a) isoleerbare (of dissocieerbare) mentale systemen, die elk een specifieke functie vervullen, en die b) domein-specifiek zijn, c) automatisch (of onwillekeurig), d) lokaliseerbaar en e) ontoegankelijk. Fodor onderscheidde negen verschillende kenmerken van modules; Carruthers houdt er daar vijf van over. De meest opmerkelijke afwezige is uiteraard het kenmerk van ingekapseldheid. Volgens Carruthers zijn niet alle modules ingekapseld

We zullen in het vervolg van dit hoofdstuk concentreren op evolutiepsychologen. Samuels neemt in zijn tekst aan dat zij claimen een robuuste invulling te hebben van modulariteit en daarbij massieve modulariteit prediken.

Evolutiepsychologie

In wat volgt concentreer ik – zoals reeds gezegd – me vooral op de argumenten die aangebracht werden door de evolutiepsychologie om de stelling te staven dat de menselijke geest massief modulair is. Die stelling volgt, althans volgens evolutiepsychologen, uit een andere stelling, met name dat de menselijke geest een product is van natuurlijke selectie. In een notendop, de volledige redenering. We bekijken elk van de premissen van naderbij.

* + Menselijke geest = product van natuurlijke selectie
  + Overleven = adaptieve problemen oplossen
  + Adaptieve problemen oplossen vereist modulariteit
  + Dus: menselijke geest is massief modulair

1ste premisse

Evolutiepsychologen beweren dus dat onze geest een van de vele verwezenlijkingen is van het proces van natuurlijke selectie. En zoals vele hedendaagse darwinisten beweren ze dat die claim helemaal niet nieuw is, maar integendeel al verdedigd werd door de godfather van het darwinisme: Darwin zelf. En daar is iets voor te zeggen. In zijn eerste werk, *On the Origin of Species* (1859) was Darwin zelf niet zo happig om ‘zijn’ evolutietheorie toe te passen op de mens. Hij sprak voornamelijk over de effecten van natuurlijke selectie op plantsoorten en niet-menselijke diersoorten. In latere werken, zoals *The Descent of Man* (1871) en *The Expression of Emotions in Man and Animals* (1872) geeft hij die terughoudendheid op. Aan het eind van *On the Origin*, in een van de laatste paragrafen, kondigt hij al een en ander aan. Hij schetst er het volgende toekomstvisioen:

In vertaling: ‘In de verre toekomst zie ik open velden voor veel belangrijker onderzoek. De psychologie zal op een nieuw fundament geplaatst worden, namelijk die van de noodzakelijke verwerving van elk mentaal vermogen en aanleg door evolutie. Er zal licht worden geworpen op de oorsprong van de mens en zijn geschiedenis’. Darwin zelf meende dus dat elk van de vermogens van de menselijke geest ‘ontworpen’ werd door natuurlijke selectie. En hij besefte dat die vaststelling de psychologie op z’n kop zou zetten, of ten minste ‘een geheel nieuw fundament’ zou bezorgen. Ter informatie: Alfred Russell Wallace, de man die min of meer gelijktijdig met Darwin het principe van natuurlijke selectie op papier zette, was het op dit punt oneens met Darwin. Wallace was in alle opzichten radicaler dan Darwin wat betreft de kracht van natuurlijke selectie. Terwijl Darwin oog had voor andere evolutionaire mechanismen dan natuurlijke selectie, en de beperkingen van natuurlijke selectie, bezong Wallace de oneindige mogelijkheden ervan. Vreemd genoeg was Wallace wel van mening dat de menselijke geest onmogelijk het product kan zijn van natuurlijke selectie. De menselijke geest is immers te complex om verklaard te worden met behulp van een in wezen dom en mechanisch proces van vallen en opstaan.

Wat bedoelt men met ‘natuurlijke selectie’? Ik fris even uw geheugen op.

Darwin was niet de eerste om te beweren dat soorten evolueren, maar hij was wel de eerste om een mechanisme aan te duiden dat die evolutie aandrijft: natuurlijke selectie. Het principe van natuurlijke selectie is zonder twijfel Darwins belangrijkste bijdrage geweest aan de intellectuele geschiedenis. Eigenlijk is het een erg eenvoudig principe. Men stelt het vaak voor als de conclusie van een logische redenering waarvan de drie basisstellingen aantoonbaar waar zijn. De drie premissen zijn:

(a) *Er is schaarste.* In principe zou eender welke biologische soort (planten, dieren of mensen) zich exponentieel kunnen voortplanten in nauwelijks een tiental generaties. Zulke bevolkingsexplosies komen echter zelden voor, omdat de ‘wonderbaarlijke vermenigvuldiging’ vroeg of laat toch op schaarste stuit – voedselschaarste of plaatsgebrek. Deze schaarste impliceert dat niet alle organismen in staat zijn om zich voort te planten.

(b) *Er is variatie.* Het is onmiskenbaar dat er een enorme variatie is in de levende wereld. Geen twee exemplaren van een soort zijn identiek. Krachtens deze variatie zullen sommige organismen beter overweg kunnen met de gegeven (natuurlijke of sociale) omstandigheden, waardoor ze ‘een stapje voor’ hebben op andere organismen, en dus meer nakomelingen kunnen voortbrengen. Met andere woorden: de variatie bepaalt welke organismen zich zullen voortplanten en welke niet.

(c) *Variatie is erfelijk.* Men kan niet ontkennen dat nakomelingen in de regel sterk op hun biologische ouders lijken. Daarom kunnen we aannemen dat die eigenschappen die de ouders zo aangepast maken aan hun omgeving, ook zullen doorgegeven worden aan de volgende generatie. Darwin wist zelf niet precies welke eigenschappen erfelijk waren en welke niet, en hoe die erfelijkheid precies in z’n werk ging (de ontdekking van de DNA-helix dateert van 1953), maar zijn observaties leerden hem wel dat er zoiets bestond als erfelijkheid.

Darwins conclusie luidt dat er een ‘onzichtbare hand’ moet bestaan die nuttige eigenschappen behoudt en doorspeelt via de erfelijkheid. Hij noemt het principe van natuurlijke selectie (de ‘onzichtbare hand’) dan ook een ‘principe van behoud’. Wat behouden wordt door natuurlijke selectie zijn variaties die nuttig zijn voor het organismen in de ‘strijd’ om te overleven en zich voort te planten. Wanneer zo’n variatie zich doorheen een aantal generaties over de populatie verspreid heeft, spreekt men van een adaptatie. Een adaptatie is een eigenschap van een organisme dat in een bepaalde omgeving een bepaalde functie vervult waardoor dat organisme een groter aantal levensvatbare (en zich voortplantende) nakomelingen kan voortbrengen dan organismen die die eigenschap *niet hebben.*

Drie belangrijke implicaties van Darwins theorie

Darwin was er dus van overtuigd dat zijn theorie, evolutie door natuurlijke selectie, een nieuw licht zou kunnen werpen op de menselijke geest. Maar wat betekent het concreet dat onze geest het product zou zijn van natuurlijke selectie? Deze claim heeft verschillende interessante implicaties, waaronder

de stelling dat onze geest een stapsgewijze ‘verfijning’ is van eerdere versies van het ding (gradualisme) – versies die we kunnen terugvinden bij verwante ‘lagere’ diersoorten. Uit deze stelling kan men allerlei voorspellingen afleiden over de cognitieve continuïteit tussen de mens en de andere dieren. Denk maar aan die voortdurende stroom van berichten in de wetenschapskaternen van dagbladen en weekbladen, waarin onderzoek gepresenteerd wordt waaruit blijkt dat deze of gene ‘uniek menselijke’ cognitieve eigenschap toch niet zo uniek blijkt te zijn in de dierenwereld.

Een andere implicatie van Darwins claim is de stelling dat de menselijke geest samenvalt met de hersenen – de basisstelling van het filosofisch materialisme (of fysicalisme). Natuurlijke selectie kan enkel psychologische eigenschappen selecteren voor zover die voortvloeien uit een materieel substraat – de hersenen. Deze implicatie verklaart veel van de katholieke weerstand tegen het darwinisme. In katholieke kringen kan men zich misschien nog net verzoenen met de gedachte dat ons lichaam niet ontworpen is door een schepper, maar zeggen dat ook de geest een product is van natuurlijke selectie is een brug te ver. Zo staat er in een pauselijke mededeling uit 1996: ‘Consequently, theories of evolution which, in accordance with the philosophies inspiring them, consider the mind as emerging from the forces of living matter, or as a mere epiphenomenon of this matter, are incompatible with the truth about man’. De bezorgdheid van de paus heeft uiteraard te maken met zijn geloof in de onsterfelijkheid van de ziel – een geloof dat moeilijk te rijmen valt met het materialisme van de evolutietheorie.

Een derde en laatste implicatie van Darwins claim is dat de verschillende onderdelen van de geest, net zoals de verschillende onderdelen van het lichaam, moeten begrepen worden als adaptaties, in die zin dat ze een antwoord bieden op een of ander probleem waarmee onze voorouders in hun omgeving geconfronteerd werden. Onze geest moet met andere woorden begrepen worden als een mentale gereedschapskist, die ons in staat stelt om allerlei cognitieve (en emotionele) klusjes te klaren. Welke klusjes, hoor ik u dan vragen?

Hier zijn enkele voorbeelden. Allerlei wetenschappen, zoals de archeologie en de biologische antropologie, geven ons een redelijk gedetailleerd beeld van de natuurlijke en sociale omgeving van onze voorouders. *Natuurlijke* omgeving of habitat: dieren (roofdieren & prooidieren), planten, temperatuur,… Het succes van onze voorouders heeft volgens evolutiepsychologen te maken met het feit dat hun geest hen in staat stelde om, onder andere, natuurlijke gevaren goed in te schatten (welke champignons kan je eten en welke niet?), en allerlei werktuigen te bedenken en te fabriceren. De *sociale* omgeving van onze voorouders was erg vergelijkbaar met de sociale wereld van die enkele *small-scale societies* (die vroeger wel eens ‘primitieve samenlevingen’ genoemd werden) die vandaag nog bestaan. Zoals de San, die eerder aan bod kwamen in het kader van de Müller-Lyer illusie. Zulke samenlevingen leven van jagen en verzamelen, vandaar dat ze vaak *hunter-gatherer societies* genoemd worden. Ze bestonden uit kleine groepen van ongeveer vijftig personen, waarin familiebanden centraal stonden. Deze sociale structuur bood, net als hun natuurlijke omgeving, allerlei uitdagingen. Het succes van onze voorouders heeft te maken met het feit dat ze mentale vermogens ‘ontwikkelden’ (het is moeilijk om zulke actieve metaforen te vermijden, maar in werkelijkheid is het proces van natuurlijke selectie geen proces van ontwerpen – het is eerder een proces waarin toevalstreffers stapsgewijs en tergend traag uitgroeien tot wijdverspreide adaptaties) om deze uitdagingen het hoofd te bieden. Aldus verwierven onze voorouders de mentale vermogens om goede partners te kiezen (dat wil zeggen: partners waarmee ze een stevig nageslacht konden uitbouwen), om gedachten en emoties te herkennen bij zichzelf en anderen (theory of mind), gezichten te herkennen van familieleden en leden van hun sociale groep, enzovoort.

Tooby en Cosmides sommen nog een aantal (hypothetische) modules op: ‘A face recognition module, a spatial relations module, a rigid objects mechanics module, a tool-use module, a fear module, a social-exchange module, an emotion-perception module, a kin oriented motivation module, an effort allocation and recalibration module, a child care module, a social inference module, a friendship module, a semantic-inference module, a friendship module, a grammar acquisition module, a communication-pragmatics module, a theory of mind module, and so on’ (geciteerd in Mithen 1996, 45).

daarentegen niet vanzelfsprekend. Ik formuleer een uitgebreide versie van de premisse: natuurlijke selectie geeft de voorkeur aan een modulaire geest omdat adaptieve problemen (waarvan ik er enkele opsomde in de vorige slide) sneller, efficiënter en betrouwbaarder opgelost kunnen worden met behulp van modulaire systemen dan met behulp van niet-modulaire systemen.

Is deze premisse plausibel? Volgens Samuels niet, maar daar kom ik dadelijk op terug. Laten we even twee argumenten bekijken die evolutiepsychologen in deze context vaak aandragen om hun zaak te bepleiten. Het eerste argument staat bekend als het ‘poverty of stimulus’-argument; het tweede argument als het ‘frame problem’.

Poverty of stimulus (aangedragen door evolutiepsychologen)

Het eerste argument luidt dat jonge kinderen zo snel greep krijgen op zoveel complexe problemen dat we eigenlijk moeten veronderstellen dat hun geest voorgeprogrammeerd is om specifieke taken te vervullen. Denk maar aan het gemak waarmee kinderen een taal leren. Hoe is het mogelijk dat ze met zo’n rotvaart zoveel complexe semantische en syntactische regeltjes leren, gewoon door naar het gemompel van hun ouders

te luisteren? Noam Chomsky noemde dit argument het ‘poverty of stimulus’-argument; er is weinig of geen stimulans nodig, zegt hij, om kinderen een taal te doen leren. Het proces verloopt automatisch, alsof het geprogrammeerd is. Evolutiepsychologen veronderstellen dan ook dat onze geest bij de geboorte uitgerust is met een ‘taalverwervingsmodule’ (Chomsky sprak van een genetisch geprogrammeerd ‘language acquisition device’) die ons in staat stelt om zonder veel moeite eender welke taal aan te leren. Deze redenering kan je vervolgens toepassen op eender welk domein in het leven, gaande van het herkennen van gezichten tot het vermijden van giftige planten.

Frame problem ( aangedragen door evolutiepsychologen om hun claim te onderbouwen)

Het tweede argument dat evolutiepsychologen aanhalen om hun geloof in massieve modulariteit te onderbouwen staat bekend als het ‘frame problem’. Het is een argument dat Fodor zelf al gebruikte om uit te leggen waarom er ‘stupid input systems’ bestaan. Het gaat over het nemen van belangrijke beslissingen. Stel dat u bij een wandeling een leeuw ontmoet. Wat doet u? Er zijn twee opties. Een eerste optie is dat u de zaken even rustig op een rijtje zet, en dat u al de mogelijkheden nagaat die u op dat moment ter beschikking staan. Het zou bijvoorbeeld kunnen dat deze leeuw rechtstreeks uit de Afrikaanse savanne komt, en honger gekregen heeft van de vliegreis. Maar misschien is het een oude leeuw, die op zoek is naar gezelschap? Misschien gaat het om een onschuldig circus-exemplaar, en hoeft u dus helemaal niets te vrezen? Of misschien... De kans is groot dat u tijdens deze overpeinzingen opgepeuzeld wordt. Een tweede optie is dat u gehoorzaamt aan een instinctieve drang om ofwel te verstijven (in de hoop dat het beest een beetje bijziend is, en u voor een boom houdt) ofwel te vluchten (of aan te vallen, maar dan moet u wel een speer bij de hand hebben). De eerste optie is gebaseerd op de veronderstelling dat onze geest verschillende problemen met eenzelfde vaste set van procedures oplost. De tweede, evolutiepsychologische optie gaat er vanuit dat onze geest een gespecialiseerde module bevat die enkel informatie over wilde dieren verwerkt, en die verbonden is met een gedragspatroon dat in de psychologische literatuur bekend staat als de ‘fright, flight or fight response’ (verstijf-, vlucht- of vecht-module) – een module die volgens evolutiepsychologen universeel is. De kans is groot, zo redeneren evolutiepsychologen, dat voorouders die uitgerust waren met een mentale *passe-partout* – een algemene set van procedures die bij heel uiteenlopende problemen ingezet wordt – snel weggeconcurreerd werden door ‘modulaire’ voorouders, die de relevante informatie veel sneller en efficiënter verwerken.

Een ander voorbeeld van hetzelfde argument: ‘The problem with general-purpose learning rules, according to Cosmides & Tooby, is that there are no bounds as to what information should be excluded from making a decision, and as to which alternative courses of action should be ignored. Every single possibility should be examined. Our pre-historic ancestors would have quietly starved as they tried to decide where and what to hunt. But if one of the hunters had a specialized mental module for making hunting decisions, which prescribed the types of information to consider and how to process them, he would have prospered. This would, no doubt, have increased his reproductive success, and soon the community would be populated with his offspring, each with this specialized mental module for making hunting decisions’ (Mithen 1996, 44).

Tegenargumenten voor evolutionaire psychologie

Samuels bespreekt nog een aantal andere (vooral niet-evolutionaire) argumenten die in het voordeel spreken van de stelling dat het oplossen van adaptieve problemen een massief modulaire geest vereist (de derde premisse uit de redenering van evolutiepsychologen, zie eerdere slide). Daarbij toont hij telkens aan hoe het argument er niet in slaagt de betreffende premisse te ondersteunen. Hij herinnert ons namelijk telkens aan de robuuste invulling van het concept van modulariteit, waarin de kenmerken ‘ingekapseld’ en ‘domein-specifiek’ een centrale rol spelen, en stelt vast dat evolutiepsychologen (en andere vertegenwoordigers van MM) vooral argumenten aandragen die bepaalde kenmerken van modulariteit ondersteunen, zoals dissocieerbaarheid (eerste en tweede argument in de tekst van Samuels, p. 42-43), maar niet de centrale kenmerken. En als ze toch de noodzaak van zulke kenmerken postuleren, dan slaagt Samuels erin om aan te tonen dat ze zich vergissen (laatste argument, p. 45). Het lijkt er met andere woorden op dat wanneer evolutiepsychologen claimen dat de geest massief modulair is, dat ze het woord ‘modulair’ op een minimale manier invullen, waardoor ze eigenlijk een theorie verdedigen die niet interessant noch nieuw is.

Je kan Samuels hakwerk aanvullen door de twee argumenten die wij zonet besproken hebben, namelijk het ‘poverty of stimulus’-argument en het ‘frame-problem’, op eenzelfde manier te fileren. Want wat toont het eerste argument eigenlijk aan? Het toont aan dat bepaalde mentale vermogens, zoals het vermogen om een taal te leren, onbegrijpelijk zijn wanneer we er niet vanuit gaan dat ze in hoge mate aangeboren zijn. ‘Aangeboren’ is uiteraard een kenmerk van modulariteit, ook in Fodors verhaal, maar het is bezwaarlijk een centraal kenmerk te noemen. Het tweede argument is in hetzelfde bedje ziek. Het toont hoogstens aan dat bepaalde (niet alle!) onderdelen van de geest van onze voorouders baat hadden bij snelle en efficiënte informatieverwerkingsmechanismen (hottentottententententoonstelling!). En zoals ik eerder al opmerkte suggereert het bestaan van snelle mentale systemen weliswaar dat er sprake is van ‘ingekapseldheid’ (encapsulation), maar het is ook niet meer dan een suggestie. Deze kritiek is min of meer analoog aan de kritiek van Samuels op het argument van ‘computational tractability’ (p. 45). Het argument stelt dat een goede werking van de menselijke geest enkel haalbaar is (‘tractable’) wanneer die geest volledig is opgebouwd uit ingekapselde modules – modules die geen tijd en energie verliezen door allerlei bronnen van informatie te raadplegen bij het verwerken van hun informatie. Samuels weerlegt dit argument door te zeggen dat haalbaarheid gerealiseerd kan worden door ‘ingekapseldheid’, maar ook door zuinigheid (‘frugality’). ‘Rather, what’s required [to engender tractability] is that the mechanism be frugal in its use of information: that it not engage in exhaustive search but only use a restricted amount of the available information’ (p. 45). Het argument toont met andere woorden niet aan dat mentale systemen robuust modulair zijn, laat staan dat alle mentale systemen robuust modulair zijn.

Hoe dan ook, het zijn niet alleen de theoretische argumenten in het voordeel van MM die falen, ook de empirische argumenten zijn volgens Samuels weinig overtuigend. Nogmaals: er is heel wat empirisch bewijs voor het bestaan van modulaire ‘input systems’, zoals taal en waarneming. Maar het gaat hier om bewijs voor de stelling dat ook centrale systemen modulair zijn – systemen zoals redeneren, beslissen, enzovoort. Naast deze onrechtstreekse aanval (het falen van0 argumenten pro), is er ten slotte ook rechtstreekse kritiek (de kracht van argumenten contra). Samuels bespreekt enkele van deze argumenten in het nadeel van MM – argumenten die aantonen dat heel wat centrale systemen niet-modulair zijn (sectie 4 van de tekst).

De vraag of de geest modulair is heeft enkele vragen voorafgaand die het antwoord mee bepalen. Met name wat we bedoelen met module, wat we bedoelen met de geest. We bekomen na deze uiteenzetting een vrij genuanceerd antwoord: ‘bepaalde delen van de menselijke geest zijn, in bepaalde levensstadia, tot op zekere hoogte modulair’

1. Psychophatologie

De vraag die ons bezighoudt sinds het begin van het semester is hoe lichaam/hersenen en geest zich tot elkaar verhouden. En bij uitbreiding: hoe ‘harde’ wetenschappen, zoals neurobiologie en fysica, en ‘zachte’ wetenschappen, zoals psychologie, zich tot elkaar verhouden. Nu zijn er mensen die dagelijks met deze vraag worstelen: psychiaters.

psychiaters hebben precies hetzelfde vertrekprobleem als philosophers of mind. Net zoals deze laatste een bepaling moeten geven van wat ze bedoelen met ‘the mind’ of ‘mental states’ (zie onze pogingen in de eerste les), moeten psychiaters een gemene deler zoeken in die wilde veelheid van fenomenen die ze ‘mental disorder’ noemen. Ze moeten met andere woorden een goede definitie geven van wat psychopathologie precies is. En die opdracht is eigen tweevoudig.

Enerzijds moeten psychiaters uitleggen waarin precies het verschil ligt tussen psychopathologie en ‘gewone’ levensproblemen. Een goed voorbeeld is het verschil tussen rouw en depressie. Rouw is een uiterst pijnlijke mentale toestand die allerlei treffende gelijkenissen vertoont met het ziektebeeld van een klinische depressie (in het Engels: *Major Depressive Disorder*). Niettemin beschouwen we rouw niet als ziekte – het is eerder een normale reactie op een ingrijpende tegenslag in het leven, zoals het verlies van een geliefde persoon. Een depressie kan optreden wanneer het rouwproces ontspoort. Maar op basis waarvan bepalen we of een normale reactie verstoord wordt? Anders gezegd: hoe beslissen we over het verschil tussen ziekte en gezondheid?

Anderzijds moeten psychiaters ook uitleggen waarin precies het verschil ligt tussen psychische stoornissen (psychopathologie) en niet-psychische stoornissen, zoals neurologische stoornissen of gewone medische aandoeningen. De vraag is dan niet zozeer hoe we ‘stoornis’ definiëren, maar wel hoe we ‘mentaal’ definiëren. Waarom noemen we geestesziekten ziekten van de geest, en niet van de hersenen?

In deze les zoeken we antwoorden op de eerste vraag: hoe onderscheiden we levensproblemen van geestesziekten? De tweede vraag beantwoorden we in de volgende les, aan de hand van de tekst van Brülde & Radovic.

*More than any other professionals on earth, psychiatrists « live » the mind-body problem in their day-to-day work. Any active psychiatric clinician who uses both pharmacological and psychological treatment methods is constantly switching perspectives, seeing the dysfunction of the patients before them as 1) a problem in brain functioning that they are seeking to treat with drugs and 2) as a result of psychological dysfunctions in a « minded » individual, who is full of hopes, wishes, fears, and desires that they hope to help with psychotherapy ’ (Kendler 2008, 9)*

* 1. **Wat maakt een psychische stoornis tot stoornis (metafysisch)**

Voorbeeld van deze vraag: hoe onderscheiden we ‘normale toestanden’ van ‘stoornissen’? In DSM-II (1968) staat homoseksualiteit in de lijst van seksuele afwijkingen, geheel in lijn met Freuds denken over homoseksualiteit (zie bijvoorbeeld diens Drei Abhandlungen zur Sexualtheorie uit 1905). En seksuele afwijkingen vormen een subcategorie van de neurosen. De American Psychiatric Association beschouwde homoseksualiteit lange tijd als een psychische stoornis. Door intensief lobbywerk van homoseksuele verenigingen verdween het in 1974 uit het handboek (DSM). Besefte men dan pas dat het fenomeen ‘homoseksualiteit’ eigenlijk niet in het door DSM gehanteerde concept van stoornis paste? Of was er een wijziging in het concept in kwestie? Je vindt er meer informatie over in Bayer, R., Homosexuality and American Psychiatry: The Politics of Diagnosis (Princeton: Princeton University Press, 1981).

Mijn punt is dat de discussie over de plaats van homoseksualiteit in DSM een goed voorbeeld is van een eerste interpretatie van de vraag ‘wat is psychopathologie?’ Je kan je immers afvragen waarom we homoseksualiteit als een stoornis zouden beschouwen, eerder dan als een ‘normale’ seksuele variant. Een ander voorbeeld van deze interpretatie van de vraag ‘wat is psychopathologie?’ is rouw. De ‘symptomen’ van iemand die in rouw is, lijken erg op de symptomen van een klinische depressie, en toch noemen we rouw geen stoornis. Het is eerder een pijnlijke mentale toestand die niettemin normaal is - een normale reactie op een tegenslag in het leven. Op basis waarvan kunnen we rouw en depressie dan onderscheiden? Waarom is rouw normaal en depressie een stoornis?

Verwarring

In wat volgt beschouw ik ‘geestesziekte’, ‘psychopathologie’, ‘psychische stoornis’, ‘mentale stoornis’ en zelfs ‘psychiatrische stoornis’ en ‘psychiatrische aandoening’ als synoniemen.

*‘The double duty done by the term « psychopathology » in customary usage. It is used to designate both mental disorders themselves (…) and also the subfield of medical practice and psychology that studies such disorders’ (Graham & Stephens 1994, 3-4).*

Men ziet dus dat met de term psychopathologie zowel de geestesziektes als de subcategorie van medische praktijken en psychologie die zulke stoornissen bestuderen. In wat volgt zal de nadruk vooral op het eerste liggen.

Filosofische psychopathologie

*‘Under the heading of philosophical psychopathology, we include both investigations of philosophical issues arising from consideration of various sorts of mental disorders and investigations of issues raised by theoretical and clinical studies of such disorders’ (Graham & Stephens 1994, 3-4).*

beschouw de term ‘filosofische psychopathologie’ als een synoniem voor ‘filosofie van de psychiatrie’ – een term waarvan de betekenis mijns inziens duidelijker is. Filosofie van de psychiatrie is een van de vele takken van de wetenschapsfilosofie, zoals F van de fysica, F van de biologie, enzovoort, die zich buigt over filosofische problemen in de psychiatrie. De filosofie van de psychiatrie bestrijkt een aantal subdomeinen, die je op verschillende manieren kan opdelen.

1. Ten eerste is er een duidelijk verband tussen filosofie van de psychiatrie en *philosophy of mind*. In deze en volgende les zullen we vooral kennismaken met dit eerste subdomein.
2. Ten tweede is er ook een uitgebreide literatuur over wetenschapsfilosofie en psychiatrie. Daarin worden vragen gesteld die gedeeltelijk ook in het eerste subdomein aan bod komen, maar de klemtoon ligt toch vooral op de wetenschappelijke waarde van de psychiatrie. Wat betekent het om een geestesziekte te verklaren? Hoe moeten we de relatie denken tussen de verschillende stromingen in de psychiatrie (psychoanalyse, biologische psychiatrie, anti-psychiatrie,…)? Zijn psychische stoornissen natuurlijke soorten?
3. Ten derde en ten laatste tonen filosofen ook een interesse in ethische problemen in de psychiatrie. Men stelt dan vragen naar de toerekeningsvatbaarheid van patiënten, naar de impliciete waarden die psychiaters hanteren in het opstellen van ziekte-definities, en de relatie tussen de geneesmiddelenindustrie en de psychiatrie.





Waarom noemen we op de eerste afbeelding de uiteenlopende toestanden ziekten?

Waarom noemen we op de tweede afbeelding de stoornissen psychisch. In welke zin verschilt pedofilie van alzeheimer of zijn ze eenzelfde soort ziekte?

De vraag ‘wat is psychopathologie?’ kan je nog verder opsplitsen in twee verschillende filosofische vragen, net zoals de vraag ‘wat is persoonsidentiteit?’. Je kan de vraag enerzijds als een epistemologische vraag behandelen. De vraag is dan: hoe weten we dat een bepaalde mentale toestand een geestesziekte is? Waaraan herkennen we een geestesziekte? Aan welke criteria moet een persoon voldoen om in aanmerking te komen voor deze of gene diagnose? Het antwoord op deze lijst vind je in een lijst met diagnostische criteria. Anderzijds kan je de vraag ook als een metafysische vraag beschouwen. De vraag is dan: wat maakt een bepaalde toestand tot een geestesziekte? Wat is de eigen aard (‘de natuur’) van psychopathologie? Het antwoord op deze vraag is geen lijst met diagnostische criteria, maar een lijst met voldoende en noodzakelijke voorwaarden.

Enkele antwoorden op de vraag.

We zullen in het vervolg van deze paragraaf drie antwoorden geven om de metafysische vraag wat maakt een psychische stoornis tot een stoornis. De volgende drie antwoorden komen aan bod: harmful dysfunction analysis (wakefield), broken brain analysis (Andreasen en griesinger) en prototype resemblance analysis (Georg graham).

Eerste antwoord: Harmful dysfucntion analysis

Volgens Wakefield moet een bepaalde toestand aan twee voorwaarden voldoen om een psychische stoornis genoemd te worden. Deze twee voorwaarden zijn elk noodzakelijk en samen voldoende. De eerste voorwaarde is dat er sprake is van ‘harm’ – de patiënt of diens entourage moet de toestand (van de patiënt) in kwestie als pijnlijk of schadelijk ervaren, ofwel voor de patiënt zelf, ofwel voor de buitenwereld. De tweede voorwaarde is dat er sprake is van een disfunctie van een psychologisch mechanisme dat zelf een product is van natuurlijke selectie. Ik gaf vorige week het voorbeeld van depressie. Depressie is een stoornis omdat er sprake is van subjectief lijden (van de patiënt) en van een disfunctie van het psychologische mechanisme dat instaat voor verdriet en neerslachtigheid.

De Amerikaanse filosoof Jerome Wakefield stelt dus voor om het concept ‘geestesziekte’ te definiëren als ‘schadelijke disfunctie’ (*harmful dysfunction*). Dat wil zeggen dat X een geestesziekte is wanneer ze aan twee voorwaarden voldoet, die elk noodzakelijke voorwaarden zijn en samen voldoende voorwaarden zijn. Beide voorwaarden zijn volgens Wakefield kenmerkend voor *alle* ziekten, dus ook voor medische aandoeningen, en *enkel* voor ziekten. De voorwaarden bepalen met andere woorden de essentie van het concept van psychopathologie. In die zin zou je Wakefields benadering als een essentialistische benadering kunnen begrijpen.

Wat betekenen die voorwaarden nu precies? De eerste voorwaarde is dat X schadelijk (‘harmful’) is voor de drager van X (het zieke individu), of als schadelijk beschouwd wordt door diens omgeving. Anders geformuleerd: de drager van X moet een zeker lijden vertonen. (Je kan ‘schadelijkheid’ natuurlijk op verschillende manieren interpreteren, maar dat doet nu even niet ter zake.) De tweede voorwaarde is dat X gebaseerd is op een biologische disfunctie (‘dysfunction’). Dat wil zeggen dat de natuurlijke functie van een orgaan of een mechanisme verstoord moet zijn – denk maar aan nieren die geen afvalstoffen meer verwerken, een hart dat niet regelmatig genoeg pompt, of modules die geen informatie meer kunnen verwerken. Wakefield begrijpt de term ‘disfunctie’ op een evolutionaire manier (er zijn ook andere definities van het functie-begrip, maar die zijn hier niet van tel). Wanneer een lichaamsorgaan of een mentale module niet meer functioneren, dan komt het reproductief succes van het individu in het gedrang. En reproductief succes wordt in de evolutiebiologie afgemeten aan het aantal levensvatbare nakomelingen dat iemand kan voortbrengen. In het geval van psychische stoornissen gaat het om disfuncties van psychologische mechanismen, zegt Wakefield. Zijn antwoord op de vraag ‘wat is psychopathologie’ luidt dus: ‘a harmful failure of a mental mechanism to perform a natural function for which it was designed by evolution’. Wakefield hoopt dat de evolutiepsychologie (zie de les over modulariteit) ons een lijst zal bezorgen van alle psychische mechanismen (of modules) waarmee natuurlijke selectie ons behept heeft.

Intermezzo

Vooralleer we Jerome wakefields theorie illustreren met enkele voorbeelden maken we een kleine omweg.

Volgens evolutiepsychologen is ons brein een zakmens dat dient om adaptieve problemen op te lossen (zie hoofstuk 6). Ze stellen dat bepaalde mechanisme dienen om die adaptieve problemen op te lossen en ons zo meer kans te geven om te overleven en meer nakomelingen te bezorgen. Verdriet en angst blijken ook zulke mechanisme te zijn, maar waar liggen hun voordelen (indien ze geen voordelen hadden, dan hadden we ze al lang niet meer gehad.) wat volgt zijn drie verklaringen.

Eerste verklaring is die van Jhon Scott Price. Hij stelt dat neerslachtigheid als voordeel heeft dat het gebruikt kan worden als Time-out. Hypothese: Neerslachtigheid is een geëvolueerde strategie die ‘losers’ toelaat om competitie (tijdelijk) stop te zetten, verlies te aanvaarden, en onderwerping te signaleren

Tweede verklaring is die van Paul J. Watson. Hij stelt dat het voordeel van neerslachtigheid is dat we het als zelf-evaluatie kunnen gebruiken. Hypothese: N stelt ons in staat om (onhaalbare) investeringen en doelstellingen opnieuw te evalueren, en belet ons om daarbij overhaast te werk te gaan.

Derde en laatste verklaring komt van Edward H. Hagen. Hij stelt dat neerslachtigheid dient als noodkreet. N is een kostbaar alarmsignaal, waarmee iemand aangeeft dat hij/zij belangrijke verliezen lijdt, en dat naasten ertoe aanzet om hulp te bieden.

Duidelijk moet zijn (via de qoutes op de ppt slides die niet in deze tekst staan) dat evolutiepsychologen neerslachtigheid en depressie door elkaar halen. Jerome Wakefield stelt dat neerslachtigheid en depressie verschillend zijn. Depressie is de disfunctie van het psychologische mechanisme neerslachtigheid.

Einde intermezzo

Voorbeelden ter illustratie van Jerome Wakefields theorie.

*Depressie*. Depressie is al enkele jaren aan een heuse blitzkrieg bezig. De Westerse wereld lijkt zowaar geteisterd te worden door een epidemie van depressieve stoornissen – een evolutie die op het eerste gezicht moeilijk te rijmen valt met de toenemende professionalisering van de geestelijke gezondheidszorg, en de ontwikkeling van steeds nieuwe generaties van psychofarmaca. Een van de ijkpunten in het debat over de depressie-epidemie is *The Loss of Sadness* (2007), waarin socioloog Allan Horwitz en filosoof Jerome Wakefield argumenteren dat de vermeende ‘epidemie’ van depressieve stoornissen een artefact is van de hedendaagse psychiatrische diagnostiek. In hun ogen is depressie wel degelijk een stoornis in de medische zin van het woord. Het probleem is dat de ziekte overgediagnosticeerd wordt. Dat wil zeggen: de diagnose ‘depressie’ wordt ook toegeschreven aan mensen wiens angst en neerslachtigheid een natuurlijke (en ‘gezonde’) reactie is op de onvermijdelijke tegenslagen in het leven (‘levensproblemen’). Daarmee gaan Horwitz en Wakefield lijnrecht in tegen diegenen die beweren dat depressieve stoornissen vandaag de dag vaker voorkomen dan vroeger, en dat de psychiatrie bovendien verfijndere technieken ontwikkeld heeft om zulke stoornissen te identificeren en te behandelen.

Maar even terug naar de les: waarom is een ‘echte’ depressie volgens Wakefield een goed voorbeeld van een psychische stoornis? Enerzijds is het duidelijk dat er sprake is van ‘schade’ – de eerste component van zijn analyse. Depressieve patiënten klagen over een gebrek aan eetlust, aan fut, aan levenslust. Ze kunnen niet meer genieten, rumineren voortdurend over het verleden, en zien de toekomst donker en dreigend. Ze verwaarlozen zichzelf en hun familie, hun werk, en komen daarom vaak in een negatieve spiraal terecht, waarin het leven zelf door hun vingers dreigt te glippen. Patiënten ervaren hun eigen toestand als pijnlijk, en vragen meestal uitdrukkelijk om hulp, samen met hun naaste familie of vrienden. De *harm*-voorwaarde lijkt dus vervuld.

Anderzijds is er in het geval van depressie ook sprake van een disfunctie van een geëvolueerd psychologisch mechanisme – Wakefields tweede voorwaarde . Sinds het einde van de jaren 1960 regent het namelijk evolutionaire verklaringen voor verdriet en neerslachtigheid (‘low mood’). Hoezo, hoor ik u vragen: hoe kan neerslachtigheid nuttig (adaptief) zijn? Evolutiepsychologen leren ons inderdaad dat de menselijke geest niet ontworpen is om ons gelukkig te maken. De menselijke geest is een zakmes dat ons toestaat om allerlei adaptieve problemen op te lossen, en zo onze kansen te vergroten in het grote gevecht om te overleven en ons voort te planten. Natuurlijke selectie maalt niet om ons geluk; ze is enkel geïnteresseerd in het aantal nakomelingen dat we voortbrengen. Verdriet en angst zijn natuurlijk geen aangename gevoelens. Maar waren ze nooit nuttig geweest, dan zouden we ze nu waarschijnlijk ook niet kunnen ervaren. (Vergelijk negatieve emoties met de vervelende verdedigingsmechanismen van het lichaam: overgeven, koorts, pijn, diarree, enzovoort. Hoogst vervelend, maar wel cruciaal om allerlei ziekmakers uit het lijf te weren.) Waarin ligt dan hun voordeel? Ik noem er snel drie (supra). Er is echter een belangrijk verschil tussen depressie en neerslachtigheid. Depressie is meer dan neerslachtigheid – het is een verstoring, een disfunctie van onze negatieve emoties, zoals angst en verdriet. In tegenstelling tot neerslachtige mensen zijn depressieve patiënten helemaal niet beter in het oplossen van sociale problemen (verklaring 1 & 2 van neerslachtigheid), of in het aantrekken van externe hulp (verklaring 3).

*Autisme* als schadelijke disfunctie van *theory of mind*-module, die ons toestaat om het gedrag van anderen te beschrijven, te verklaren en te voorspellen.

*Homoseksualiteit* werd om allerlei redenen uit de psychiatrische handboeken verwijderd, in de jaren 1970. Een van die redenen, zo zou Wakefield argumenteren, is dat homoseksualiteit voldoet aan geen van de voorwaarden die hij in zijn analyse noemt. Ten eerste is er geen sprake van ‘harm’ – de grote meerderheid van de homoseksuelen had en heeft geen problemen met hun eigen seksuele voorkeur. En als ze hun voorkeur toch als schadelijk (of pijnlijk) ervaren, heeft dat meestal te maken met familiale of culturele verwachtingspatronen en afwijzende reacties van de buitenwereld. Er is ook geen sprake van schade aan derden, ook al zijn er nog steeds hardnekkige geruchten dat homoseksualiteit een besmettelijke ziekte is. Ten tweede is er ook geen sprake van een disfunctie. Conservatieve filosofen hebben getracht om homoseksualiteit alsnog als een ziekte te bestempelen door te argumenteren dat homoseksualiteit een verstoring is van de natuurlijke functie van seksualiteit, die er dan in zou bestaan zich voort te planten, of banden te smeden tussen man en vrouw. Die argumenten houden echter geen steek. Er is namelijk een uitgebreide en uitstekend gedocumenteerde evolutionaire literatuur over de vele functies van homoseksualiteit. Denk maar aan het smeden van strategische allianties (in gevechten), het bekrachtigen van vriendschappen en loyauteiten (‘diddling’), enzovoort



Nog een laatste keer aantonen dat depressie geen adaptatie is.

Problemen met de harmful disfunction analysis

Om te beginnen steunt zijn analyse op de resultaten van een welbepaalde wetenschap, met name de evolutiepsychologie. Om te bepalen of een mentale toestand inderdaad een mentale stoornis is, moeten we nagaan of er sprake is van een disfunctie van een natuurlijk geselecteerde mentale module. Maar dan moeten we eerst weten of de module in kwestie voorkomt in het lijstje van de evolutiepsychologen. En daar wringt het schoentje. Er is namelijk niet veel eensgezindheid over zo’n lijst. Anders gezegd: de resultaten van de evolutiepsychologie zijn voorlopig eerder pover, geeft Wakefield zelf toe: ‘Because of the complexity of the inferences involved in judgments of dysfunction *and our relative ignorance about the evolution of mental functioning*, it is easy to arrive at differing judgments about mental dysfunction even on the basis of the same data (Wakefield 1992 , p. 386, mijn cursivering). Wakefield lijkt dit probleem als een tijdelijk probleem te zien, maar in de vorige les hebben we aangestipt dat er fundamentele conceptuele problemen zijn met bepaalde aspecten van de evolutiepsychologie, onder andere haar geloof in de massieve modulariteit van de menselijke geest. En er zijn critici die bovendien beweren dat onze onwetendheid over lichaam en geest van onze verre voorouders zodanig groot is dat we wellicht nooit een lijst van geëvolueerde mentale modules zullen kunnen opstellen.

Vervolgens zijn er een aantal voorbeelden van toestanden die wij geestesziekten noemen, en die niettemin niet voldoen aan de twee voorwaarden in Wakefields analyse. Enerzijds zijn er een aantal psychopathologieën die door de ‘patiënt’ zelf niet als schadelijk worden ervaren. Psychopathie is een voor de hand liggend voorbeeld. Psychopaten typeren zich door een totaal gebrek aan moreel gevoel – zij kunnen anderen bestelen en vermoorden zonder enige gewetenswroeging. De vraag is natuurlijk wat Wakefield precies bedoelt met ‘schadelijk’. Zelf besteedt hij nauwelijks aandacht aan deze voorwaarde, en er is een duidelijke ambiguïteit in de term. Je kan je immers afvragen of patiënten zelf hun toestand als schadelijk (in de zin van: pijnlijk, ongemakkelijk, onaangenaam) moeten ervaren, of eerder hun onmiddellijke omgeving. Sommige psychotische patiënten verzetten zich hevig tegen een opname, omdat ze hun toestand niet als problematisch ervaren (of beschouwen), terwijl familieleden of vrienden diezelfde toestand wel als problematisch ervaren (en daarom aansturen op een ‘gedwongen opname’ of collocatie – een opname zonder toestemming van de patiënt). Anderzijds zijn er ook een aantal psychopathologieën die niet voldoen aan de tweede voorwaarde – de disfunctie-voorwaarde. Dyslexie bijvoorbeeld. Wanneer we Wakefields analyse toepassen moeten we concluderen dat dyslexie geen stoornis is, want er is geen sprake van een disfunctie. Welk psychologisch mechanisme zou er immers disfunctioneren? Het zou idioot zijn om het bestaan te claimen van zoiets als een ‘leesmodule’ of een ‘schrijfmodule’, want lezen en schrijven zijn evolutionaire nieuwigheden. Het zijn met andere woorden vermogens die niet geëvolueerd zijn. (Bovendien voldoet dyslexie ook niet altijd aan de eerste voorwaarde van de HDA: de toestand is namelijk enkel schadelijk in een cultuur die lezen en schrijven hoog in het vaandel voert, wat uiteraard niet van elke cultuur in het verleden kan gezegd worden.)

Deze theorie is door en door essentialistisch. Wakefield denkt de essentie van psychophatologie gevat te hebben. Nadeel is (zoals vorige twee voorbeelden illustreren) dat het een geestesziekte is ja of nee, tussenin bestaat niet.

Tweede antwoord: Broken brain analysis

Een tweede antwoord op de vraag ‘wat is psychopathologie’ is eveneens een essentialistisch antwoord, en geeft als voldoende en noodzakelijke voorwaarde dat X een geestesziekte is dan en slechts dan als X een uiting is van een verstoord brein (‘broken brain’). Anders gezegd: X is een geestesziekte dan en slechts dan als het een neurologische stoornis is. Geestesziekten zijn hersenstoornissen, zoals epilepsie. Het is duidelijk dat deze zienswijze aansluit bij het reductionistisch fysicalisme. U herinnert zich nog dat aanhangers van deze theorie de stelling verdedigen dat wij ons brein zijn. Dat we samenvallen met onze hersenen. Dat alles wat er te vertellen valt over iemand, verteld *is* wanneer we diens brein uitputtend (exhaustief) beschreven hebben. Aanhangers van de BBA menen dat het net zo is met geestesziekten. Het probleem van de Rattenman, in Freuds gevalstudie die ik aanhaalde in de eerste les, is een probleem van zijn brein – een hersenstoornis. Om het probleem van de Rattenman te doorgronden, hebben we in principe voldoende aan een ‘brainoscope’ (zoals Smart die beschrijft in zijn tekst). Het brein van de Rattenman bevat alle relevante informatie over zijn psychopathologie.

U vraagt zich misschien af wat het verschil is tussen de HDA en de BBA? Is er überhaupt een verschil? Wakefields analyse vereist toch ook dat er sprake is van een stoornis of een disfunctie? Het grote verschil bestaat erin dat Wakefields analyse geen oordeel inhoudt over wat er zich in het brein afspeelt. Geen van de voorwaarden van de HDA verwijst naar de verhouding tussen een geestesziekte en het brein. Wakefield spreekt namelijk enkel over een disfunctie van *psychologische* mechanismen, ook al zal hij niet ontkennen dat die mechanismen op een of andere manier gerealiseerd zijn in het brein. Ik vergat tijdens de vorige les te vermelden dat evolutiepsychologen zich onthouden van uitspraken over hoe ‘hun’ modules zich verhouden tot het menselijk brein. Ze beweren in ieder geval niet, zoals Fodor wel doet, dat elke module overeenkomt met een specifiek deel van de hersenen. Ze vermijden daarmee het probleem van meervoudige realiseerbaarheid, dat in een eerdere les al aan bod kwam. Wakefield schaart zich ook op dit punt achter de evolutiepsychologen, en daarin onderscheidt zijn analyse van het concept van geestesziekte zich van de BBA.

De *broken brain analysis* van het concept van psychopathologie is erg populair bij zowel psychiaters als patiënten. Het is een analyse die centraal staat in de zogenaamde ‘biologische psychiatrie’ – een stroming die sinds de jaren 1970 de dominante ideologie vormt onder psychiaters in de meeste landen ter wereld. De eerste biologische psychiaters waren bovendien ook biologische psychiaters. De Duitse psychiater Wilhelm Griesinger, bijvoorbeeld, is bekend om zijn credo dat ‘Geisteskrankheiten sind Gehirnkrankheiten’. De eerste psychiaters hoopten dan ook de essentie te vinden van een aantal categorieën van geestesziekten (de essentie van schizofrenie, bijvoorbeeld) door intensief onderzoek te doen naar de hormoonhuishouding van patiënten, hun neurofysiologie, neuro-anatomie, of andere aanwijsbare lichamelijke disfuncties. Voor een uitgebreide discussie van de theoretische fundamenten van de biologische psychiatrie kan u terecht in het derde hoofdstuk van *Het Nut van Waanzin* (verkrijgbaar in de betere boekhandel tegen een scherpe prijs).

De term ‘Broken brain analysis’ is min of meer afkomstig van Nancy Andreasen, een bekende Amerikaanse psychiater-neurowetenschapper, die zich onder andere in haar boeken *The Broken Brain* (1984) en *Brave New Brain* (2004) een aanhanger toont van de claim dat geestesziekten eigenlijk niets anders zijn dan hersenziekten. Andreasen is een voorvechter van de biologische psychiatrie, en zou zich filosofisch outen als reductionistisch fysicalist.

Nog iets over verschillende motieven achter *Broken brain analysis*: a) wetenschappelijke motieven, b) morele motieven, zoals het bestrijden van de stigmatisering die nog steeds gepaard gaat met psychopathologie. De stelling dat geestesziekten niets anders zijn dan hersenziekten, en bij uitbreiding niets anders dan medische ziekten, zou volgens Andreasen helpen om de negatieve attitudes tegenover psychiatrische patiënten weg te nemen. Zo zegt ze dat zulke patiënten het verdienen om behandeld te worden ‘as human beings who deserve as much sensitivity and love as people who suffer from cancer, muscular dystrophy, or heart disease’.

Problemen met Broken brain analysis

De BBA is duidelijk schatplichtig aan het reductionistisch fysicalisme in de filosofie. De onderliggende gedachte is namelijk dat geestesziekten niet meer zijn dan hersenstoornissen, net zoals mentale toestanden uiteindelijk niet meer zijn dan hersentoestanden. Zoals de Amerikaanse psychiater-filosoof Kenneth Kendler opmerkt: 'Driven in part by exciting advances in neuroscience and molecular genetics, a reductionist "brain-focused" paradigm is currently becoming increasingly dominant. Advocates of this paradigm argue vigorously that all "real" explanations for psychiatric illness can be found at the molecular level. Some of this persuasion might tolerate other explanatory approaches (say, at the level of systems neuroscience or even psychology) but only as a holding pattern until the necessary scientific advances render these perspectives obsolete' (Kendler 2008, 3). Het doel van de BBA is niet om psychologische benaderingen van psychopathologie aan te vullen, maar om ze overbodig te maken (‘obsolete’). Biologische psychiaters zijn geen eliminativistische fysicalisten, denk ik, ook al zijn er die op de grens tussen beide vormen van fysicalisme balanceren. Die grens is overigens niet altijd even duidelijk. Ik heb geargumenteerd dat het grote verschil tussen R en E fysicalime te maken heeft met hun waardering voor de niet-fysische wetenschappen, zoals de psychologie. In deze context zou je je de vraag kunnen stellen: is de psychologie een belemmering voor biologische psychiaters – iets waarmee ze moeten afrekenen voor ze aan het echte werk kunnen beginnen? Of kan de psychologie ons de weg wijzen naar ‘het echte werk’? In Kendlers opvatting over de biologische psychiatrie lijkt de psychologie nog een tijdelijke rol te kunnen spelen. ‘Verdringing’, zo zouden reductionistische fysicalisten zeggen, is uiteindelijk *niets anders dan* een strikt neurofysiologisch proces. Maar dat betekent niet dat de term ‘verdringing’ opeens waardeloos wordt. Denk aan het voorbeeld van ‘water’ en ‘H²O’ – we weten weliswaar dat water niets anders is dan een samenstelling van waterstofatomen en zuurstofatomen, maar het is wel een verdomd handige term in het dagelijks leven. Eliminativistische fysicalisten hebben een harder oordeel voor de psychologie: het is wartaal die het echte (lees: neurowetenschappelijke) onderzoek belemmert, en waar we dus zo snel mogelijk van af moesten. De psychologie zal, net als de flogiston-theorie over ontbranding, vroeg of laat in de bak met pseudowetenschappen belanden. De meeste biologische psychiaters zouden niet zover gaan, denk ik. Ze zouden zich eerder vinden in een reductionistisch-fysicalistisch verhaal, waarin vroeg of laat zal aangetoond worden dat alle psychische stoornissen eigenlijk neurologische stoornissen zijn – stoornissen van het brein. Als het waar is dat de BBA schatplichtig is aan het reductionistisch fysicalisme, dan kan je deze analyse bestoken met dezelfde argumenten waarmee we het R fysicalisme bestookt hebben.

Daarnaast is er nog een ander probleem met de BBA. Het grote probleem met dit antwoord (op de vraag ‘wat is psychopathologie?’) is dat het onduidelijk wordt waarom we nog over *geestes*ziekten zouden spreken. Wat is er precies ‘psychisch’ aan psychische stoornissen als het eigenlijk hersenstoornissen zijn? In een cynische bui zou je kunnen zeggen dat het onderscheid tussen neurologische en psychiatrische stoornissen een tijdelijk onderscheid is, dat vroeg of laat zal verdwijnen. Het is een onderscheid dat gebaseerd is op onwetendheid over de ware aard van geestesziekten, en dat onvermijdelijk zal verbrokkelen naarmate de hersenwetenschappen dichter bij hun einddoel geraken.

Maar die claim, dat psychische stoornissen niet meer zijn dan hersenstoornissen, is fundamenteel tegen-intuïtief.

Een voorbeeld om deze intuïtie te expliciteren.

In de film *Turks Fruit* (1973) (naar de gelijknamige roman van Jan Wolkers) trouwt de bohémien-kunstenaar Eric Vonk met de ravissante Olga. De relatie is nogal turbulent, en eindigt tijdens een etentje dat georganiseerd werd door de burgerlijke familie van Olga. Tijdens het etentje begint Olga zich vreemd te gedragen. In het bijzijn van haar echtgenoot flirt ze ostentatief met een haar onbekende zakenman, wat Eric danig op de zenuwen werkt, en een knetterende ruzie veroorzaakt, waarin hij haar slaat. Game over. Uiteindelijk komt er een scheiding, en in de periode die daarop volgt wordt Olga’s flirtgedrag steeds extremer. Ze kleedt zich erg flamboyant, en haar spreken mist coherentie. Op dit punt in de film beantwoordt Olga’s gedrag aan een aantal criteria die in DSM-III (1980) in de richting wijzen van *Nymphomania* - een subcategorie van de diagnostische groep ‘Psychosexual Disorders’. Een soort seksverslaving, zeg maar. En een seksverslaving is een voorbeeld van een psychische stoornis.

Even later in de film blijkt echter dat Olga een ernstige hersentumor heeft, die volgens de dokters een voldoende verklaring biedt voor haar ‘vreemde’ gedrag van de voorbije maanden. Olga heeft met andere woorden een hersenstoornis. En onze intuïtie zegt ons dat zo’n stoornis fundamenteel verschilt van een psychische stoornis. Het verschil is dat nymfomanie ingebed is in Olga’s persoonlijke geschiedenis – we verzinnen meteen allerlei eenvoudige psychologische verklaringen voor zulk gedrag, in de trant van: ‘ze is manisch op zoek naar haar verloren vader’ of ‘ze probeert andere mannen betaald te zetten wat haar eigen man haar aandeed’, enzovoort. Olga’s gedrag is dus beladen met allerlei betekenissen. Zulke verklaringen vervallen wanneer er sprake is van een breintumor. Op dat moment wordt haar gedrag zuiver mechanisch; ze wordt voortgedreven door wild knetterende neuronen, door de wildgroei van cellen in haar brein. Olga’s intenties verdwijnen als sneeuw voor de zon; haar gedrag wordt betekenisloos, net als haar spreken.

Laatste antwoord: Prototype resemblence analysis

In *The Disordered Mind* (2010) bekritiseert Graham allerlei essentialistische antwoorden op de vraag ‘wat is psychopathologie?’ Zijn alternatief is een antwoord dat schatplichtig is aan de filosofie van de latere Wittgenstein – ook wel ‘Wittgenstein II’ genoemd. Ik noem het de ‘prototype-resemblance analysis’ van het concept van psychopathologie. De HDA en de BBA gaan er allebei vanuit dat er een bescheiden lijst van noodzakelijke en voldoende voorwaarden bestaat waaraan een mentale toestand moet voldoen om een psychische stoornis genoemd te worden. Volgens Wakefield en Andreasen bestaat er zoiets als de essentie van het concept van psychopathologie. Daarom noemen we hun benaderingen essentialistisch. Geen van deze benaderingen is echter onproblematisch, en in een Wittgensteiniaanse zienswijze is daar ook een goede reden voor. Het concept van psychopathologie heeft namelijk geen essentie, evenmin als tal van andere concepten, zoals het concept van taal, of het concept van spel. Er zijn waanzinnig veel soorten spelen en spelletjes, gaande van gezelschapsspelen over gokspelen en volksspelen tot Olympische spelen. Sommige spelen lijken in bepaalde opzichten erg op elkaar: bordspelen en kaartspelen zijn bijvoorbeeld allebei gezelschapsspelen. Andere spelen, zoals een volksspel en de Olympische spelen, hebben nauwelijks iets gemeen. Er is dus niet zoiets als een unieke gemene deler voor alle dingen die wij ‘spelen’ noemen. (Probeer er maar eens eentje te vinden! Een tegenstander? Bij steltenlopen is er geen tegenstander. Een publiek? Bij computerspelletjes is er geen publiek. Vrije tijd? Maar er zijn ook mensen die professioneel spelen – voetballers bijvoorbeeld.) En zo is het ook met geestesziekten. Er zijn zoveel toestanden die we bestempelen als geestesziekten, dat het onmogelijk is om in die veelheid een unieke gemene deler te vinden. De vraag ‘wat is psychopathologie?’ zou in zijn ogen een verkeerde vraag zijn, precies omdat ze essentialistisch is. Ze gaat er vanuit dat alle psychische stoornissen een uniek definiërende eigenschap hebben, en dat die eigenschap ons toestaat om psychische stoornissen als stoornissen te identificeren, en ze te onderscheiden van andere psychische stoornissen, neurologische stoornissen (in het geval van HDA), en van normale levensproblemen (‘problems in living’).

Vindt u dat allemaal nogal pessimistisch? Dan bent u wellicht blij met de volgende bedenkingen. Ten eerste is het zo dat er wel degelijk fenomenen zijn die wel op een essentialistische manier kunnen begrepen worden. Een goed voorbeeld van een fenomeen dat wel gedefinieerd kan worden met behulp van een lijst van noodzakelijke en voldoende voorwaarden zijn de elementen in de tabel van Mendeljev. Elk van deze elementen heeft een vaste plaats in de tabel, die bepaald wordt door het precieze aantal protonen waaruit één enkel atoom van het element in kwestie bestaat. Het element ‘goud’ (of ‘aurum’) kan je definiëren als een substantie waarvan elk afzonderlijk atoom samengesteld is uit precies 79 protonen. Niet meer of niet minder. En alle atomen waarvan kan aangetoond worden dat ze uit 79 protonen bestaan zijn goudatomen. Er is met andere woorden geen vergissing mogelijk. Het hebben van 79 protonen is de essentie, de uniek definiërende eigenschap van goud. Ze staat ons toe om goud te identificeren, en vervolgens te onderscheiden van andere elementen in de tabel. Je kan overigens een gelijkaardig argument opzetten in verband met menselijke bloedgroepen. Bloedgroepen (A, B, AB en O) laten ons toe om mensen op een eenduidige manier te groeperen op basis van een uniek definiërende eigenschap, met name hun bloedgroep. Een eenvoudige bloedtest wijst uit tot welke bloedgroep je behoort. Omdat goudatomen en mensen met een bepaalde bloedgroep een natuurlijke verzameling vormen, dat wil zeggen een verzameling van dingen die een natuurlijke essentie delen, noemt men goud en bloedgroepen *natuurlijke soorten*. In het Engels spreekt men over *natural kinds*.

Een tweede bedenking bij Wittgensteins kritiek op het essentialisme is dat ze niet hoeft te leiden tot volstrekte chaos in de psychiatrie. Volgens Wittgenstein is het namelijk zo dat sommige instanties van een fenomeen als prototype kunnen dienen. Prototypes zijn gevallen waarover de meeste betrokken partijen het eens zijn dat ze dat tot een bepaald fenomeen behoren. Een voorbeeld: het concept van spel. Wittgensteins opmerking dat er geen uniek definiërend kenmerk bestaat voor alle instanties van het concept van spel (voor alle spelen dus) hoeft niet tot wanhoop te leiden, want er bestaan minstens een aantal prototypische spelen – dingen waarover iedereen het eens is dat het spelen zijn. Monopoly bijvoorbeeld. Wie zou er durven beweren dat Monopoly geen spel is? Idem dito voor het concept van psychopathologie. Het is misschien onmogelijk om de essentie vast te leggen van het concept van psychopathologie, maar dat neemt niet weg dat bepaalde mentale toestanden duidelijk geestesziekten zijn. Er zijn bijvoorbeeld weinig psychiaters die durven beweren dat schizofrenie geen psychische stoornis is. Schizofrenie is met andere woorden een prototype van een geestesziekte. En zulke prototypes kunnen ons helpen om een aantal typische kenmerken te vinden van een bepaald concept. Daarover dadelijk meer.

Ten derde en ten laatste is het zo dat instanties van een fenomeen misschien geen unieke gemene deler hebben, maar wel allerlei verwantschappen. Wittgenstein sprak van ‘familiegelijkenissen’. Elke familie (in de ruime zin: geen kerngezin, maar een aantal generaties, met neven en nichten etc) kenmerkt zich door bepaalde eigenschappen – een haakneus, bijvoorbeeld, of een neiging tot drankzucht, of luidruchtigheid. Niet alle familieleden hebben die eigenschap, en de eigenschap komt ook niet enkel in die familie voor, maar de eigenschap in kwestie vormt wel een soort rode draad in een bepaalde familie. Sommige familieleden vertonen een aantal van die eigenschappen, en kunnen daarom beschouwd worden als prototypes van die familie. Punt is dat de meeste familieleden gelijkenissen vertonen, met elkaar en met de prototypes. De kans is groot dat er in de kern van de familie (whatever that may be) een aantal prototypes zullen gevonden worden, en dat de gelijkenissen langzaam uitsterven naarmate we verdergaan in de tijd (bedovergrootouders, achter-achterkleinkinderen) en in de verwantschap (achter-achterneven). Wanneer we een familie willen typeren, zegt Wittgenstein, gaan we niet op zoek naar de essentie van die familie, maar naar prototypes en familiegelijkenissen. En die analyse zou je kunnen doortrekken naar psychopathologie. De verschillende (categorieën van) geestesziekten, zoals depressie, alcohol-verslaving, psychopathie, en pyromanie, hebben geen unieke gemene deler, zoals Wakefield en Andreasen beweren, maar er zijn wel prototypes, en de meeste categorieën vertonen allerlei gelijkenissen. Over sommige categorieën, zoals eetstoornissen of ADHD, zal er heel wat geruzied worden – zij bevinden zich aan de rand van de familie. Zij vertonen nauwelijks nog gelijkenissen met het prototype.

Intermezzo: Het rijk van wittgenstein

Enkele voorbeelden van fenomenen die je kan begrijpen met behulp van een Wittgensteiniaanse *prototype-resemblance analysis*. In de vorige les zagen we bijvoorbeeld dat Fodor het concept van module niet op een essentialistische manier invult, maar eerder op een gradualistische manier. Sommige mentale systemen zijn op een prototypische wijze modulair – denk maar aan de visuele waarneming – terwijl andere systemen ‘op het randje van modulair’ zijn. De mate waarin een systeem modulair is kan bepaald worden aan de hand van het aantal kenmerken uit Fodors lijst die op het systeem van toepassing zijn, waarbij het relatieve gewicht van elk kenmerk bovendien verschillend is.

Graham legt uit waarom het concept van boek ook op een Wittgensteiniaanse manier kan begrepen worden: ‘The term “book” is not messy or contestable in all of its uses. Some applications are prototypical or examplary. Consensus applications, if you will. War and Peace, for example, is a consensus application of the term. No truly informed person would deny that Tolstoy’s masterpiece is a book. Others are foil applications, by which I mean clear cases of non-books (e.g. a single blank piece of paper on a desk). No truly informed user of the word “book” would be tempted to classify a single blank piece of paper as a book. A characterization of the very idea of a book can be extracted from exemplary applications. Then, when instances of use of the term drift toward foils (clear non-cases), however, the wisdom of the applicability of the concept becomes increasingly contestable or debatable and eventually dissolves. Here’s a description of what the word “book” means (extrapolating from consensus applications): “A set of written, printed, or blank pages fastened along one side and encased between protective covers.” That’s how my desk dictionary defines the term. The description applies both to the object you are reading at present and to War and Peace. But it fails to apply to, say, a single blank piece of paper. True, some candidate books (e.g. a manuscript in your computer file) are not covered by such a characterization. But uncontroversial or exemplary cases are covered. Borders between exemplary books and non-books (foils) are, as noted, contestable or debatable. We can debate rights of semantic immigration (should comics be counted in as books?) and emigration (should the manuscript in your computer be counted out as a book?). But clear-cut cases are clear-cut and to condemn a whole or entire concept to catch-all status just because various instances of application provoke controversy and doubt is to misunderstand how sound and empirical notions often work’ (Graham 2010, 22-3).

Problemen met PRA

Grahams Wittgensteiniaanse antwoord op de vraag ‘wat is psychopathologie?’ heeft in mijn ogen een zekere aantrekkingskracht, maar het heeft ook enkele zwakheden.

Ten eerste is het een vaag antwoord – ze laat ons niet toe om een strakke lijn te trekken tussen het concept van psychische stoornis en het concept van neurologische stoornis. Graham zou wellicht zeggen dat deze twee Wittgensteiniaanse verzamelingen (de verzameling van alle instanties van het concept van psychische stoornis, en de verzameling van alle instanties van het concept van neurologische stoornis) elkaar overlappen. Idem dito voor de grens tussen psychopathologie en levensproblemen, zoals rouw of faalangst. Het gebrek aan eenduidige criteria staat haaks op de ‘filosofie’ van de hedendaagse gezondheidszorg, waarin bijvoorbeeld enkel behandelingen voor psychische stoornissen worden terugbetaald, maar niet voor levensproblemen.

Ten tweede, en aansluitend bij het eerste bezwaar, kan de PRA misbruikt worden door allerlei lobby-groepen die een bepaalde mentale toestand willen laten herkennen als een psychopathologie. In de voorbije halve eeuw zijn er tal van fenomenen tot stoornissen gebombardeerd onder druk van actiegroepen. Een voorbeeld: in 1973 beslisten de leden van de *American Psychiatric Association* (APA)*,* onder leiding van Robert Spitzer, dat homoseksualiteit officieel niet langer beschouwd werd als een geestesziekte (wat wel het geval was in DSM-II)*.* Gelijkaardig lobbywerk en schimmige stemmingen zorgen ervoor dat ook andere ziektebeelden in het handboek opgenomen worden: *post traumatic stress disorder, repressed memory syndrome, multiple personality disorder –* de lijst is haast eindeloos*.* De Canadese historicus Edward Shorter zegt daarover het volgende: ‘Toen eenmaal bekend was geworden hoe vlot de Nomenclatuurcommissie van de APA bezweken was voor homoseksualiteit, was het duidelijk dat de psychiaters gemakkelijk om de vinger te winden waren. (...) Deze zaken [d.i. seksuele voorkeur, stress of maandstonden] konden kennelijk allemaal gepathologiseerd en ontpathologiseerd worden volgens de wil van de meerderheid, of door pressie van campagnes en hardnekkige groepen. Men had de wetenschap dus niet als gids kunnen nemen, en dat verklaart waarom DSM-III en zijn opvolgers, bedoeld om de psychiatrie uit het moeras van de psychoanalyse te voeren, haar in werkelijkheid naar de wildernis heeft geleid’ (Shorter 1998, 325-6). De grens tussen levensproblemen en psychopathologie is erg vaag, en zulke vaagheid werkt in de hand van drukkingsgroepen. (Je zou je natuurlijk kunnen afvragen waarom dat een probleem zou zijn. Lobbywerk leidt vaak tot goede resultaten, zoals in het geval van de verwijdering van homoseksualiteit uit de psychiatrische handboeken. Even vaak leidt het echter tot een pathologisering van het dagelijks leven, zoals Shorter opmerkt in het citaat hierboven, waarbij er minstens de schijn is dat de werkelijkheid geweld wordt aangedaan.)

Een derde en laatste kritiek op Grahams PRA is dat het in wezen een gemakzuchtige oplossing is. Het ontslaat ons namelijk van de verantwoordelijkheid om verder te zoeken naar een unieke gemene deler. Deze kritiek is echter niet helemaal terecht. Graham besteedt bijvoorbeeld heel wat tijd aan het omschrijven van enkele familiegelijkenissen van (categorieën van) psychische stoornissen.

Enkele prototypes

Een interessante vraag is (en eigenlijk een vierde kritiek op Grahams analyse): hoe bepalen we welke instanties van een fenomeen prototypisch zijn? In het geval van geestesziekten: hoe bepalen we welke geestesziekten prototypische ziekten zijn? De zoektocht naar prototypes is niet onbelangrijk, want de kenmerken van prototypes bepalen welke toestanden ook in aanmerking komen om ‘ziekten’ genoemd te worden. Graham noemt enkele criteria waarmee we op zoek kunnen gaan naar prototypes van psychopathologie:

*Universaliteit*. Een stoornis die aantoonbaar voorkomt in verschillende, geografisch verspreide gebieden, en misschien zelfs overal ter wereld, lijkt een goede kandidaat te zijn voor een prototypische psychische stoornis. Maar: waarom zouden alle prototypische stoornissen universeel moeten zijn? Denk even aan de geneeskunde: sikkel-cel anemie is een ziekte die voornamelijk voorkomt in gebieden rond de Middellandse Zee, en in specifieke regio’s in Afrika. Nochtans zijn er goede redenen om aan te nemen dat sikkel-cel anemie een *prototypische* ziekte is (disfunctie van bloedcellen, mortaliteit, erfelijkheid,…), en er is in ieder geval niemand die zal betwijfelen dat sikkel-cel anemie een ziekte is. We moeten ons hoeden voor de gedachte dat enkel universeel voorkomende stoornissen echte, prototypische stoornissen kunnen zijn. En een tweede bedenking bij Grahams criterium van universaliteit: sommige psychische stoornissen komen overal ter wereld voor omdat de Amerikaanse psychiatrie, meer bepaald de Diagnostic and Statistical Manual of Mental Disorders, erin geslaagd is om haar ziektecategorieën te exporteren naar tal van niet-Westerse landen. Ik citeer even uit een eigen paper-in-wording over dit thema: ‘some disorders may well be found throughout the world, but only because they were exported by American psychiatry to various non-Western societies. It has been argued, for example, that international pharmaceutical companies went to great lengths to introduce specific DSM disorder categories in Argentina, because the conditions at which their products were targeted were simply not being seen there. By sponsoring training events and distributing drug samples and key texts, they helped shape the way mental health professionals think about mental illnesses, and how patients experience such illnesses (Lakoff 2005). Similar accounts have been given of the marketing of depression in Japan (Applbaum 2006), and the sudden epidemic of anorexia nervosa in the 1990s in Hong Kong (Watters 2010)’.

*Frequentie*. Een stoornis die in verschillende landen, en volgens verschillende prevalentiestudies, vaker voorkomt dan andere stoornissen, lijkt ook een goede kandidaat. Maar ook bij dit criterium zijn er een aantal caveats. Ik noem er eentje. Sinds het begin van de 21ste eeuw spreekt men van een heuse depressie-epidemie in het Westen. Onze welvaart kan blijkbaar niet beletten dat maar liefst een kwart van de bevolking ten prooi valt aan depressie. Volgens de Wereldgezondheidsorganisatie zijn er wereldwijd niet minder dan 121 miljoen mensen klinisch depressief, en de voorspelling is dat depressie binnen twintig jaar naar de top klimt van het lijstje van doodsoorzaken bij de mens. Enkel hartziekten doen momenteel ‘beter’. De prevalentiecijfers van depressie zijn zonder meer duizelingwekkend, en weerspiegelen zich in de steeds stijgende verkoop van antidepressiva. In België vormen zulke geneesmiddelen de op een na grootste uitgavenpost van het Rijksinstituut voor Ziekte- en Invaliditeitsverzekering (RIZIV), en wereldwijd zijn maar liefst drie van de zeven best verkopende geneesmiddelen-op-voorschrift antidepressiva (Prozac, Paxil en Zoloft). Wat is er eigenlijk aan de hand? Een groeiend leger van critici betwijfelt dat de zogenaamde depressie-epidemie een reëel fenomeen is, en zoekt eerder naar antwoorden in het ontrafelen van de *liaison dangereuse* tussen psychiatrische diagnostiek en farmaceutische industrie. Depressie is trouwens niet de eerste epidemie van psychische stoornissen in het Westen. In de jaren vijftig van de vorige eeuw leed een groot deel van de Amerikaanse bevolking aan angststoornissen; enkele decennia later, aan het eind van de jaren tachtig, werden de psychiatrische praktijken overspoeld door patiënten met stemmingsstoornissen. Historicus Edward Shorter merkt op: ‘Precies zoals een natie die onder angsten leed, zich door Valium had laten kalmeren, zo had het bestaan van een nieuw antidepressivum [Prozac] een ziektepatroon geschapen dat met dit medicijn behandeld kon worden’ (Shorter 1998, 345; zie ook Valenstein 1998). Als depressie al een prototypische psychische stoornis is, dan is het in ieder geval niet omwille van haar prevalentie-cijfers. Daarmee bedoel ik: de frequentie waarmee bepaalde stoornissen zich voordoen, kan ook een artefact zijn, geen *natuurlijk feit*

* 1. **Wat maakt een psychische stoornis tot psychisch (metafysisch)**

De vraag ‘wat is psychopathologie’ kan je op twee verschillende manieren interpreteren. Enerzijds kan je de vraag lezen als een vraag naar wat psychische stoornissen tot *stoornissen* maakt. Waarom noemen we sommige toestanden stoornissen, en andere niet? Tijdens de vorige paragraaf besprak ik drie antwoorden op deze vraag. Anderzijds kan je de vraag, ‘wat is psychopathologie?’, ook lezen als een vraag naar wat er *psychisch* is aan psychische stoornissen. Het beantwoorden van deze laatste vraag brengt ons opnieuw naar het eerste hoofdstuk, waarin we onder andere probeerden uit te vlooien waarom we bepaalde toestanden *mentale* toestanden noemen. Het is deze laatste vraag die centraal staat in deze paragraaf en in de tekst van Brülde & Radovic.

hoe onderscheiden we ‘neurologische stoornissen’ van ‘psychische stoornissen’? Waarom staat een ziekte als neurosyfilis (categorie 292.0: Psychosis with general paralysis) (de eindfase van een syfilis-infectie, waarin het hersenweefsel aangetast wordt door de syfilis-bacterie Treponema Pallidum) geboekt als een psychische stoornis – bijvoorbeeld in het Amerikaanse Diagnostic & Statistical Manual of Mental Disorders? De oorzaak van de aandoening lijkt immers allesbehalve psychisch te zijn – een bacteriële infectie. De symptomen zijn deels van somatische aard (blindheid, incontinentie, stijfheid van de nek, spierzwakte,…), deels van psychische aard (depressie, verwardheid, dementie,…).

De discussie over de plaats van neurosyfilis in DSM is een goed voorbeeld van een tweede interpretatie van de vraag ‘wat is psychopathologie?’ Je kan je immers afvragen waarom we neurosyfilis als een psychische stoornis zouden moeten beschouwen, eerder dan als een neurologische stoornis.

Keren we even terug naar het eerste. We besloten toen dat er ruwweg drie soorten van mentale toestanden zijn: ervaringen, propositionele attitudes en mentale handelingen. Deze drie categorieën zou je kunnen zien als prototypes van het concept van geest, net zoals schizofrenie en depressie prototypes zijn van het concept van psychopathologie. (In wat volgt concentreer ik me vooral op ervaringen & propositionele attitudes, en laat ik mentale handelingen gemakshalve buiten beschouwing.) We stelden toen ook de vraag wat zulke prototypes maakt tot prototypes van mentale toestanden. Anders gezegd: wat maakt bepaalde toestanden tot *mentale* toestanden? Ik herneem een slide over deze vraag, uit de eerste les, en ik vul aan. (Nota: wanneer het onderstaande de slide uit de eerste les tegenspreekt, dan mag u vergeten wat er in de eerste les gezegd werd…) Zie ook B&R, p. 101-3.

Marks of the mental

*Intentionaliteit* (*intentionality* , *directionality* of *aboutness*): ‘aboutness’ is typisch voor attitudes en voor ervaringen. Een waarneming is altijd gericht op hetgeen waargenomen wordt: een voorwerp, een geluid, een geur,… Attitudes zijn altijd gericht op proposities, die je kan begrijpen als gedachte-inhouden. Brentano wordt vaak beschouwd als de bedenker van de term ‘intentionaliteit’, en volgens hem is intentionaliteit de essentie van het mentale: ‘Kenmerkend voor alle psychische verschijnselen is wat de middeleeuwse scholastici de intentionele (of mentale) inexistentie van een object noemden, en wat wij (…) de betrekking op een inhoud zouden noemen, de gerichtheid op een object (die hier niet moet worden opgevat als een realiteit) (…). Elk van deze verschijnselen draagt iets in zich als object (…). In de voorstelling wordt iets voorgesteld, in het oordeel iets aanvaard of verworpen, in de liefde iets bemind, in de haat iets gehaat, in het begeren iets begeerd enzovoort. Deze intentionele inexistentie is uniek voor de psychische verschijnselen. Bij geen enkel fysisch verschijnsel doet zich iets dergelijks voor’ (geciteerd in Dooremalen et al. 2010, 42). Intentionaliteit lijkt in mindere mate van toepassing op mentale *ervaringen*. Je kan je bijvoorbeeld afvragen waarop stemmingen of gemoedstoestanden (‘moods’) gericht zijn. Het lijkt net typisch voor zulke mentale toestanden dat ze niet gericht zijn op iets specifiek. Het zijn eerder vage, diffuse toestanden. Of neem een ander voorbeeld: pijn. Is pijn intentioneel? Waarop zou mijn ervaring van pijn dan gericht zijn? Sommige filosofen menen dat ook pijn intentioneel is, in die zin dat het gericht is op de schade aan het pijnlijke orgaan. Verlangens, geuren, herinneringen en overtuigingen zijn daarentegen duidelijk intentioneel: ze verwijzen naar iets anders dan zichzelf. Je kan je echter de vraag stellen of intentionaliteit enkel typisch is voor mentale toestanden. Er zijn immers zoveel dingen die ‘naar iets anders dan zichzelf verwijzen’ – denk maar aan woorden, bouwplannen of wegwijzers. Zijn wegwijzers intentioneel? Om die vraag te beantwoorden maken sommige filosofen een onderscheid tussen twee vormen van intentionaliteit: intrinsieke en extrinsieke intentionaliteit. Mentale toestanden zijn intrinsiek intentioneel – hun intentionaliteit is met andere woorden eigen aan de toestand zelf. Wegwijzers en woorden worden daarentegen door iets of iemand anders beladen met intentionaliteit. Diegene die een woord gebruikt bepaalt in zekere zin waarnaar het woord verwijst. Een voorbeeld: het woord ‘chips’. Wanneer het woord gebruikt wordt door een Engelsman, dan verwijst het naar frieten; wanneer het woord gebruikt wordt door een Vlaming, dan verwijst het naar krokant gebakken aardappelschijfjes. De intentionaliteit van het woord ‘chips’ wordt dus bepaald door diegene die het woord hanteert (of de context waarin het woord gebruikt wordt) – daarom noemen we zulke intentionaliteit ‘extrinsiek’ (zie ook Brülde & Radovic, p. 101, rechterkolom bovenaan).

*Fenomenaal bewustzijn* (*phenomenal consciousness*): lijkt ook te gelden voor zowel ervaringen als attitudes. Het enige probleem met dit kenmerk van het mentale is dat er ook zoiets bestaat als onbewuste mentale toestanden. Sinds Freud nemen veel filosofen aan dat de geest bestaat uit deels bewuste toestanden en (groten)deels onbewuste toestanden – twee systemen met daartussen een grens die, nog steeds volgens Freud, streng bewaakt wordt door de zelfcensuur. Als bewustzijn een ‘mark of the mental’ is, zijn onbewuste verlangens, herinneringen en overtuigingen dan geen mentale toestanden? Die claim lijkt tegen-intuïtief. Een oplossing is dat we ‘bewustzijn’ herdefiniëren als ‘potentieel bewustzijn’. Onbewuste mentale toestanden zijn dan mentale toestanden in de mate dat ze potentieel toegankelijk zijn voor het bewustzijn. Bijvoorbeeld met behulp van een intensieve psychotherapie. Een van de belangrijkste kenmerken van het bewustzijn is dat het door en door subjectief is. Bewuste ervaringen, zo heb ik uitgelegd in de voorbije lessen, hebben een kwalitatief of fenomenaal karakter – je kan over zulke ervaringen spreken door te vragen hoe het is om ze te hebben. Hoe is het om een vleermuis te zijn? Om blind te zijn? Om de geur van versgemaaid gras te ruiken op een mooie lente-ochtend? Bewuste ervaringen hebben qualia, en qualia zijn, volgens sommige filosofen ten minste, prototypische mentale toestanden. Ze zijn prototypisch in die zin dat ze (nog steeds volgens sommige filosofen) onmogelijk tot hersentoestanden kunnen herleid worden. Merk op dat B&R zichzelf tegenspreken over de vraag of ook attitudes qualia kunnen hebben. Enerzijds: ‘The characterization in terms of “what it is like to be” in a mental state (…) seems to include thought and emotion…’ (p. 102, linksbovenaan). Anderzijds: ‘Thoughts (...) may be characterized as subjective *despite the nonphenomenal status of such traits*’ (102-3).

*Subjectiviteit*: lijkt ook te gelden voor zowel ervaringen als propositionele attitudes. Ik ben op een bepaalde manier verbonden met mijn ervaringen en attitudes – het zijn niet alleen intentionele, bewuste toestanden, maar ook *mijn* mentale toestanden. Epistemologisch vertaald dit kenmerk van het mentale zich als oncorrigeerbaarheid (*incorrigibility*): door introspectie heb ik een gepriviligieerde toegang tot mijn eigen geest, mijn eigen ervaringen, gedachten/attituden en handelingen. Introspectie zorgt ervoor dat ik betere (rechtstreekse) kennis heb over mijn eigen geest dan over de geest van andere mensen (of dieren), en dat mijn kennis over mijn eigen geest beter (of in ieder geval onmiddellijker of directer) is dan mijn kennis over de buitenwereld. Oncorrigeerbaarheid betekent meer bepaald dat niemand kan aantonen dat mijn ervaringen of attitudes onwaar zijn. (Varianten: onbetwijfelbaarheid, onfeilbaarheid,…)

Er zijn verschillende manieren om te antwoorden op de vraag wat er juist psychisch is aan psychische stoornissen. Ik het verdere verloop van deze paragraaf haal ik er vier aan. Met name the internal caus view (psychische oorzaken), the symptom view (psychische symptomen), Pluralist view (psychologische oorzaken en symptomen= niet behandendelen) en the minimalist view (een hopeloze zaak)

Psychische oorzaken

Sommige filosofen oordelen dat de aard van de oorzaak van pathologie een bepalende rol speelt in het onderscheiden van psychische en neurologische (of lichamelijke) stoornissen. B&R: ‘A disorder is mental only if it is necessary to explain the symptoms in psychological terms’. Wakefield is volgens B&R een vertegenwoordiger van deze positie. Zijn HDA gaat er immers vanuit dat psychopathologie een schadelijke disfunctie of verstoring is van een geëvolueerd *psychologisch* mechanisme. ‘The internal cause view’ is met andere woorden de zienswijze ‘that a disorder can only be mental if the proximate or immediate cause (the underlying pathology) of the symptoms involved is (at least in part) mental’ (B&R, p. 103). Graham verdedigt deze zienswijze als volgt: ‘Mere mental symptoms are completely and utterly insufficient for calling a condition a “mental disorder”. Did [Virginia] Woolf kill herself and intentionally so because she felt hopelessly trapped or did she do so accidentally because of a defective alteration of neurotransmitter function due to Parkinson’s disease? It’s important to know what causes or helps to cause a set of symptoms if we are to grasp what sort of disorder a disorder is and what sorts of behavior its symptoms are. Depression-related deliberate suicide is one thing; accidentally killing oneself because of a prefrontal lobe dysfunction in an episode of lost motor control is another’ (Graham 2010, 36).

Elke psychische stoornis is de uitkomst van een lange causale keten, en het is duidelijk dat het fysicalisme ons niet toestaat om de oorspronkelijke of eerste oorzaak van een stoornis als een psychische oorzaak te beschouwen. B&R doen een halfslachtige poging (p. 103, rechts onderaan) om twee soorten van oorzaken te onderscheiden: ‘nabije’ (*proximate*) oorzaken en verre (*distal*) oorzaken. Het onderscheid doet denken aan het onderscheid tussen nabije en ultieme oorzaken in de evolutiebiologie (een onderscheid dat geformuleerd werd door de filosoof Ernst Mayr). Ultieme (of evolutionaire) oorzaken geven een antwoord op de vraag *waarom* een organisme een bepaalde eigenschap (bijvoorbeeld een taalmodule) heeft (met name: omdat de eigenschap een voordeel opleverde in de strijd om te overleven en zich voort te planten), terwijl nabije oorzaken een antwoord geven op de vraag hoe een organisme die eigenschap ontwikkelt (met name: omdat zijn genen of hormonen of wat dan ook daartoe aanleiding geven). De vraag die B&R hier aankaarten (en niet echt uitwerken) is een fundamentele vraag in de philosophy of mind: bestaat er zoiets als *mental causation* – mentale veroorzaking? Het is vanzelfsprekend dat een mentale toestand de oorzaak kan zijn van een (andere) mentale toestand, maar kan het ook de oorzaak zijn van hersentoestanden? Mentale veroorzaking lijkt enkel mogelijk in een dualistisch raamwerk, en B&R kiezen resoluut voor een fysicalistisch raamwerk. Zie ook de claim van Wakefield, die geciteerd wordt in de tekst van B&R op p. 106, links bovenaan. Wakefield geeft aan dat een mentale toestand een psychische stoornis is dan en slechts dan als de toestand veroorzaakt wordt door een disfunctioneel psychisch mechanisme (een psychische oorzaak dus). Maar dat neemt niet weg, voegt Wakefield toe, dat de disfunctie van een psychisch mechanisme (nabije oorzaak, in de woordenschat van B&R) het gevolg kan zijn van een lichamelijke disfunctie (verre oorzaak, in de woordenschat van B&R).

Om te verduidelijken het voorbeeld van Anna O.

Anna O. was een van de patiënten die Breuer en Freud opvoerden in hun gezamenlijk werk *Studien Uber Hysterie* (1895). Anna (aka Bertha Pappenheim) was een jonge en intelligente vrouw, die gedurende meer dan twee jaar een reeks van zowel lichamelijke als psychische symptomen ontwikkelde, waaronder verlammingsverschijnselen (rechterarm & rechterbeen), verstoorde oogbewegingen, eetstoornissen, een nerveuze hoest, spraakstoornissen (soms was ze niet in staat om in haar moedertaal te spreken), en zogenaamde ‘absences’ – toestanden van afwezigheid en verwarring, waarin ze in zichzelf begon te mompelen. Anna O. vertoonde met andere woorden alle symptomen van een psychische stoornis die ondertussen de geest gegeven heeft, en geen rol van betekenis meer vervuld in de psychiatrie: hysterie.

Wat veroorzaakte deze symptomen volgens Breuer & Freud? Het is niet helemaal duidelijk, maar wel waarschijnlijk dat Anna O. door talrijke dokters onderzocht is voor ze kennismaakte met Breuer. Geen van die dokters vond een aanwijsbare organische oorzaak voor haar klachten. Tegenwoordig zou men zulke patiënten screenen voor neurologische problemen, zoals een hersenbloeding, een tumor of epilepsie, of endocrinologische afwijkingen, zoals een schildklierprobleem, enzovoort. Wanneer elk van die onderzoeken negatief blijkt (dat wil zeggen: wanneer er geen organische afwijkingen gevonden worden), verwijst men de patiënt door naar de psychiater. Zo verging het Anna O. wellicht ook. Haar lichamelijke symptomen (verlammingsverschijnselen, gezichtsstoornissen,…) wezen niet op een lichamelijke oorzaak, maar op een niet-lichamelijke, en dus psychische oorzaak, en daarom lijdt ze aan een psychische stoornis. Tegenwoordig spreekt men van een conversiestoornis – een DSM-categorie die, samen met bijvoorbeeld Illness Anxiety Disorder (voorheen hypochondrie) deel uitmaakt van de zogenaamde Somatic Symptom Disorders (voorheen Somatoform Disorders). In DSM-IV luidde een van de diagnostische criteria voor conversiestoornis: ‘The symptom or deficit cannot, after appropriate investigation, be fully explained by a general medical condition, or by the direct effects of a substance, or as a culturally sanctioned behavior or experience’.

Drie definities van het psychische

Het antwoord dat psychische stoornissen psychisch zijn omdat ze psychische oorzaken hebben is uiteraard louter een herformulering van het probleem. De vraag is immers wat er psychisch is aan de oorzaken van een psychische stoornis. Dat het oorzaken zijn die de ruggengraat vormen van een ‘typisch psychologisch’ theorie, zoals de psychoanalyse, verlegt het probleem opnieuw. Want dan is de vraag: wat maakt de concepten waarvan Freud zich bedient in ‘zijn’ psychoanalyse tot ‘psychologische concepten’. B&R onderscheiden drie invullingen van de term ‘psychisch’ in ‘psychische oorzaken’.

Een negatieve invulling: psychische oorzaken zijn oorzaken die niet-lichamelijk zijn. Anders geformuleerd: wanneer alle mogelijke lichamelijke oorzaken uitgesloten zijn, dan nemen we aan dat de oorzaak psychisch is. B&R geven twee argumenten in het voordeel van deze zienswijze. Beide argumenten zijn gebaseerd op de realiteit van de medische wereld. Ten eerste is het een dagelijkse realiteit in de medische wereld dat patiënten doorverwezen worden naar een psychiater wanneer de medische testen geen lichamelijke oorzaak opleveren. Ook vandaag nog zijn er mensen met gelijkaardige problemen als Anna O., getuige de ziektecategorie van Conversion Disorder in DSM. Ten tweede, en omgekeerd, is het een historische realiteit in de medische wereld dat psychische stoornissen van status veranderen, en neurologische stoornissen worden, wanneer ze gekoppeld worden aan een aanwijsbare lichamelijke oorzaak. Een voorbeeld: Encephalitis Lethargica is een ernstige aandoening die in het decennium na de eerste wereldoorlog een ware ravage aanrichtte in Europa. Een ouderwetse epidemie. Patiënten met EL veranderen in een ijltempo in levende standbeelden, niet meer in staat om te spreken of te bewegen. Men spreekt vaak van een ‘catatoon’ ziektebeeld, waarbij patiënten volledig verstijven, en jarenlang niet in staat zijn om te bewegen. Daarnaast worden ook vaak hallucinaties gemeld, soms zelfs comateuze toestanden. Mijn punt: sommige historici van de psychiatrie menen dat psychiaters tijdens de hoogdagen van de epidemie redeneerden volgens een logica die eigen is aan de zienswijze dat psychische oorzaken niet-lichamelijke oorzaken zijn. Zij kenden de oorzaak van de aandoening niet, en patiënten werden bijgevolg vaak opgenomen in psychiatrische ziekenhuizen. Toen ‘duidelijk’ werd dat EL veroorzaakt wordt door een zeldzame variant van het griepvirus, die tijdens de eerste wereldoorlog vele slachtoffers eiste, veranderde EL van een psychische stoornis in een neurologische stoornis. (Het voorbeeld van EL toont echter ook de grenzen aan van deze zienswijze, omdat er nog steeds geen consensus bestaat over de oorzaak van de aandoening. Blijkbaar is het *vermoeden* van een lichamelijke oorzaak voldoende om een ziekte te classificeren als een neurologische aandoening. Dat vermoeden is er nochtans ook bij schizofrenie, en die aandoening wordt niettemin nog steeds beschouwd als een prototypische psychische stoornis.)

Een invulling op basis van exemplarische voorbeelden (in de tekst van B&R: de sectie ‘Lists of examples’ op p. 106-7): psychische stoornissen zijn stoornissen waarvan alle symptomen veroorzaakt worden door psychische oorzaken, en prototypische voorbeelden van zulke oorzaken zijn verstoringen van psychologische mechanismen, zoals perceptie (in het geval van hallucinatie), emotie (in het geval van depressie), en cognitie (in het geval van schizofrenie). Maar waarom noemen we perceptie een *psychologisch* mechanisme? De hete aardappel wordt steeds doorgeschoven. Wittgensteinianen, zoals Graham, zullen zeggen dat perceptie een psychologisch mechanisme is omdat nagenoeg iedereen het daar over eens is (‘iedereen’ = ‘iedereen die een relevante rol vervult in het debat over perceptie’). Maar is consensus een noodzakelijke en voldoende voorwaarde om van een mechanisme een psychologisch mechanisme te maken? Het debat over de plaats van homoseksualiteit in de DSM (zie eerdere lessen) werd beslist met een referendum, waarbij een nipte meerderheid van de Amerikaanse psychiaters zich uitsprak tegen de psychiatrisering van homoseksualiteit. Sommige historici menen dat de psychiatrie zich belachelijk maakte door een referendum te organiseren over een wetenschappelijk probleem. Niemand haalt het in het hoofd om te stemmen over de vraag of de aarde aan het opwarmen is, wel? De feiten moeten over zo’n probleem beslissen. We willen niet weten hoeveel mensen voor of tegen een theorie stemmen, maar wel waarom ze dat doen – op welke feiten beroepen ze zich?

Een intentionele invulling: psychische oorzaken zijn oorzaken die moeten voldoen aan een belangrijke voorwaarde. De oorzaken moeten namelijk het kenmerk van intentionaliteit vertonen. We keren even terug naar Anna O., de patiënt van Breuer en Freud die daarnet al aan bod kwam. Anna O. lijdt onder andere aan een soort ‘eetstoornis’ – gedurende een bepaalde periode weigerde ze te drinken. Breuer & Freud spreken van ‘hydrofobie’. Tijdens een hypnose-sessie (een erg populaire vorm van psychotherapie in die tijd) komt aan het licht dat deze weigering te maken heeft met een herinnering aan een bezoek aan een vriendin met een hondje. Tijdens het bezoek ziet Anna O. het hondje stiekem uit een glas drinken. Het gedrag van de hond wekt walging en woede op, en deze twee emoties zijn volgens Breuer & Freud de motor of de oorzaak van Anna’s ‘hydrofobie’. En walging en woede zijn uiteraard intentionele mentale toestanden: de walg is een walg voor de lebberende hond, de woede is gericht op diezelfde hond (en de eigenaar van de hond). Een ander voorbeeld van een symptoom met een intentionele oorzaak: Anna’s verlammingsverschijnselen. Ik lees voor uit *Studien Uber Hysterie* (zie tekst op Toledo). De verlamming van haar rechterarm wordt volgens Breuer en Freud veroorzaakt door het verlangen om haar doodzieke en hulpeloze vader te verdedigen (in een droom) tegen de aanval van een slang. Anna’s verlangen werd echter gefnuikt doordat haar arm tijdens het indutten ‘in slaap gevallen was’ – ze kon haar arm niet gebruiken om haar vader te beschermen. Opnieuw: als de verklaring van Breuer & Freud klopt, dan is de oorzaak van het verlammingsverschijnsel een propositionele attitude, namelijk een verlangen, en PA’s zijn, zoals we weten, intentionele mentale toestanden. En intentionaliteit is, volgens sommigen, hét kenmerk van het mentale. Ook de problemen van de Rattenman kan je volgens Freud verklaren met behulp van allerlei intentionele mentale toestanden: liefde (voor zijn verloofde), haat (voor zijn vader), enzovoort. B&R lijken deze zienswijze (over de vraag waarom psychische oorzaken psychisch zijn) toegenegen te zijn. De enige opmerking die ze maken is dat het een zienswijze is met een beperkte reikwijdte: je kan er volgens hen niet mee verklaren waarom depressie en schizofrenie psychische stoornissen zijn (zie p. 107 links onderaan). De verborgen premisse in hun redenering is de volgende: niet alle vormen van een klinische depressie, en niet alle vormen van schizofrenie worden veroorzaakt door intentionele mentale toestanden. Anders geformuleerd: er zijn vormen van depressie die niet veroorzaakt worden door intentionele mentale toestanden. B&R geven geen voorbeeld, en het is een feit dat de psychiatrische wereld verdeeld is over de vraag of er depressies bestaan zonder psychische (intentionele) oorzaak (zie volgende slide).

In het algemeen vind ik deze laatste zienswijze, namelijk dat psychische stoornissen psychisch zijn omdat hun oorzaken intentionele mentale toestanden zijn, erg aantrekkelijk. Symptomen kunnen, op zichzelf genomen, bedrieglijk zijn. Psychische symptomen kunnen veroorzaakt worden door neurologische problemen (en omgekeerd kunnen psychische stoornissen lichamelijke symptomen voortbrengen, zoals verlammingsverschijnselen). Denk even terug aan Olga in *Turks Fruit*. Olga lijdt niet aan een psychische stoornis, zoals we eerst dachten (op basis van haar bizar gedrag), maar aan een neurologische stoornis, namelijk een tumor, en een tumor is een hersentoestand die helemaal nergens op gericht is, geen intentionaliteit heeft. Daarom, en enkel daarom, is Olga niet psychisch ziek. Haar symptomen geven geen uitsluitsel over de aard van haar ziekte. Enkel de oorzaak helpt ons om daarover te beslissen, en we spreken enkel over een psychische stoornis wanneer de symptomen veroorzaakt worden door intentionele mentale toestanden. Het grote probleem is dat sommige symtomen ‘overgedetermineerd’ zijn. Dat wil zeggen: sommige symptomen worden bepaald (veroorzaakt, zeg maar) door een reeks van oorzaken. Sommige van die oorzaken kunnen lichamelijk zijn (neurologisch, hormonaal, genetisch,…), en andere psychisch. In die zin is het moeilijk om een sluitend antwoord te geven op de vraag of een stoornis psychisch is, tenzij we een rekensommetje doen: hoeveel symptomen zijn er (en dan laten we nog buiten beschouwing dat ziekten de neiging hebben om te clusteren – een fenomeen dat bekend staat als ‘comorbiditeit’ – zodat het moeilijk wordt om te bepalen welk symptoom bij welke ziekte hoort); hoeveel oorzaken zijn er voor elk van deze symptomen; en hoeveel oorzaken zijn psychische oorzaken? En hoever gaan we terug in de tijd? Waar eindigen de ‘nabije’ oorzaken (in B&R’s terminologie) en waar beginnen de ‘verre’ oorzaken? Het lijkt niet zo eenvoudig om in de praktijk te beslissen of we te maken hebben met een psychische dan wel een neurologische stoornis.

Interessante vraag op p. 107 rechts onderaan: ‘Moreover, if neurophysiological research would discover the physical causes of some disorder that is at present regarded as prototypically mental, and if this would result in a highly effective drug treatment, this would not imply that we would no longer regard the disorder as mental’. Bent u het daarmee eens? Wat zegt uw intuïtie over dit probleem?

Bestaan er depressies zonder psychische (dat wil zeggen) intentionele oorzaak? Zie tekst van Antoon vergrote op toledo

Psychologische symptomen

Sommige filosofen oordelen dat niet de aard van de oorzaak van een stoornis, maar wel de aard van de symptomen van een stoornis bepalend is voor de aard van de stoornis. Zij menen dus dat de aard van de symptomen een stoornis al dan niet tot een psychische stoornis maakt. We spreken enkel van een psychische stoornis wanneer er sprake is van (voornamelijk) psychische symptomen. B&R: ‘A disorder is mental if the distress and/or disability involved is of a “mental kind”’ (p. 108).

b) De vraag is dan: wat maakt een symptoom tot een *psychisch* symptoom? Ook hier is het criterium van intentionaliteit de sterkhouder (zie ook B&R, p. 111, links midden, die zichzelf helaas klemrijden door symptomen van meetaf aan te definiëren als *disabilities –* de vraag is immers of een symptoom altijd een gevolg is van een disfunctioneel mechanisme). Een symptoom is een symptoom van een psychische stoornis dan en slechts dan als het een intentionele mentale toestand is. Neem het voorbeeld van een fobie. Een fobie is een (prototypische) psychopathologie die zich in de eerste plaats kenmerkt door een hevige angst voor een specifiek ding (een injectiespuit), een organisme (een specifiek dier of en stalker, bijvoorbeeld), of een setting (pleinvrees of agorafobie). Een fobische angst is intentioneel in de mate dat ze gericht is op iets anders dan zichzelf: het ding, het organisme, of de setting. De angst van de fobie is een intentionele mentale toestand en maakt een fobie daarom tot psychische stoornis.

c) Een van de grote problemen van deze zienswijze is dat ze psychiaters noopt tot een gigantische herziening van bestaande classificaties van psychische stoornissen. Heel wat psychische stoornissen in zulke classificaties uiten zich immers in lichamelijke symptomen. Fobie is een goed voorbeeld. Sommige fobische symptomen (en symptomen van een paniekstoornis) maken het moeilijk om te bepalen of er sprake is van een fobie of eerder van een hartaanval. Ook bij een fobische aanval klagen patiënten over pijn in de borststreek, flauwvallen, verstijving, enzovoort. Strikt genomen zijn deze symptomen, als lichamelijke symptomen, nergens op gericht. De pijn in de borststreek van fobische patiënten heeft geen (intrinsieke) intentionaliteit. Is een fobie dan geen psychische stoornis? Die conclusie lijkt absurd. Of moeten we opnieuw het relatieve belang van psychische versus lichamelijke symptomen gaan meten, en op basis van de uitkomst van die calculus besluiten of er sprake is van een psychische dan wel een lichamelijke stoornis? De welwillendheid van B&R tegenover deze zienswijze (i.e. *the symptom view*), is aandoenlijk, maar onterecht. Het vermeende voordeel van deze zienswijze, dat ze aanhalen op p. 111 links onderaan, geldt niet alleen voor de *symptom view (SV)*, maar ook voor de *internal cause view (ICV)*. Het argument luidt dat de SV ons toelaat om ruimte te maken voor nabije oorzaken van een neurologisch type. Anders geformuleerd: in de SV is het mogelijk dat symptomen van een psychische stoornis ook gedeeltelijk veroorzaakt wordt door hersentoestanden. Maar die gedachte kan je ook verwerken in de ICV, als je er vanuit gaat dat de meeste symptomen overgedetermineerd zijn

Een hopeloze zaak

P= psychisch en L = lichamelijk

* + P oorzaken 🡪 L stoornissen 🡪 P symptomen
  + L oorzaken 🡪 P stoornissen 🡪 L symptomen

Een vierde en laatste visie stelt dat het zinloos is om ook maar een poging te wagen om criteria te vinden die ons zouden toestaan om een strikt onderscheid te maken tussen psychische stoornissen en neurologische stoornissen. Zo’n zoektocht is zinloos, want die criteria bestaan niet. De keuze voor de kwalificatie ‘nihilistisch’ is een beetje eigenaardig, vind ik, en drukt het afgrijnzen van B&R uit ten opzichte van deze zienswijze. Ze hadden immers ook het woord ‘pessimisme’ kunnen gebruiken, zoals bijvoorbeeld McGinn doet om te verwijzen naar de hopeloze complexiteit van het probleem van lichaam en geest. B&R laten geen gelegenheid voorbijgaan om te benadrukken dat ze op zoek zijn naar ‘a coherent and nonarbitrary way’ (p. 111, rechts onderaan) om psychische stoornissen van neurologische stoornissen te onderscheiden. Ze zijn op zoek naar de essentie van het concept van psychopathologie, om daarmee de grens te trekken tussen psychopathologie en andere soorten van stoornissen. Tegelijkertijd herhalen ze ook enkele keren dat ‘mental and somatic disorders might not be mutually exclusive categories’ (p. 105, rechts onderaan; ook op p. 100, links onderaan; en p. 114, links bovenaan). Anders gezegd: het zou kunnen dat er overlappingen bestaan tussen beide categorieën (of soorten) van stoornissen. In dat geval is het echter zinloos om te spreken over de essentie van de ene of de andere categorie, of over noodzakelijke en voldoende voorwaarden, want het is eigen aan een essentie dat ze alle leden van een verzameling kenmerkt, en enkel de leden van die verzameling (en niet leden van andere verzamelingen). Kortom: er zit een zekere ambivalentie in de tekst van B&R.

Waarom zou het futiel zijn om te zoeken naar een strikt onderscheid tussen psychische stoornissen en neurologische (of, bij uitbreiding, lichamelijke) stoornissen? B&R geven twee argumenten.

Ten eerste hebben zowel psychische stoornissen als neurologische stoornissen zowel lichamelijke als psychische symptomen én oorzaken. Dat is een heleboel informatie in een enkele zin. Ik schematiseer de mogelijkheden in net onder het titeltje. Zie ook B&R, p. 112, midden links. Probeer zelf een aantal voorbeelden te bedenken van de verschillende opties. Bijvoorbeeld: de claim dat lichamelijke stoornissen psychische symptomen veroorzaken, kan geïllustreerd worden met behulp van de ziekte van Alzheimer (een neurologische stoornis met psychische symptomen, zoals depressie en verwardheid). Nog eentje: de claim dat lichamelijke oorzaken tot psychische stoornissen kunnen leiden, is een van de basisclaims van de biologische psychiatrie, en kan je plausibel maken door te verwijzen naar genen die mensen een duwtje in de rug geven om een welbepaald temperament te onwikkelen, waarbij dat temperament een rol speelt in de neiging om bijvoorbeeld angstig te reageren op nieuwe situaties. Er bestaat volgens mij niet zoiets als ‘genen voor angst’, tenzij men zo’n uitdrukking genuanceerd leest als ‘dingen die het een beetje waarschijnlijker maken dat iemand een beetje angstiger zal reageren op (bijvoorbeeld) nieuwe situaties dan anderen die zulke genen ontberen’. Wij zijn niet het slachtoffer van onze genen, zoals sommige populariserende wetenschapsboeken ons soms doen geloven.

Een tweede argument dat B&R aanhalen om het ‘nihilisme’ te verdedigen (p. 112 links onderaan) is dat er in de medische wereld geen theoretisch onderscheid gemaakt wordt tussen beide soorten van stoornissen. Of een stoornis beschouwd wordt als neurologische stoornis, dan wel als een psychische stoornis, hangt eerder af van pragmatische overwegingen, zoals: wie behandelt zulke stoornissen meestal? Wie is er bij gebaat of een stoornis als psychisch dan wel als neurologisch beschouwd wordt? Heeft de patiënt bijvoorbeeld meer kans om terugbetaald te worden voor een behandeling wanneer zijn of haar stoornis gepresenteerd wordt als een P dan wel een N stoornis? B&R brengen daar tegenin dat er in sommige situaties een pijnlijk verschil bestaat tussen hoe het proces van classificeren de facto verloopt, en hoe het zou moeten verlopen. Een bepaalde praktijk kan geen excuus zijn om alle theoretische reflectie te staken.

Merk op: vlak voor de conclusie, op p. 113 links bovenaan, leggen B&R hun kaarten op tafel. Hun antwoord op de vraag waarom psychische stoornissen *psychisch* zijn, weerspiegelt het antwoord van Graham op de vraag waarom psychische stoornissen *stoornissen* zijn. Wat is het antwoord van B&R? Het is plausibel, zeggen ze, dat er geen sluitende criteria zijn om te bepalen of een stoornis een *psychische* stoornis is. Of nog: er zijn geen ‘noodzakelijke en samen voldoende voorwaarden die ons kunnen helpen om psychische stoornissen van andere soorten van stoornissen te onderscheiden’ (letterlijke vertaling van passage op p. 113 bovenaan). En even verder, op het einde van diezelfde paragraaf: ‘it seems that the conditions we presently classify as mental disorders are systematically interconnected in ways that make some kind of family resemblance analysis reasonable’. Met andere woorden: B&R pleiten voor een wittgensteiniaans antwoord op de vraag waarom psychische stoornissen *psychisch* zijn. Het antwoord: omdat ze een gelijkenis vertonen met bepaalde prototypische *psychische* stoornissen. En hoe bepalen we welke stoornissen prototypische psychische stoornissen zijn? Je zou kunnen zeggen: sommige stoornissen hebben zowel hoofdzakelijk psychische oorzaken en hoofdzakelijk psychische symptomen, en zulke stoornissen definiëren we (op basis van consensus) als prototypische psychische stoornissen. Deze gedachte wordt nog eens herhaald op p. 113, rechts onderaan: ‘If one considers it important that … jointly sufficient’. Waarom B&R deze wittgensteiniaanse benadering als ‘nihilistisch’ bestempelen is me niet helemaal duidelijk. In mijn ogen is het een eerlijk en opbouwend antwoord op de hamvraag waarom sommige stoornissen ‘psychisch’ genoemd worden – een antwoord dat bovendien geen afbreuk doet aan de verwezenlijkingen van de twee heersende theorieën over het onderwerp in kwestie.

* 1. **Zijn psychische stoornissen natuurlijke soorten?**

In de eerste twee lessen van het drieluik over psychopathologie hebben we twee vragen beantwoord. We hebben een aantal antwoorden geformuleerd op de vraag waarom psychische stoornissen *stoornissen* zijn, en op de vraag waarom het *psychische* stoornissen zijn. Beide vragen hielpen ons om psychische stoornissen te onderscheiden van andere soorten van stoornissen (voornamelijk neurologische stoornissen) en van ‘de normaliteit’. In deze derde les van het drieluik over psychopathologie zoomen we een beetje uit, en zullen we ons afvragen hoe een classificatie van psychische stoornissen zich precies verhoudt tot classificaties van andere ‘dingen’ in de wereld. Denk bijvoorbeeld aan de classificatie waarmee biologen werken, waarin individuele dieren deel uitmaken van een soort; soorten deel uitmaken van een familie, families van een orde; ordes van een klasse; klassen van een stam; stammen van een rijk; en rijken van een domein. Of denk aan de bekende classificatie van chemische elementen die er in de werkelijkheid te vinden zijn – de zogenaamde Tabel van Mendeljev. Waarmee kunnen we psychische stoornissen best vergelijken? Is het concept ‘psychische stoornis’ vergelijkbaar met het concept ‘goud’ of ‘walvis’? Zijn de verschillende ziektecategorieën van de DSM vergelijkbaar met de categorieën die beschreven staan in de handboeken van biologen en chemici?

We kunnen de vraag ook stellen met behulp van een typisch filosofische term: natuurlijke soort (*natural kind*). ‘Goud’ en ‘walvis’ zijn volgens veel filosofen voorbeelden van natuurlijke soorten. Straks wordt duidelijk waarom dat zo is. Ik probeer eerst onze uitgangsvraag scherp te formuleren. De vraag luidt: zijn (categorieën van) psychische stoornissen natuurlijke soorten? En als het geen natuurlijke soorten zijn, wat zijn psychische stoornissen dan wel?

Let op: het onderwerp van deze les staat niettemin in verband met het onderwerp van de vorige les, want sommige filosofen beweren dat chemische en biologische classificaties vergelijkbaar zijn met neurologische classificaties. Zij beweren met andere woorden dat neurologische stoornissen, in tegenstelling tot psychische stoornissen, wel natuurlijke soorten zijn. Zachar lijkt zoiets te suggereren: ‘Schizophrenia and the personality disorders are not natural kinds in the way that Alzheimer’s disease is supposedly a natural kind’ (p. 168 rechts bovenaan). Het woord ‘supposedly’ (of ‘zogezegd’) gooit een beetje roet in het eten – bedoelt Zachar dat hij het oneens is met mensen die Alzheimer als een natuurlijke soort beschouwen? Uit het vervolg van het artikel zal blijken dat Zachar ook neurologische stoornissen niet als natuurlijke soorten beschouwt.

Bij wijze van springplank naar de centrale problematiek van vandaag, en ook als algemene achtergrond, geef ik een ruwe schets van twee paradigmatische denkstijlen over wat geestesziekten zijn (ook al wordt die vraag vaak verward met de vraag naar de oorzaak van geestesziekten): een essentialistische denkstijl, en een pragmatische (volgens Zachar) of een sociaal-constructivistische (volgens Hacking) denkstijl.

Daarna verbind ik deze denkstijlen met het filosofische denken over soorten of *kinds*. Ik doe een bescheiden poging om een classificatie van soorten in de filosofie op te stellen.

Vervolgens reik ik een aantal argumenten aan TEGEN de stelling dat geestesziekten natuurlijke soorten zijn, waarbij ik me concentreer op twee verschillende claims. Enerzijds de claim dat ze praktische soorten zijn, zoals Peter Zachar beweert in de begeleidende tekst – soorten die een rol kunnen spelen in de psychiatrische praktijk. Anderzijds de claim dat psychische stoornissen interactieve soorten zijn, zoals de filosoof Ian Hacking beweert. Daarmee bedoelt hij dat het soorten zijn die, in tegenstelling tot natuurlijke soorten, op een ingenieuze manier vormgeven aan de individuen die er deel van uitmaken, en die tegelijkertijd door die individuen vormgegeven wordt.

* + 1. **Inleiding: twee paradigma’s**

Essentialisme

De eerste denkstijl (of paradigma) noem ik een essentialistische denkstijl, omdat de voorvechters ervan (zoals de Amerikaanse psychiater Nancy Andreasen, die we eerder al ontmoetten) er impliciet of expliciet van overtuigd zijn dat geestesziekten een natuurlijke essentie hebben, en dat het mogelijk moet zijn om die ‘essentie’ bloot te leggen door middel van zogenaamde ‘harde’ wetenschappen, zoals de neurowetenschappen en de genetica. In deze denkstijl gebruiken ze graag de uitdrukking carving nature at the joints. Ze bedoelen hier mee dat ze da natuur kunnen ontleden bij haar naden (joints). Ze kunnen de essentie vatten door middel van concepten die ze analyseren aan de hand van voldoende en noodzakelijke voorwaarden.

Het is opvallend dat Zachar heel erg gekant is tegen een essentialistische benadering van psychische stoornissen. Hij vindt zo’n benadering niet gewoon verkeerd, hij vindt ze ‘malignant’, dat wil zeggen: kwaadaardig (p. 169 bovenaan; en p. 180). Waarom hij zo’n sterke term gebruikt is niet helemaal duidelijk. Hij lijkt te zeggen dat essentialisme eigenlijk een simplistische denkwijze is: ‘Reality is always going to be more complex than what is captured by our categories’ (p. 180).

(bekijk zeker ook de powerpoint van dit hoofstuk en meerbepaald de woordenwolken)

Pragmatisme/constructivisme

De tegenstanders van het essentialisme in het debat over psychische stoornissen worden op verschillende manieren benoemd (ook al hebben ze vaak een gelijkaardige achtergrond in de ‘zachte’ wetenschappen, zoals de filosofie, historische wetenschappen, culturele antropologie, enzovoort). Ik onderscheid twee vormen van anti-essentialisme: pragmatisme en sociaal-constructivisme.

Peter Zachar is een goed voorbeeld van een pragmatist inzake psychische stoornissen. ‘Pragmatisme’ definieert hij losweg als de stelling dat bruikbaarheid een belangrijke theoretische deugd is. U herinnert zich wellicht nog dat er verschillende theoretische deugden bestaan, zoals verklaringskracht, consistentie, vruchtbaarheid en vooruitgang (zie hoofdstuk over eliminativistisch fysicalisme). Bruikbaarheid vult het lijstje verder aan, en verwijst naar de mate waarin een theorie ons helpt om onze weg te vinden in de wereld (‘to help us navigate through the world’, zegt Zachar op p. 170). De DSM (het Amerikaanse psychiatrische handboek waarover we in de vorige lessen gesproken hebben) is volgens Zachar een goed voorbeeld van een theorie waarin betrouwbaarheid (‘reliability’) en bruikbaarheid centraal staan. Daarmee bedoelt hij dat het handboek gebruikers toestaat om op een betrouwbare manier psychische stoornissen te identificeren. Psychiaters uit verschillende hoeken van de wereld, met verschillende staat van dienst, verschillende interesses, en zelfs een verschillende onderliggende ideologie (biologisch-psychiatrisch versus psychoanalytisch, bijvoorbeeld) zullen eenzelfde patiënt eenzelfde diagnose toekennen wanneer ze de richtlijnen en diagnostische criteria in het handboek nauwkeurig toepassen. De DSM bevordert dus de communicatie tussen psychiaters onderling, en tussen psychiaters en andere actoren in de geestelijke gezondheidszorg. Het is dus een nuttig instrument, zegt Zachar. (Dat is overigens niet altijd zo geweest. De eerste twee edities van DSM (DSM-I, 1952; DSM-II, 1968) bevatten enkel een rudimentaire definitie en een korte beschrijving van elke ziektecategorie. Of een patiënt aan die beschrijving beantwoordde werd overgelaten aan de ‘willekeur’ van de psychiater. De betrouwbaarheid van DSM-I en DSM-II was bijgevolg niet erg groot. Zo bleek uit een groot internationaal onderzoek, het *US/UK Diagnostic Project* (Cooper *et al.* 1972), dat de diagnose ‘schizofrenie’ haast tweemaal vaker door Amerikaanse psychiaters gesteld werd dan door hun Britse collega’s. Een jaar later organiseert de *World Health Organisation* een vervolgstudie,*The International Pilot Study of Schizophrenia* (1973). Daaruit bleek dat er nog steeds geen eensgezindheid bestond over de diagnose ‘schizofrenie’, zelfs niet wanneer alle psychiatrische teams gebruikmaakten van dezelfde diagnostische criteria, die vooraf door een computerprogramma waren opgesteld (Claridge 1985, 105).

Vervolgens is het opvallend dat Zachar de DSM-benadering van psychische stoornissen typeert (en looft) als een *Wittgensteiniaanse* benadering. De verzameling van psychische stoornissen is een verzameling met onduidelijke grenzen. Er zijn prototypes, vreemde vogels en randgevallen (p. 169, rechts bovenaan), die verbonden worden door familiegelijkenissen, en niet door lijstjes van voldoende en noodzakelijke voorwaarden. Zachar is daarmee de derde Wittgensteiniaan in ons drieluik over psychopathologie. Graham outte zich als Wittgensteiniaan in het debat over wat psychische stoornissen tot stoornissen maakt, terwijl Brülde & Radovic een Wittgensteiniaans antwoord gaven op de vraag wat er psychisch is aan psychische stoornissen. Net als Graham legt Zachar uit waarom pragmatisme geen relativisme hoeft te zijn (zie p. 170 links onderaan & rechts bovenaan). Een Wittgensteiniaanse benadering van psychische stoornissen houdt niet in dat eender welke toestand gebombardeerd kan worden tot psychische stoornis. Net zoals een Wittgensteiniaanse benadering van biologische soorten (diersoorten en plantensoorten) niet in chaos hoeft te eindigen (er is bijvoorbeeld niemand die orangutans als een reptielensoort zou classificeren, zegt Zachar – dat zou ‘whimsical’ (idioot) zijn; p. 174 links). We zullen dadelijk zien dat psychische stoornissen volgens Zachar geen natuurlijke soorten zijn, maar *praktische soorten*.

De Frans-Canadese filosoof Ian Hacking is dan weer een goed voorbeeld van een constructivist – al moet ik er eerlijkheidshalve bij vermelden dat Hacking het daar niet helemaal mee eens zou zijn. De term ‘sociaal-constructivisme’ is in de voorbije decennia zo vaak gebruikt, zegt Hacking, dat niemand nog weet wat hij (de term) betekent. Vaak verwijst de term naar een of andere theorie waarin geestesziekten niet in de genen geschreven staan, maar ingebed zijn in allerlei complexe sociale, culturele en economische netwerken. Hacking wil voorbij zulke vage claims. Psychische stoornissen ‘sociale constructen’ noemen is niet voldoende – we moeten precieser zijn. We moeten ons de vraag durven stellen wat er geconstrueerd wordt (symptomen of syndromen?), wat construeren precies inhoudt, en wie er construeert. Hacking is erg geïnteresseerd in deze laatste vraag. Hij beschouwt psychische stoornissen als *interactieve soorten* – het zijn categorieën die oprijzen uit de veelvuldige en complexe interacties tussen, onder andere, psychiaters en hun patiënten (Nota: een van de meest extreme vormen van sociaal-constructivisme is de zogenaamde ‘anti-psychiatrie’ – een beweging die furore maakte in de jaren 1960, als reactie tegen de gedachte dat psychische stoornissen medische stoornissen zijn. Denk daarbij aan de Amerikaanse psychiater Thomas Szasz, die eerder al aan bod kwam in de context van het fysicalisme (les 4, slide 19)).

*‘Constructionisten beweren dat classificaties niet bepaald worden door hoe de wereld is, maar dat ze handige manieren zijn om de wereld voor te stellen. Ze beweren dat de wereld niet verpakt wordt in feiten (‘wrapped up in facts’). Feiten zijn de gevolgen van de manier waarop we de wereld voorstellen (in classificaties)’ (Hacking 1999, 33).*

* + 1. **Soorten van soorten**

In het debat tussen ‘essentialisten’ en ‘pragmatisten’ of ‘sociaal-constructivisten’ staat de notie ‘natuurlijke soort’ vaak centraal. In wat volgt zal ik vier soorten van soorten bespreken: natuurlijke soorten, arbitraire soorten, praktische soorten en interactieve soorten

Natuurlijke soorten

Zachar zegt: ‘Thinkers who believe in the widespread existence of natural kinds are called essentialists’ (p. 167). Om te weten wat essentialisme inhoudt moeten we dus eerst nagaan wat natuurlijke soorten precies zijn. Een groep van individuen, organismen, entiteiten of eigenschappen is een natuurlijke soort als ze van nature een groep vormen. Ze moeten met andere woorden vaste, interne en natuurlijke eigenschappen delen, die het dan mogelijk maken om ze op een betrouwbare manier te identificeren, en om ze van andere, verwante entiteiten te onderscheiden. Typische voorbeelden van natuurlijke soorten zijn chemische elementen (zoals bijvoorbeeld goud), biologische soorten (planten- en diersoorten, zoals bijvoorbeeld tijger) en fysische entiteiten (zoals bijvoorbeeld quark). Een goudatoom is een element uit de befaamde tabel van Mendelejev dat uit precies 79 protonen bestaat – een uniek definiërende natuurlijke eigenschap. Het hebben van 79 protonen is een *vast* kenmerk van een goudatoom – het is een kenmerk dat niet gewijzigd is sinds de oerknal. ‘Vast’ staat in deze context tegenover ‘variabel’. Verder is het hebben van 79 protonen een *intern* of *inherent* kenmerk, omdat het deel uitmaakt van de structuur van een goudatoom. Elk goudatoom bestaat uit 79 protonen, op eender welk moment (vast) en in eender welke omstandigheden (intern). De essentie van het goudatoom ligt vervat in het goudatoom zelf. Geen gezeur over extimiteit en excentrische identiteiten in de chemie – chemische elementen ontlenen hun identiteit aan zichzelf, en niet aan iets anders (wat volgens sommige filosofen wel het geval is bij de mens). Een ander voorbeeld is controversiëler, zoals we dadelijk zullen aantonen: tijgers. Sommige filosofen zeggen dat de tijgers op aarde ook een natuurlijke soort vormen, omdat ze hun erfelijk materiaal met elkaar delen. Ze delen onderling in ieder geval meer erfelijk materiaal dan met andere diersoorten uit de familie van de katachtigen, zoals de leeuw en de jaguar. Erfelijk materiaal is zeker een natuurlijk kenmerk, maar is het ook een vast en intern kenmerk? Zachar brengt in zijn tekst een aantal argumenten aan om het tegendeel te beweren (zie verderop).

Sommige filosofen beweren dat er ook in de menswetenschappen natuurlijke soorten bestaan – ze noemen die dan ‘natuurlijke menselijke soorten’. Bloedgroepen zijn goede voorbeelden van natuurlijke soorten. Ze laten ons immers toe om mensen op een eenduidige manier te groeperen op basis van een vaste, interne en natuurlijke eigenschap, met name hun bloedgroep (A, B, AB en O). Een eenvoudige bloedtest wijst uit tot welke bloedgroep je behoort. Daarnaast kan je mensen bijvoorbeeld ook indelen in natuurlijke soorten op basis van erfelijke kenmerken, zoals het verschil tussen geslachtschromosomen. Met een eenvoudige genetische test kan je uitzoeken of iemand twee X-chromosomen heeft of een combinatie van een X- en een Y-chromosoom, en kan je met een behoorlijke zekerheid voorspellen welke primaire en secundaire geslachtskenmerken een individu zal ontwikkelen.

Waarom is een bloedgroep een natuurlijke soort? Omdat elke bloedgroep gedefinieerd wordt door de aanwezigheid van specifieke antigenen die zich hechten aan het celmembraan van rode bloedcellen. Antigenen zijn een goed voorbeeld van vast, interne en natuurlijke kenmerken.

Let op: bloedgroep A is een natuurlijke soort, bloedgroep B is een natuurlijke soort, enzovoort. Een bloedgroep is een categorie, een verzameling van individuen waarvan de bloedcellen op een welbepaalde manier gevormd zijn.

Maar wat betekent het concreet wanneer we zeggen dat X een natuurlijke (menselijke) soort is? De stelling dat X een natuurlijke soort is is meestal gebaseerd op drie voorwaarden:

Ten eerste veronderstelt ze dat X totaal onverschillig is ten opzichte van onze theorieën over X. De verzameling van quarks in de wereld is een natuurlijke soort, onder andere omdat geen enkele quark zich ook maar iets aantrekt van wat we over hem denken of zeggen. Een quark verandert niet door hem een quark te noemen. Zoals de Canadese filosoof Ian Hacking zegt: ‘The classification “quark” is indifferent in the sense that calling a quark a quark makes no difference to the quark’ (I. Hacking, *The Social Construction of What?*, p. 105*).* Hetzelfde geldt uiteraard voor bloedgroepen of chromosomale combinaties.

Een tweede veronderstelling van de claim dat X een natuurlijke soort is houdt in dat X een homogene en begrensde categorie vormt. Anders gezegd: alle leden van X (homogeniteit), en enkel de leden van X (begrensdheid), kenmerken zich door een unieke eigenschap die hen onderscheidt van leden van Y. Een voorbeeld: alle goudatomen bestaan uit 79 protonen, terwijl alle platina-atomen uit 78 protonen bestaan, en alle kwik-atomen uit 80 protonen. Ook bloedgroepen kan je op een eenduidige manier identificeren en van andere bloedgroepen onderscheiden. De bloedgroep wordt namelijk bepaald door de aanwezigheid van specifieke antigenen die zich hechten aan het celmembraan van rode bloedcellen. En nogmaals: de kenmerken die ons toestaan om de leden van X te identificeren en te onderscheiden van leden van Y zijn vaste en interne kenmerken.

Een derde en laatste voorwaarde luidt dat de unieke eigenschap die alle leden van X gemeenschappelijk hebben een natuurlijke eigenschap moet zijn. Protonen, chromosomen en antigenen zijn goede voorbeelden van zulke natuurlijke eigenschappen.

Arbitraire soorten

Het is duidelijk dat niet alle verzamelingen van dingen in de werkelijkheid *natuurlijke* verzamelingen zijn. Zulke niet-natuurlijke verzamelingen zou je ook ‘sociale soorten’ kunnen noemen. Het is natuurlijk een zeer brede categorie, en daarom op zichzelf genomen niet erg interessant. Kunnen we verschillende soorten van sociale soorten onderscheiden? Zachar suggereert van wel. Hij maakt een onderscheid tussen aan de ene kant ‘artificial kinds’, ‘pseudo-kinds’, en ‘arbitrary kinds’ (drie synoniemen), en aan de andere kant ‘practical kinds’, die volgens hem soms bruikbaar, soms onbruikbaar zijn, maar niet arbitrair (p. 168). In een voetnoot nuanceert hij het onderscheid tussen beide soorten van soorten (‘artificial’ versus ‘practical’), met de opmerking dat ‘artificial kinds’ eigenlijk gewoon onbruikbare ‘practical kinds’ zijn (p. 180, voetnoot 1). Daarmee spreekt hij zichzelf tegen, want elders in de tekst beweert hij dat de eigenzinnige classificatie die de jongen maakt in de papieren van zijn vader (p. 175-6) ‘impractical’ is, maar niet ‘arbitrary’. De vader had zijn documenten geordend volgens thema. Er was een map met documenten die te maken hadden met onderzoeksadministratie (projectaanvragen, contracten van doctoraatsstudenten,…), een map met onderwijsadministratie (ziektebriefjes van studenten, examenformulieren,…), een map met wetenschappelijke artikels over onderwerp X, enzovoort. De jongen verving de ordening of classificatie van zijn vader door een nieuw classificatiesysteem, dat hij baseerde op de kleur en de vorm van het papier. Donker (gerecycleerd) A4-papier belandde op een hoop, licht (niet-gerecycleerd) A4-papier een andere hoop, enzovoort (A5-papier, papier uit notaboekjes, gelijnd & geruit papier,…). Volgens Zachar is deze classificatie niet arbitrair – de verschillende hoopjes waarmee de jongen eindigt zijn hoopjes van papieren die op een of andere manier op elkaar lijken (zelfde kleur, zelfde vorm,…). Ze delen dus minstens één kenmerk. Een arbitraire classificatie of verzameling is een verzameling van entiteiten, organismen en individuen die *geen enkel* kenmerk met elkaar delen.

Maar bestaan er zulke (arbitraire) verzamelingen? Een vuilnisbelt, misschien? Of een hoop rommel die door een orkaan in een zwembad beland is: stoelen, dakpannen, geiten, een hoogspanningskabel? Maar zelfs zulke hoogst toevallige samenraapsels delen nog een kenmerk, met name dat ze door de orkaan in het zwembad geworpen zijn. Idem dito met de waanzinnige verzameling spulletjes in de handtas van mijn vrouw: kruimels van koekjes, een geplette en overrijpe banaan, een agenda van het jaar 2010, potloodstompjes, een kam, lippenstift, autosleutels,… De gemene deler in deze veelheid is dat ze door mijn vrouw in de tas geduwd zijn. Het is een leuke denkoefening: kan je een verzameling van dingen bedenken die geen enkel kenmerk delen? Zachar typeert arbitraire soorten in de eerste plaats op een negatieve manier (ook al doet hij dat niet op een consequente manier, zie hierboven): het zijn onbruikbare soorten. Zachar definieert de term ‘onbruikbaar’ niet, maar het lijkt te maken te hebben met ‘wetenschappelijk nut’ (zie volgende slide). Arbitraire soorten zijn categorieën die geen enkele rol kunnen spelen in een wetenschappelijke classificatie. Maar ook die definitie helpt ons niet veel verder. Wetenschappers bestuderen de gekste categorieën. De verzameling van gescheiden vaders in het arrondissement Leuven, bijvoorbeeld. Is die verzameling of soort een arbitraire soort? Ik ken niemand die onderzoek doet naar deze populatie, maar de kans is reëel dat een of andere student sociologie bezig is aan een thesis over het onderwerp – waarin onderscheiden deze vaders zich van gelukkig getrouwde vaders, bijvoorbeeld? Dan maar een stapje verder: wat met de categorie van gescheiden vaders in het arrondissement Leuven, waarvan de achternaam rijmt op ‘toespijs’? Is dat een arbitraire soort? Het hangt er een beetje vanaf hoe je het concept ‘arbitraire soort’ definieert. Als verzameling van ‘elementen’ die geen enkel kenmerk delen? In dat geval is het antwoord negatief (want alle individuen van de verzameling zijn mensen). Als verzameling of categorie die geen rol kan spelen in wetenschappelijk onderzoek? In dat geval misschien wel.

Praktische soorten

Zachar breekt in zijn tekst een lans voor een derde soort van soort: de praktische soort (‘practical kind’). Hij zegt: ‘Practical kinds are fuzzier than natural kinds, but they are not arbitrary’ (p. 168, rechts onderaan). Praktische soorten zitten dus ergens tussen arbitraire en natuurlijke soorten in. En ook al is het onderscheid tussen praktische en arbitraire soorten misschien onscherp, het onderscheid tussen praktische en natuurlijke soorten is, volgens Zachar tenminste, haarscherp. Natuurlijke soorten zijn namelijk hoogst zeldzaam, zegt Zachar. De verzameling van natuurlijke soorten is heel klein geworden sinds Darwin – er voldoen nauwelijks nog dingen in de werkelijkheid aan de drie strikte voorwaarden die ik daarnet noemde. De elementen in de tabel van Mendeljev uiteraard, maar verder? Zachar wijdt een groot deel van zijn tekst aan het verdedigen van de stelling dat ziekten (in de brede zin van het woord, ‘diseases’) en biologische soorten (‘species’) geen natuurlijke soorten zijn. Hij gebruikt daarbij eigenlijk twee verschillende argumenten. Enerzijds toont hij aan dat ziekten en biologische soorten niet voldoen aan alle voorwaarden waaraan een verzameling moet beantwoorden om een natuurlijke soort te zijn. Anderzijds is er ook een positief luik, waarin Zachar aantoont dat ziekten en biologische soorten desondanks (dat wil zeggen ondanks het feit dat ze geen natuurlijke soorten zijn) een belangrijke rol kunnen spelen in de wetenschappen. Praktische soorten zijn praktisch in de mate dat ze bruikbaar zijn. Het is niet meteen duidelijk wat Zachar bedoelt met ‘bruikbaarheid’ (‘usefulness’). Sommige filosofen menen dat natuurlijke soorten zich onderscheiden van andere soorten van soorten door hun wetenschappelijke bruikbaarheid. De Australische filosoof Paul Griffiths, bijvoorbeeld, definieert ‘natural kind’ als ‘a productive object of scientific inquiry’. Een natuurlijke soort is een categorie die een rol kan spelen in wetenschappelijk onderzoek. Griffiths’ definitie zou volgens Zachar te weinig specifiek zijn, want psychische stoornissen zijn wel degelijk het voorwerp van ernstig (en nuttig) wetenschappelijk onderzoek, en toch zijn het geen natuurlijke soorten.

Waarom zijn ziekten en biologische soorten geen natuurlijke soorten? Om die claim van Zachar te begrijpen, moeten we ons concentreren op twee van de drie kenmerken die leden van een natuurlijke soort met elkaar delen: vaste en interne kenmerken. Beginnen we met ‘interne kenmerken’. Het hebben van 79 protonen is een intern of inherent (letterlijk: ‘eigen aan’ of ‘innig verbonden aan’) kenmerk van elk goudatoom; het is dit kenmerk dat er voor zorgt dat goud de eigenschappen heeft die het heeft (kleur, hardheid,…). De setting waarin het atoom zich bevindt speelt geen enkele rol bij het identificeren van het atoom. Het is niet zo dat een atoom met 79 protonen in een radioactieve omgeving plots andere kenmerken zal vertonen – goud blijft goud, *no matter what*. Daarin ligt meteen een belangrijk verschil tussen goudatomen en (medische) ziekten, zoals infectieziekten. Neem syfilis als voorbeeld. Syfilis wordt veroorzaakt door de bacterie *Treponema Pallidum*, de spirocheet die wellicht verantwoordelijk was voor Nietzsches trieste laatste levensjaren. Je zou kunnen argumenteren dat syfilis een natuurlijke soort is, waarvan de essentie bepaald wordt door de bacterie in kwestie – net zoals het hebben van 79 protonen de essentie uitmaakt van een goudatoom. Maar die redenering klopt niet, zegt Zachar, want infectieziekten worden niet enkel bepaald door pathogenen (zoals bacterieën en virussen), maar door de interactie tussen die pathogenen en hun dragers (de patiënt). ‘Infections exist in hosts, and disease processes result from the interaction between infection and host. *They are relational rather than inherent properties*’ (p. 171, rechts onderaan, mijn cursivering). Het symptoombeeld en het verloop van de ziekte, en zelfs de werkzaamheid van een behandeling voor de ziekte, zijn in mindere of meerdere mate afhankelijk van kenmerken die niet inherent zijn aan de bacterie, maar wel, bijvoorbeeld, aan de manier waarop het lichaam reageert op de bacteriële invasie. Ziekten zijn praktische soorten, zegt Zachar, omdat patiënten met een bepaalde ziekte niet louter interne kenmerken delen, maar ook relationele kenmerken. Deze ‘eigenaardigheid’ maakt ziektecategorieën als syfilis niet minder bruikbaar in wetenschappelijk onderzoek. Infectieziekten kunnen immers op een zeer betrouwbare manier geïdentificeerd worden (‘reliability’), en hun verloop kan min of meer voorspeld worden (‘predictive validity’; p. 172). Homoseksualiteit is een ander voorbeeld van de claim dat praktische soorten gedefinieerd worden met behulp van relationele kenmerken. Er zijn mensen die beweren dat minstens een deel van de verzameling van homoseksuele individuen op aarde een natuurlijk kenmerk delen (een biologische ‘marker’, zoals dat dan heet), maar dat kenmerk is niet voldoende om de categorie ‘homoseksualiteit’ te definiëren. Die definitie wordt namelijk gedeeltelijk ingevuld door hoe een samenleving zich verhoudt tot homoseksuelen. Anders gezegd: wat homoseksualiteit is, wordt niet alleen bepaald door natuurlijke kenmerken, maar ook door relationele kenmerken (d.i. kenmerken die voortvloeien uit de interactie tussen het individu en diens omgeving).

Dan een tweede verschil tussen natuurlijke en praktische soorten: variabele versus vaste kenmerken. Ik zei daarnet dat sommige filosofen van mening zijn dat er nauwelijks nog natuurlijke soorten overblijven sinds Darwin zijn gedachten op papier zette. Anders gezegd: het darwinisme gaf de doodsteek aan een lange traditie van essentialisme in de filosofie, waarvan Plato uiteraard een van de bekendste vertegenwoordigers was. In een eerdere les (les 6, slide 20) legde ik het verschil uit tussen twee manieren van denken over biologische soorten. Enerzijds is er het typologische denken, anderzijds het populatie-denken. Het typologische of essentialistische gedachtegoed houdt in dat er een vast aantal onveranderlijke types of ideeën schuilgaat onder de wilde veelheid van fenomenen in de werkelijkheid. Enkel deze ideeën zijn reëel; de rest zijn louter afschaduwingen van de ideeën. Plato’s ideeën kan je begrijpen als de essenties van de dingen in de werkelijkheid – het zijn eeuwige, onveranderlijke en volmaakte vormen. Het zijn de sjablonen waarmee de demiurg werkte toen hij de materie vormgaf (in Plato’s kosmologie). Samen met de vele andere aspecten van Plato’s filosofie, drong zijn doorgedreven essentialisme moeiteloos door in de Christelijke wereld. Christelijke filosofen moesten een beetje knoeien met de details (in tegenstelling tot de God van de christenen heeft de demiurg de wereld niet geschapen, maar enkel geboetseerd naar het beeld van de eeuwige ideeën), maar het essentialisme bleef behouden: God heeft elk van de biologische soorten (plantsoorten en diersoorten) op aarde geschapen; hij voorzag de oerkoppels van de nodige kenmerken, en die zijn sindsdien niet meer veranderd. Generaties komen en gaan, maar de essentie van elke soort blijft behouden. Darwins populatiedenken maakte definitief komaf met deze essentialistische denkwijze. Ten eerste deed Darwin dat door de radicale uniciteit van ieder levend wezen te benadrukken. Er is zoveel variatie in de levende wereld, zegt Darwin, dat geen twee organismen identiek zijn. Populaties zijn verzamelingen van individuen, en enkel deze individuen zijn reëel. En het is onmogelijk een soort te vinden waarvan alle leden, en enkel deze leden, een reeks van vaste, interne en natuurlijke kenmerken delen. Zachar formuleert het zelf als volgt: ‘Rather than what an essentialist would call “imperfections”, “errors”, or “accidents”, individual variation is central for understanding the long-term fluid nature of species’ (p. 174). Ten tweede toonde Darwin aan dat soorten evolueren – de kenmerken van een verzameling van individuen veranderen doorheen de tijd. Wanneer we lang genoeg wachten zien we dat soorten zich opsplitsen, wanneer een groep van individuen voldoende ver verwijderd is van de ‘moedergroep’ (in afstand of in uiterlijke kenmerken). Op die manier vormen zich voortdurend nieuwe soorten. Mayr formuleert het verschil tussen Plato en Darwin als volgt: ‘the ultimate conclusions of the population thinker and of the typologist are precisely the opposite. For the typologist, the type (eidos) is real and the variation an illusion, while for the populationist the type (average) is an abstraction and only the variation is real. No two ways of looking at nature could be more different’ (Mayr 1975, 27). Als Darwin gelijk heeft (en daar gaan we nu even van uit), dan zijn biologische soorten inderdaad geen natuurlijke soorten.

De implicatie van deze variabiliteit van biologische soorten is dat het mogelijk is om verschillende classificaties op te stellen van diersoorten en plantensoorten. Twee classificaties kunnen soms zelfs verrassend verschillend zijn, zegt Zachar, en hij geeft het voorbeeld van fenetische en cladistische classificaties van biologische soorten (p. 174, links bovenaan). Een fenetische classificatie is een classificatie waarbij soorten geordend worden met behulp van gelijkenissen – uiterlijke gelijkenissen, en/of genetische, moleculaire of anatomische gelijkenissen. Een cladistische classificatie kijkt niet zozeer naar de gelijkenissen tussen soorten om ze te ordenen, maar naar hun fylogenetisch traject – hun evolutionair verleden. Twee soorten belanden dan in eenzelfde groep omdat ze een gemeenschappelijke voorouder hebben. Geen van deze twee classificaties is dé classificatie – ze hebben allebei hun nut bewezen in het wetenschappelijk onderzoek van biologen. Het zijn eigenlijk twee manieren om naar de variatie in de levende wereld te kijken, en er structuur in aan te brengen. De bruikbaarheid staat daarbij centraal – Zachar is een pragmatist, remember. Waarin verschillen biologische soorten (plantsoorten en diersoorten), als praktische soorten (in Zachars terminologie) van natuurlijke soorten? Zachar geeft zelf het antwoord: ‘The very possibility of the evolution of species contradicts the idea of *a fixed inherent structure* that defines all members of a species’ (p. 174, mijn cursivering). Biologische soorten kan je niet typeren met behulp van vaste interne kenmerken, maar enkel met behulp van variabele kenmerken – er is immers zoveel variatie in een bepaalde soort, op een bepaald moment (synchroon) én doorheen de tijd (diachroon). Plantsoorten en diersoorten zijn variabel, en ze veranderen, dat is het punt. Daarom zijn het geen natuurlijke soorten. Daarentegen zijn het ook geen arbitraire soorten. ‘Species are not just inventions’, zegt Zachar, ‘the distinctions accepted by taxonomists are far from arbitrary’ (p. 175).

Interactieve soorten

Het unieke aan sommige niet-natuurlijke (menselijke) soorten is dat de leden van zulke soorten niet onverschillig zijn ten opzichte van onze theorieën over die soorten. Dat klinkt een beetje ingewikkeld, maar het betekent gewoon dat (groepen van) mensen zich vaak bewust zijn van de manier waarop ze geconceptualiseerd worden, en dat ze hun doen en laten (en hun zelfervaring) daarop afstemmen. Mensen veranderen wanneer onze theorieën over hen veranderen. En soms veranderen die theorieën omdat de mensen waarover die theorieën handelen veranderen. Er is met andere woorden sprake van een interactie tussen theorieën en mensen, en daarom noemt men sommige sociale menselijke soorten ook wel ‘interactieve soorten’. Hacking noemt zulke soorten ‘interactieve soorten’, omdat ze geconstrueerd worden in en door het krachtenspel tussen aan de ene kant een theorie of een classificatie, die op haar beurt ook een product is van de interactie tussen verschillende actoren (psychiaters, verzekeringsmaatschappijen, farmaceutische bedrijven,…), en aan de andere kant het voorwerp van die theorie – in casu de leden van de verzameling in kwestie.

Een van Hackings bekende voorbeelden van een interactieve soort is de ziektecategorie ‘multiple personality disorder’ (MPD). Het looping effect dat optreedt in het geval van MPD kan je als volgt weergeven:

1. Het concept van MPD en de term worden geïntroduceerd in de psychiatrie. Soms heeft zo’n concept een aantal begeleidende (theoretische) overtuigingen, maar dat is niet noodzakelijk.
2. Sommige mensen krijgen de diagnose MPD toegeschreven, en worden overeenkomstig behandeld.
3. Sommige van deze ‘patiënten’ (want dat zijn ze dus ondertussen) identificeren zich met de nieuwe diagnose die ze gekregen hebben (op een bewust of onbewust niveau; het wordt een ‘tweede natuur’), en bevestigen aldus de diagnose; anderen verzetten zich tegen de diagnose, vechten hun diagnose aan.
4. In het tumult (identificatie & verzet) worden ‘typische’ eigenschappen van het concept van MPD afgegooid of bijgesteld (bijvoorbeeld het gemiddelde aantal persoonlijkheden), en worden nieuwe eigenschappen verworven.
5. Deze verdwenen en nieuwe eigenschappen dwingt psychiaters ertoe om hun concept van MPD aan te passen, alsook hun overtuigingen over MPD

Er is dus sprake van twee bewegingen. Enerzijds beïnvloeden we een fenomeen, en dwingen we veranderingen af; anderzijds evolueert het fenomeen, als antwoord op deze beïnvloeding, in een bepaalde richting, waardoor we onze opvattingen over het fenomeen in kwestie moeten herzien. De filosofische waarde van Hackings theorie over interactieve soorten ligt hierin dat ze kanttekeningen plaatst bij de gedachte dat classificaties enkel dienen om fenomenen te beschrijven en te verklaren, en dat het proces van classificeren op geen enkele manier invloed uitoefent op die fenomenen. Classificeren, zegt Hacking, is meer dan beschrijven; het is ingrijpen.

* + 1. **Wat met schizofrenie?**

In het derde luik van deze les wil ik nagaan waar psychische stoornissen eigenlijk thuishoren in deze classificatie van soorten. Zijn psychische stoornissen natuurlijke soorten? Of misschien eerder arbitraire soorten? Of interactieve soorten? Praktische soorten? Het veld van psychische stoornissen is behoorlijk groot, dus ik richt mijn pijlen op een van de psychische stoornissen waarvan Graham zei dat het een prototypische soort is: schizofrenie. Ik begin met een weerlegging van de gedachte dat schizofrenie een natuurlijke soort is. Mijn kritiek leidt tot de suggestie dat schizofrenie eerder een interactieve soort is. Daarna bespreek ik enkele filosofen die schizofrenie eerder als een arbitraire of een praktische soort zouden zien.

Sommige psychiaters, voornamelijk biologische psychiaters, beweren dat schizofrenie een natuurlijke soort is. Denk even terug aan de eerste paragraaf van het drieluik over psychopathologie waarin ik de *broken brain analysis* van het concept van psychopathologie beschreef. Deze analyse stelt dat er slechts één voorwaarde is waaraan een toestand moet voldoen om een stoornis genoemd te worden, en deze voorwaarde is tegelijkertijd voldoende en noodzakelijk. De voorwaarde: er moet sprake zijn van een hersenstoornis. Veel psychiaters menen dat een hersenstoornis een goed voorbeeld is van een vast, intern en natuurlijk kenmerk, en dat psychische stoornissen daarom natuurlijke soorten zijn. Schizofrenie wordt inderdaad vaak voorgesteld als een hersenstoornis – een stoornis in de chemische huishouding van de hersenen. Medicatie zou deze ontregeling tegengaan, en het feit dat de meeste antipsychotica (geneesmiddelen die gebruikt worden in de behandeling van psychotische stoornissen, zoals schizofrenie) ernstige nevenwerkingen hebben (libidoverlies, motorische problemen, lethargie,…) toont enkel aan dat we de ontregeling nog niet helemaal ontcijferd hebben.

Een goed voorbeeld van een context waarin schizofrenie gepresenteerd wordt als een natuurlijke soort, is het hersenonderzoek naar schizofrenie. Sommige onderzoekers beweren dat het mogelijk is (of wordt) om louter en alleen op basis van een scan van iemands hersenen, te voorspellen of die persoon al dan niet de stoornis ‘schizofrenie’ zal ontwikkelen.

Drie bedenkingen

Eerste bedenking: onverschillig

Is het zo dat schizofrenie onverschillig is tegenover een steeds veranderende psychiatrische diagnostiek? Of, anders geformuleerd: verandert er iets aan de schizofrene patiëntenpopulatie wanneer het concept ‘schizofrenie’ verandert? Hackings antwoord op deze vraag is dat de soort ‘schizofrenie’ onmiskenbaar een interactieve soort is. Het is immers duidelijk dat er een merkwaardige coïncidentie bestaat tussen conceptuele veranderingen in de psychiatrie aan de ene kant, en de epidemiologie van schizofrenie aan de andere kant. Onderzoek toont aan dat er sterke schommelingen zijn in de prevalentie van schizofrenie, zowel cross-cultureel als doorheen de tijd. Maar ook op microniveau doen zich zulke schommelingen voor. Zo toonde de Engelse psychologe Mary Boyle aan dat de prevalentie van dementia praecox in Kraepelins psychiatrische instelling in minder dan tien jaar tijd van 5 procent (in 1892) naar 50 procent (in 1900) steeg. Het is natuurlijk mogelijk dat deze stijging een weerspiegeling is van een stijgend aantal gevallen van schizofrenie, zoals voorstanders van een *natural kind view* zouden volhouden. Maar in dit geval, zegt Boyle, lag de oorzaak in de eerste plaats bij een graduele verbreding van Kraepelins diagnostische criteria (Boyle 1990). Als schizofrenie een interactieve soort is, zo vervolgt Hacking, dan betekent dat niet alleen dat het concept ‘schizofrenie’ de psychiatrische patiënten-populatie beïnvloedt, maar ook dat, omgekeerd, patiënten met schizofrenie een invloed uitoefenen op de inhoud van het concept. Volgens Hacking heeft zo’n dubbele interactie zich inderdaad reeds voorgedaan in de geschiedenis van schizofrenie. Om die claim te begrijpen moeten we even terugkeren naar het begin van de twintigste eeuw. De Zwitserse psychiater Bleuler, een tijdgenoot van Freud, was van mening dat hallucinaties en waanbeelden geen prominente plaats zouden mogen innemen op de lijst van diagnostische criteria voor schizofrenie. Zijns inziens moesten psychiaters zich vooral concentreren op de zogenaamde ‘vier A’s’: verbrokkelende *associaties*, *ambivalentie*, onaangepaste *affecten*, en *autisme* (Bleuler 1911). Op die manier werden hallucinaties gedestigmatiseerd, zegt Hacking, en veroverden ze vreemd genoeg opnieuw een ereplaats in de psychiatrische diagnostiek: “Bleuler allowed fairly free expression of auditory hallucinations. They were not important; there were other aspects of one’s life to come to grips with. He took hallucinations in stride and paid little heed to them. Hallucinations became ordinary, not to be worried about, neither to be the voice of God to be proud of, nor something to hide from the doctor. Hallucinations became so freely available, unproblematic, that schizophrenics said they had them. So Schneider made them almost a sine qua non of schizophrenia (...). The schizophrenic, as a kind of person, is a moving target, and the classification is an interactive kind.” (Hacking 1999, 114) Er is dus sprake van een complexe interactie tussen het concept ‘schizofrenie’ en de schizofrene patiëntenpopulatie. Daarom kan men met recht zeggen dat schizofrenie een interactieve soort is. Evolutiepsychiaters lijken zich niet te bekommeren om dergelijke subtiliteiten, en neigen er in het algemeen toe om

schizofrenie eerder als een onverschillige soort af te schilderen, die onafhankelijk is van wijzigingen in de psychiatrische diagnostiek. “There is a constant drive in the social and psychological sciences to emulate the natural sciences, and to produce true natural kinds of people. This is evidently true for basic research on pathologies such as schizophrenia (...). There is a picture of an object to be searched out, the right kind, the kind that is true to nature, a fixed target if only we can get there. But perhaps it is a moving target...?” (Hacking 1999, 105)

De stelling is dus niet zozeer dat theorieën over schizofrenie een *impact* hebben op patiënten met schizofrenie (die stelling is triviaal); de stelling is dat deze impact van die aard is dat ze wijzigingen aanbrengt in het denken en het gedrag van patiënten, en dat die wijzigingen op hun beurt een impact hebben op de theorievorming over schizofrenie. Er ontstaat met andere woorden een interactie tussen theorievorming en gedrag, die we als volgt kunnen schematiseren.

Tweede bedenking: homogeen en begrensd

Er zijn ook nog andere argumenten om te beweren dat schizofrenie geen natuurlijke soort is. Zo is het onduidelijk of het onderscheid tussen schizofrenie en normaliteit wel zo strikt en dwingend als harde wetenschappers ons willen doen geloven. Er is in ieder geval een grote grijze zone tussen beide extremen. Ik geef een voorbeeld. Naast hallucinaties zijn ‘bizarre wanen’ ook een belangrijk diagnostisch criterium voor schizofrenie. Maar net als hallucinaties komen ook wanen behoorlijk vaak voor bij ‘gewone mensen’ – kijkt u maar even naar deze man.



Een kwart van de Amerikanen gelooft in het bestaan van geesten; een zevende beweert ooit een ufo gezien te hebben; en één op honderd Amerikanen meent door buitenaardse wezens ontvoerd te zijn... Kan men met die informatie volhouden dat enkel patiënten met schizofrenie wanen vertonen? Het lijkt erop dat schizofrenie alvast geen strak begrensde categorie is. Is het een homogene categorie? Zijn er kenmerken die alle patiënten met schizofrenie typeren? Is het mogelijk om alle patiënten met schizofrenie bij elkaar te brengen op basis van een unieke en observeerbare set van kenmerken (‘markers’), zoals bepaalde neuropsychologische of neurofysiologische afwijkingen? Bij gebrek aan dergelijke unieke kenmerken zou het toch minstens mogelijk moeten zijn om statistisch significante verschillen te ontwaren tussen drie groepen – patiënten met schizofrenie, gezonde individuen, en niet-schizofrene psychiatrische patiënten. Tot op heden is er geen overeenstemming over een of andere unieke marker voor schizofrenie, ook al zijn de claims op zo’n marker niet langer op twee handen te tellen. Bovendien wees Heinrichs (2001) op het feit dat ook statistisch significante verschillen eerder zeldzaam zijn, vooral tussen zieke en gezonde individuen. Heinrichs maakte een stand van zaken van het onderzoek naar schizofrenie in verschillende disciplines, en zijn conclusie is niet bepaald bemoedigend: “[C]lose to 40% of the biological findings are so weak and variable that they may represent minor, unimportant, or chance abnormalities with no intrinsic link to schizophrenia. Clearly, the cognitive evidence wins the neuroscience sweepstakes. Yet even the cognitive findings do not characterize all patients with the illness.” (Heinrichs 2001, 259) Kortom: nagenoeg alle zogenaamd unieke markers voor schizofrenie kan men ook terugvinden bij ofwel de gezonde bevolking ofwel andere psychiatrische patiënten. De stelling dat schizofrenie een natuurlijke soort is wordt niet alleen ondermijnd door het bestaan van zulke overlappingen. Feit is ook dat er in de praktijk schizofrene patiënten bestaan die helemaal geen symptomen gemeenschappelijk hebben. Zelfs moleculaire psychiaters kunnen niet anders dan toegeven dat de schizofrene patiëntenpopulatie bijzonder heterogeen is: ‘Indeed, experienced clinicians often remark that it is quite possible to encounter patients with no features in common’ (Fanous & Kendler 2005, 6). Het wordt dus moeilijk om vaste en interne kenmerken te vinden die alle leden van de soort ‘schizofrenie’ typeren, en enkel deze leden.

Derde bedenking: een arbitraire soort?

Zachar verwijst naar enkele psychologen die de huidige ziektecategorieën van de DSM als arbitraire soorten beschouwen. De DSM beroept zich inderdaad uitsluitend op uiterlijke kenmerken van ziektebeelden, dat wil zeggen de symptomen, om categorieën te vormen en te ordenen. *Mood disorders*, bijvoorbeeld, is een overkoepelende categorie in DSM waarin alle ziektebeelden besproken worden die te maken hebben met *moods*: depressie, bipolaire stoornis, dysthymie, enzovoort. Angststoornissen is een overkoepelende categorie voor categorieën als fobie,… Is zo’n ordening arbitrair? Of anders gezegd: zijn de categorieën waarmee zo’n ordening werkt arbitrair? Volgens Zachar is dat niet het geval, maar zijn concept van ‘arbitraire soort’ is, zoals ik daarnet al aanstipte, eerder eng. Arbitraire soorten zijn immers verzamelingen van entiteiten, organismen of individuen die geen enkel gemeenschappelijk kenmerk hebben. Er zijn wellicht weinig verzamelingen die aan die voorwaarde voldoen, en het lijkt erg waarschijnlijk dat arbitraire en praktische soorten verbonden zijn met elkaar in een continuum – dat er een vloeiende lijn loopt tussen beide categorieën. Het is met andere woorden erg waarschijnlijk dat de huidige editie van DSM een aantal categorieën bevat die kunnen getypeerd worden als arbitraire soorten. Zachar houdt nochtans vol dat dat niet het geval is: ‘Traditional psychiatric categories are not arbitrary’ (p. 176). Ik bespreek een tegenvoorbeeld: evolutiepsychiaters Michael McGuire & Alfonso Troisi steken de draak met de diagnostische criteria voor ‘Borderline Personality Disorder’. DSM-IV bevat een lijst van 9 diagnostische criteria, en geeft aan dat een patiënt ‘recht heeft’ op de diagnose in kwestie wanneer hij of zij aan 5 criteria uit die lijst voldoet. Deze werkwijze ontlokt McGuire & Troise de volgende opmerking: ‘Given that only five of the nine disorder-classifying criteria are required, it follows that two persons can be classified with borderline personality disorder yet share only one phenotypic attribute. This is far from precise classification’ (McGuire & Troisi 1998, 24).

Er zijn zelfs onderzoekers die stellen dat eigenlijk elk geval van schizofrenie verschillend is, zowel wat symptomatologie betreft, als wat aanvang, verloop, prognose, en reactie op psychofarmaca betreft. Die heterogeniteit zou het concept ‘schizofrenie’ volgens hen overbodig maken (zie bijvoorbeeld Bentall et al. 1988; Sarbin 1990).

Lees ook zeker de tekst van Zachar (dat geldt natuurlijk voor alle teksten)

1. Vrije wil

Ik begin met een kleine anekdote (die ik ontleen aan Maureen Sie, in het boek *Hoezo Vrije Wil* uit 2011, p. 18ev): in zijn memoires schrijft Albert Einstein dat als de maan bewustzijn zou hebben, zij zou denken dat zij haar gang om de aarde uit vrije wil maakt, en dat een intelligenter wezen dan wijzelf zou glimlachen om de gedachte dat wij ‘uit vrije wil handelen’. Mensen ervaren vanuit het eerstepersoonsperspectief de noodzaak om hun leven vorm te geven, keuzes te maken en beslissingen te nemen, maar ons leven voltrekt zich zoals de maan om de aarde draait, namelijk volgens een ijzeren wetmatigheid. De hedendaagse neurowetenschappen treden Einstein bij. Er gaat geen week voorbij waarin deze of gene wetenschapper niet uitpakt met een verpletterend bewijs dat wilsvrijheid niet bestaat. En toch kunnen we het niet laten om onszelf en anderen verantwoordelijk te houden voor wat ze doen (of niet doen). Er lijkt met andere woorden een spanning te bestaan tussen twee perspectieven: het perspectief van een of ander goddelijk wezen dat gniffelt bij de gedachte dat we onszelf met enige trots als ‘vrije wezens’ beschouwen, en ons eigen perspectief – het perspectief van een individu dat ingekapseld zit in allerlei sociale praktijken.

* 1. **Het probleem**

Het probleem dat centraal staat in het debat over vrije wil, is een probleem waarvan de structuur u ondertussen welbekend is. Het is namelijk een probleem dat ontspringt uit een conflict tussen enerzijds onze intuïties en anderzijds de inzichten van de (harde) wetenschappen over hoe de wereld in elkaar zit, en meer bepaald over de determinanten van ons eigen en andermans gedrag. Het probleem van de vrije wil is met andere woorden een probleem dat voortvloeit uit een clash tussen *folk psychology* en fysica. En die clash is, zoals u zich wellicht herinnert, ook de drijfveer van het probleem van lichaam en geest.

Het probleem van vrije wil is bij uitstek een probleem voor het fysicalistische gedachtegoed, en vult dus het rijtje aan van andere problemen die we ontmoetten: het probleem van intentionaliteit (hoe kunnen mentale toestanden identiek zijn aan hersentoestanden als hersentoestanden nergens over gaan – geen intentionaliteit hebben?), en het probleem van qualia (hoe kunnen we de kwalitatieve aspecten van mentale toestanden vertalen in een fysicalistische woordenschat?).

Folk psychology

Ons dagelijks leven is gebaseerd op de gedachte dat we meestal, of toch soms, vrij zijn om te beslissen wat we doen (actie), waarom we het doen (motivatie), en hoe we het doen (hoeveel ik er voor over heb). Wilsvrijheid houdt in dat we op zulke momenten een vrije keuze maken tussen een aantal alternatieven. Alma of Thai? Te voet of te fiets? Eerste of tweede zit? Het aspect van vrijheid bij zulke keuzes is gebaseerd op twee elementen.

Enerzijds is er een element van keuzevrijheid – de gedachte dat we ook de andere optie hadden kunnen kiezen. Vrijheid moet je in deze context met andere woorden begrijpen op een positieve manier, namelijk als keuzevrijheid. Het element van keuze bepaalt ook de morele waarde van een handeling. Wanneer we andere mensen verantwoordelijk houden voor hun daden, dan gaan we er impliciet vanuit dat ze ook iets anders hadden kunnen doen. Dat doen we overigens zowel wanneer iemand het goede doet, als wanneer iemand kwaad berokkent. Lof komt iemand enkel toe wanneer hij de vrijheid had om ook iets minder lovenswaardig te doen. En omgekeerd hebben we niet de neiging om iemand op te sluiten wanneer hij of zij geen andere keuze had, wanneer hij of zij ‘niets anders had kunnen doen’. Iemand morele verantwoordelijkheid toeschrijven houdt dus in dat we hem of haar ook keuzevrijheid toeschrijven. Verplaetse verwijst naar een roman van Imre Kertesz, *Onbepaald door het Lot*, om het fenomeen van keuzevrijheid te illustreren. De hoofdpersoon van het boek keert na de oorlog uit een concentratiekamp terug naar huis, en ergert zich daar aan het fatalisme van zijn buren. ‘Achteraf, als we terugblikken, lijkt alles zo afgesloten en voltooid, zo onveranderlijk en definitief, zo razendsnel en onverklaarbaar, zo vanzelf “gekomen”. Een verliefde kus lijkt een even noodzakelijk en logisch gebeuren als het bevel om mensen te vergassen’. Maar, vervolgt de hoofdpersoon, die terugblik is bedrieglijk. De geschiedenis kan immers op elk moment een andere wending nemen; elk moment biedt de ruimte om iets anders te doen.

Anderzijds is er een element van oorspronkelijkheid – de gedachte dat we zelf de oorsprong zijn van onze handelingen en gedachten. Dat we vrij zijn van oorzaken die ons beïnvloeden bij bepaalde beslissingen. Vrijheid wordt dan eerder op een negatieve manier begrepen, als ‘vrij zijn van’ (versus ‘vrij zijn om’). Men spreekt in deze context vaak van ‘broncontrole’ als een wezenskenmerk van wilsvrijheid. ‘Broncontrole’ betekent dat de genomen beslissing, gedachte of handeling *jouw* beslissing, gedachte of handeling is, en in die zin oorspronkelijk is. Ook dit kenmerk van wilsvrijheid speelt een belangrijke rol in het toekennen van morele verantwoordelijkheid. We houden iemand enkel verantwoordelijk voor iets dat hij of zij zelf bedacht of beraamd heeft, en niet voor iets dat hem of haar opgedrongen is.

Onze intuïtie zegt ons dat er beslissingen, handelingen en gedachten zijn die vrij zijn in beide betekenissen van het woord: dat we ook iets anders hadden kunnen denken of doen (keuzevrijheid), en dat we er zelf de bron van zijn (broncontrole). En onze intuïtie zegt ons ook dat deze vrijheid ons toestaat om onszelf en anderen te loven (wanneer ze iets verwezenlijkt hebben) en ter verantwoording te roepen (wanneer ze iets verprutst hebben).

Wat is determinisme?

Het probleem is dat onze intuïties over wilsvrijheid en verantwoordelijkheid, die volgens vele filosofen erg dwingend zijn, niet of nauwelijks te rijmen vallen met inzichten uit de Newtoniaanse fysica. De wetenschappen spreken onze intuïties tegen omdat ze *deterministisch* zijn. (Men spreekt soms ook over een ‘deterministische kosmos’ – Verplaetse noemt zo’n kosmos ‘een spoorwegnet zonder wissels’.) Wat verstaat men onder de term ‘determinisme’? Er zijn in de filosofie heel wat definities van ‘determinisme’ in omloop. Ik geef twee definities, en breng onderweg enkele nuanceringen aan.

Een eerste definitie van de term ‘determinisme’ kijkt vooral naar het verleden, en stelt dat ‘alle gebeurtenissen en standen van zaken gedetermineerd worden door voorafgaande gebeurtenissen en standen van zaken’. De term ‘determineren’ wordt vervolgens gedefinieerd als ‘veroorzaken’. Er zijn echter verschillende types van oorzaken. Zo maakt men bijvoorbeeld een onderscheid tussen probabilistische oorzaken en deterministische oorzaken. Genen zijn goede voorbeelden van probabilistische oorzaken. ‘If a geneticist speaks loosely of a “gene for behaviour X”, he does not mean that there is a gene which would cause behaviour X if inserted into any animal. He means that there is a gene which, if it is present in an animal of a given species, along with all the other genes and environmental circumstances characteristic of that species, will make that animal somewhat more likely to do X’ (Maynard Smith 1986, pp. 57). Genen zijn oorzaken die het waarschijnlijker maken dat een organisme een bepaald kenmerk zal vertonen (of zich op een bepaalde manier zal gedragen). Genen zijn geen deterministische oorzaken. Deze laatste oorzaken *garanderen* immers een bepaald resultaat. Gegeven deze oorzaak is het onvermijdelijk dat het resultaat volgt. Deterministen hebben het uiteraard over deze laatste soort van oorzaken. De eerste definitie van ‘determinisme’ luidt dus dat alle gebeurtenissen en standen van zaken voorafgaande deterministische oorzaken hebben. De Newtoniaanse fysica is een goed voorbeeld van een deterministische theorie, maar niet alle fysische theorieën zijn deterministisch. De quantummechanica is een voorbeeld van een fysische theorie die stelt dat de meest fundamentele bouwstenen van het universum geen deterministische oorzaken hebben. (Er zijn ook fysici die beweren dat zelfs de Newtoniaanse fysica ruimte laat voor een zekere mate van indeterminisme, maar die laat ik gemakshalve buiten beschouwing.) Als het waar is dat alle gebeurtenissen en standen van zaken in het universum deterministische oorzaken hebben, hoe kan er dan sprake zijn van keuzevrijheid en broncontrole? Determinisme sluit keuzevrijheid uit (er kan geen sprake zijn van alternatieven in een deterministisch universum), maar ook broncontrole. De gedachte dat alles in het universum door iets anders veroorzaakt wordt (denk aan het bekende dictum uit de Oudheid: ‘omne quod movetur, ab alio movetur’, of ‘alles wat beweegt, wordt door iets anders bewogen’) sluit de mogelijkheid uit dat ik het absolute begin (of de bron) ben van een beslissing, een gedachte of een handeling. Want ook de eigenschappen van het ‘ik’ – de vermeende bron van – worden door iets anders veroorzaakt...

Een tweede definitie van ‘determinisme’ kijkt niet naar het verleden, maar wel naar de toekomst. Deze definitie luidt dat, indien we alles zouden weten over het universum op tijdstip 1, en indien we alle details zouden weten over de natuurwetten in het universum, we dan in staat zouden zijn om te voorspellen hoe het universum zal zijn op een (of elk) volgend tijdstip. Sommige filosofen definiëren de term ‘determinisme’ dan ook als volgt: ‘de stelling dat er voor elke toestand van de wereld slechts een enkele toekomstige toestand bestaat’. Neem bijvoorbeeld Jaworski: ‘Determinism is the thesis that for any given state of the universe at a time there is exactly one possible resultant state that comes about because of the laws of nature’. In deze context spreekt men vaak over de ‘demon van Laplace’.

De demon van Laplace

De demon van Pierre-Simon Laplace wordt ook wel eens ‘de superman van Laplace’ genoemd, afhankelijk van de theoretische overtuigingen van diegene die het beeld gebruikt. Laplace voert zijn superman ten tonele om het determinisme van de wetenschappen uit te leggen. Superman weet iets wat geen enkele gewone sterveling weet, en kan iets wat geen enkele sterveling kan. Superman kent, op eender welk moment, de positie van *alle* deeltjes van de werkelijkheid, en de wetmatigheden (natuurwetten) waaraan ze onderworpen zijn. Als Superman een voldoende krachtig brein heeft, of gewoon een supercomputer, dan kan hij dus, krachtens zijn kennis, enerzijds precies beschrijven hoe de wereld er op een vroeger tijdstip uitzag (pakweg op de dag van mijn geboorte, of op zondag 13 mei van het jaar 17823 BC, of op de dag van de oerknal), en anderzijds precies voorspellen hoe de wereld er zal uitzien op eender welk toekomstig tijdstip. Laplace formuleert het zo: ‘Als er een verstand zou bestaan dat in staat zou zijn alle krachten die de natuur beheersen te omvatten en tevens de toestand waarin elk wezen uit de natuur zich op één gegeven ogenblik bevindt, als dat verstand verder voldoende groot zou zijn om al deze gegevens aan analyse te onderwerpen, dan zou hij één enkele formule kunnen opstellen die tegelijk de bewegingen van de grootste lichamen van het universum en de kleinste atomen zou omvatten. Voor dit verstand zou niets onzeker zijn, en zowel het verleden als de toekomst zouden voor zijn ogen open liggen’.

Je kan de gedachte ook op een eenvoudiger manier formuleren. Ik denk dat we het eens zijn over het feit dat quarks geen keuzevrijheid hebben (laat staan broncontrole). Quarks zijn de speelbal van andere quarks en andere deeltjes in de werkelijkheid, en de natuurwetten waaraan alle deeltjes in de werkelijkheid gehoorzamen. Natuurwetten zijn onbuigzaam en meedogenloos – ze laten geen uitzonderingen toe. Dus als we de positie van enkele quarks op tijdstip 1 kennen, en we kennen de natuurwetten, dan kunnen we met zekerheid voorspellen waar elk van die quarks zich op tijdstip 2 zal bevinden. U volgt nog? Dan construeer ik een tweede premisse, die wellicht even onbetwijfelbaar is als de voorgaande. De premisse luidt: wij zijn allemaal opgebouwd uit quarks. Het fysicalisme beweert dat alles in de werkelijkheid beschreven en verklaard kan worden met behulp van de fysica. Dat wil zeggen: met behulp van de entiteiten die de fysica postuleert, zoals quarks. De ontologie van de fysica bevat uiteraard ook nog andere entiteiten, zoals protonen. Maar voor al die entiteiten geldt, net als voor quarks, dat ze geen keuzevrijheid (noch broncontrole) hebben. Het fysicalisme beweert dus dat de mens volledig opgetrokken is uit entiteiten zonder wilsvrijheid.

Ergo: de mens heeft geen vrije wil. In een wetenschappelijke wereld – een wereld die beschreven en verklaard wordt met behulp van de (fysische) wetenschappen – is er geen plaats voor keuzevrijheid of broncontrole, en dus ook niet voor vrije wil, en dus ook niet voor morele verantwoordelijkheid. Net als zijn elementaire bouwstenen is de mens een speelbal van andere bouwstenen en natuurwetten. De superman van Laplace weet al van bij onze geboorte wat er van ons zal worden, welke kleren we zullen dragen op onze plechtige communie, met wie we zullen trouwen, de namen van onze kinderen, en uiteraard onze sterfdatum. Hoe kan onze wil vrij zijn als het waar is, zoals het determinisme stelt, dat alle gebeurtenissen en standen van zaken (dus ook mijn handelingen en gedachten) in het universum voorafgaande deterministische oorzaken hebben? Er kan toch geen sprake zijn van keuzevrijheid in een gesloten causaal systeem zoals het universum?

* 1. **Enkele oplossingen**

De gedachte dat determinisme en wilsvrijheid elkaar uitsluiten is echter niet voor alle filosofen vanzelfsprekend, zoals we dadelijk zullen zien. In wat volgt bespreek ik drie verschillende oplossingen voor het probleem van determinisme en wilsvrijheid (en het probleem is dat beide fenomenen elkaar lijken uit te sluiten), en enkel de eerste twee oplossingen zijn gebaseerd op de gedachte van incompatibiliteit. De derde positie, het compatibilisme, stelt dat determinisme en wilsvrijheid wel degelijk verzoenbaar zijn.

Hard determinisme

Een eerste positie in het debat over wilsvrijheid en determinisme is de positie van Jan Verplaetse in *Zonder Vrije Wil*. Je kan de positie omschrijven met behulp van een eenvoudig syllogisme met twee premissen. A) Ofwel is er vrije wil, ofwel determinisme. B) Determinisme is waar. De conclusie is onvermijdelijk als je het eens bent met de twee premissen: er is geen vrije wil. Deze positie staat bekend als ‘hard determinisme’ of ‘hard incompatibilisme’ (omwille van premisse A: determinisme en wilsvrijheid zijn incompatibel, dat wil zeggen: ze sluiten elkaar uit). ‘Hard’ staat hier tegenover ‘zacht’, en ‘zachte incompatibilisten’ zijn filosofen die geloven dat niet alles gedetermineerd is, en dat daarom wilsvrijheid en verantwoordelijkheid toch nog mogelijk zijn.

Er zijn overigens belangrijke verschillen in het enthousiasme waarmee harde incompatibilisten hun conclusie (‘wilsvrijheid bestaat niet, evenmin als morele verantwoordelijkheid’) begroeten. De meesten voelen zich er eerder ongemakkelijk bij – *an inconvenient truth*, zeg maar. Sommigen proberen de bittere pil te verzachten door te beweren dat we de illusie van wilsvrijheid en morele verantwoordelijkheid nodig hebben, ook al bestaan ze niet Zoals Daniel Wegner doet.

Verplaetse verdedigt naar eigen zeggen ‘een radicaal standpunt dat waarschijnlijk botst met je diepste intuïties’ (p. 8). Zijn standpunt in het debat over wilsvrijheid komt min of meer overeen met het standpunt van eliminativistische fysicalisten in het debat over lichaam en geest: wilsvrijheid bestaat niet. Het concept van wilsvrijheid is niet vergelijkbaar met het concept van water – het is geen concept dat zijn nut kan behouden in het dagelijks leven, ondanks het feit dat de wetenschappen een andere (betere?) beschrijving (‘H²O’) geven van hun beider referent (de vloeistof die uit de kraan komt). ‘Water’ en ‘H²O’ zijn verschillende beschrijvingen van eenzelfde fenomeen, en beide beschrijvingen hebben hun bestaansrecht. Harde incompatibilisten zouden ‘wilsvrijheid’ eerder vergelijken met ‘flogiston’ – een concept dat volstrekt geen bestaansrecht heeft, omdat het naar niets verwijst.

Verplaetse beseft natuurlijk dat een wereld zonder wilsvrijheid en morele verantwoordelijkheid er anders uitziet dan onze wereld. Het is duidelijk dat we een en ander verliezen wanneer we wilsvrijheid en verantwoordelijkheid overboord gooien. Kunnen we vrede nemen met een wereld waarin Ratko Mladic principieel niet verantwoordelijk gehouden kan worden voor het bloedblad in Srebrenica, waarbij duizenden Moslim-mannen (volgens de aanklacht) vermoord werden door de troepen van Mladic? Het gaat dan voor alle duidelijkheid niet over de vraag of Mladic feitelijk verantwoordelijk is voor de slachting (een vraag waarover het proces uitsluitsel moet geven), maar wel over de vraag of het überhaupt zinvol is om eender wie verantwoordelijk te houden voor zijn daden (vandaar: ‘principieel verantwoordelijk’). Maar volgens Verplaetse hoeft een wereld zonder wilsvrijheid en verantwoordelijkheid er niet fundamenteel anders uit te zien. Als u hier meer over wil weten, dan kan u de laatste twee hoofdstukken van het boek van Verplaetse lezen.

Libertarisme

Hun redenering: Ofwel is er vrije wil, ofwel determinisme. Determinisme is onwaar dus er is vrije wil.

In het eerste hoofdstuk over dualisme heb ik aangekondigd om nog iets te zeggen over emergentisme . Net als epifenomenalisme (de theorie die volgens Frank Jackson volgt uit zijn verhaal over Mary) is emergentisme een variant van het zogenaamde eigenschapsdualisme. Volgens eigenschapsdualisten kunnen mentale toestanden niet beschreven worden met behulp van de fysica (contra fysicalisten); nochtans worden ze wel veroorzaakt door fysische toestanden (contra substantiedualisten). Het emergentisme is een inspiratiebron voor sommige libertaristen in het debat over wilsvrijheid en determinisme. Zulke libertaristen menen dat wilsvrijheid een emergente eigenschap is – een eigenschap die oprijst uit de menselijke natuur, uit allerlei complexe interacties in de menselijke hersenen.

Denk even terug aan het verhaal over quarks (supra). Het is duidelijk dat quarks geen keuzevrijheid hebben; wij zijn volledig opgebouwd uit quarks, dus waarom zouden wij keuzevrijheid, en bijgevolg wilsvrijheid hebben? Het is deze redenering die betwist wordt door emergentisten en libertaristen. Emergentie is namelijk een fenomeen waarbij een grote geheel eigenschappen krijgt die je onmogelijk aan de delen kunt toeschrijven. Ik citeer Verplaetse: ‘Denk aan “intelligentie”. Je hebt geen intelligente zenuwcel of neuron. Je hebt geen intelligente hersenkwab. Je hebt zelfs geen intelligent brein. Maar je hebt wel intelligente mensen. “Intelligentie” is een macro-eigenschap die pas ontstaat wanneer je complex menselijk gedrag hebt. Je schrijft die eigenschap niet toe aan allerlei delen die wel essentieel zijn voor intelligentie. De eigenschap verdwijnt pas vanaf een zeker niveau. Deze verrassende emergentie wordt samengevat in het motto “het geheel is meer dan de delen”’ (p. 100). Even recapituleren. Libertaristen proberen wilsvrijheid en morele verantwoordelijkheid te redden door te beweren dat het determinisme onwaar is. Niet alle gebeurtenissen en standen van zaken, zo zeggen ze, worden bepaald door deterministische oorzaken. Sommige beslissingen onttrekken zich aan de keten van oorzaken – zulke beslissingen zijn vrije beslissingen. (Men spreekt daarom over ‘acausale libertaristen’.) Maar hoe kunnen beslissingen (of gedachten of handelingen) zich onttrekken aan causale netwerken? Om die vraag te beantwoorden verwijzen libertaristen dus naar het bestaan van emergente eigenschappen. Maar, hoor ik u opwerpen, emergente eigenschappen zijn toch superveniërende eigenschappen? Dat wil zeggen: zelfs emergente mentale eigenschappen zijn toch onlosmakelijk verbonden met fysische eigenschappen? Superveniëntie-theorie leert ons dat macro-eigenschappen (mentale toestanden) asymmetrisch afhankelijk zijn van micro-eigenschappen (hersentoestanden). Dat wil zeggen: mentale toestanden kunnen enkel wijzigen wanneer hersentoestanden wijzigen (verschillen in intelligentie gaan altijd gepaard met neurale verschillen), maar omgekeerd kan eenzelfde mentale toestand wel voortkomen uit (gerealiseerd worden door) verschillende hersentoestanden (dezelfde intelligentie kan voortkomen uit verschillende neurale structuren). Vertalen we superveniëntie naar het debat over wilsvrijheid, dan zitten we met een probleem. Want het superveniëntie-denken leert ons dat eenzelfde hersentoestand onmogelijk kan leiden tot verschillende mentale toestanden. Als we de hersenen beschouwen als een causaal netwerk, en een wilsbeslissing als een mentale toestand, dan gunt het superveniëntie-denken ons geen wilsvrijheid. Wat we nodig hebben is een theorie die ons toestaat dat twee identieke breinen in dezelfde neurale toestand verschillende gedachten hebben. Die mogelijkheid wordt uitgesloten door het superveniëntie-denken, dus moeten we kiezen voor een niet-superveniënt emergentisme…

Libertaristen beweren dat een individu vanuit een identieke hersentoestand tot verschillende intenties kan komen. Het individu wordt met andere woorden niet gedetermineerd door het causale netwerk van zijn hersencellen. Hij of zij kan in alle vrijheid tot een beslissing, een gedachte of een handeling komen. Het probleem is dat libertaristen daarbij beroep doen op een sterke vorm van emergentisme, die veel dualistischer is dan bijvoorbeeld het epifenomenalisme. Eigenlijk verschillen ze dan ook niet zoveel van de substantiedualisten, die het probleem van wilsvrijheid en determinisme uiteraard niet kennen. De geest staat volgens hen los van het fysisch gebeuren, en kan dus in alle vrijheid beslissingen nemen. Met andere woorden: alle argumenten tegen sterke vormen van dualisme (zowel eigenschapdualisme als substantie-dualisme) kan je ook inbrengen tegen deze libertaristen. De libertaristische oplossing voor het probleem van wilsvrijheid en determinisme is volgens velen een magische oplossing, en kan daarom niet ernstig genomen worden.

Compatibilisme

In tegenstelling tot deterministen en libertaristen doen compatibilisten geen uitspraak over de waarheid van het determinisme; volgens hen kunnen we iemand moreel verantwoordelijk houden voor zijn of haar daden, ongeacht of het determinisme waar is of niet. ‘Compatibilisme is de meest sympathieke positie’, zegt Verplaetse. ‘Zij verzoent wetenschappelijk denken en verantwoordelijkheid. Er is geen vuiltje aan de lucht. Al werken kosmos en brein causaal deterministisch, zelfs dan blijft verantwoordelijkheid bestaan. Hoe de kosmos functioneert, maakt voor verantwoordelijkheid niets uit’ (*Zonder Vrije Wil*, p. 42). Maar hoe verzoenen we onze intuïtie (dat onze wil vrij is) met de waarheid van de wetenschappen (dat alles, en dus ook onze wil, gedetermineerd is)?

Er zijn twee opties. De eerste optie, waarover ik verder niets zal zeggen, is te ontkennen dat morele verantwoordelijkheid keuzevrijheid en broncontrole vereist. Ik heb daarnet uitgelegd dat morele verantwoordelijkheid wilsvrijheid veronderstelt (iemand die gedwongen wordt om een misdaad te plegen, kan niet verantwoordelijkheid gehouden worden voor die misdaad), en dat wilsvrijheid op haar beurt keuzevrijheid en broncontrole vereist. Sommige compatibilisten, zoals de Amerikaanse filosoof Harry Frankfurt, proberen deze laatste claim te ondermijnen – volgens hen is keuzevrijheid geen noodzakelijke voorwaarde voor morele verantwoordelijkheid. De tweede optie is een Wittgensteiniaanse optie, die verdedigd wordt door, onder andere, Peter Strawson.

De filosoof Peter Strawson is een goed voorbeeld van een compatibilist. Strawson verdedigt de stelling dat de zin van de toekenning van verantwoordelijkheid helemaal niet kan afhangen van de vraag of het determinisme al dan niet waar is. Volgens Strawson weerspiegelen vrije wil en determinisme twee menselijke attitudes, respectievelijk de reactieve en de objectiverende attitude. ‘Wanneer we het gedrag van iemand wetenschappelijk proberen te verklaren en daarbij bijvoorbeeld een beroep doen op hersenonderzoek, nemen we tegenover dit gedrag een objectiverende attitude aan (Burms 2011, 130). Denk even terug aan Olga uit *Turks Fruit*. Zodra we begrijpen dat Olga’s vreemde gedrag toe te schrijven is aan een hersentumor, verdwijnt onze neiging om haar gedrag moreel te veroordelen, en haar tot verantwoordelijkheid op te roepen. Het verhaal over de hersentumor activeert een objectiverende attitude. We geven onze afkeuring of verontwaardiging op. Voordien teerden we eerder op een reactieve attitude – een houding die eerder gebaseerd is op persoonlijke betrokkenheid. ‘Onze houding is “reactief” wanneer we ons bijvoorbeeld beledigd voelen door wat iemand anders zegt’ (Ibid.). Het verhaal van Olga toont aan dat we in bepaalde gevallen onze reactieve attitude kunnen vervangen door een objectiverende attitude, maar volgens Strawson is het absurd om te beweren dat we *al* onze reactieve attitudes moeten vervangen door objectiverende attitudes, zoals harde deterministen beweren. ‘Dat een bepaald menselijk gegeven (gedrag, emotie, verlangen) wetenschappelijk verklaarbaar is, impliceert helemaal niet dat we het uitsluitend met een objectiverende, wetenschappelijke attitude zouden moeten benaderen en dat het dus onredelijk zou zijn er een reactieve attitude tegenover aan te nemen. Het is niet omdat een objectiverend standpunt mogelijk is, dat het ook vereist zou zijn’ (Ibid., 131-2). Volgens Strawson sluiten beide soorten van attitudes elkaar niet uit. Je kan eenzelfde vrouwenlichaam op een reactieve manier benaderen, waarbij je geraakt wordt door de schoonheid en sensualiteit, én op een objectiverende manier, waarbij je vooral oog hebt voor verhoudingen (zoals de befaamde waist-hip ratio), en de vorm van bepaalde onderdelen (grootte van de neus, stand van de voeten,…). En in de mate dat de verhouding tussen beide attitudes een weerspiegeling vormt van de verhouding tussen determinisme en vrije wil, kan je concluderen dat deze laatsten elkaar ook niet uitsluiten.

Of met andere woorden: vrije wil en determinisme sluiten elkaar niet uit. Het zijn twee ‘sferen’, zou je kunnen zeggen, die min of meer onafhankelijk van elkaar bewegen (ik leg de nadruk op ‘min of meer’ omdat het voorbeeld van Olga aantoont dat er geen totale onafhankelijkheid bestaat; soms worden we gedwonen om reactieve attitudes te vervangen door objectiverende attitudes), omwille van het feit dat ze elk hun eigen logica hebben. Aan de ene kant is er de sfeer van het sociale leven, en aan de andere kant de sfeer van de wetenschappen. Het toeschrijven van morele verantwoordelijkheid gebeurt in de eerste sfeer, in praktijken waar we aan deelnemen omdat ze voldoen aan onze meest fundamentele menselijke behoeften, zoals erkenning en waardering. En deze praktijken zijn niet afhankelijk van de waarheid of onwaarheid van het determinisme (de tweede sfeer). We kunnen mensen dus moreel verantwoordelijk houden voor hun gedrag, ongedacht of het determinisme waar is of niet.

Interessant in deze context is een uitspraak van Verplaetse, in *Zonder Vrije Wil*: ‘Zelfs wanneer ik denk dat ik het bij het rechte eind heb, valt het tegen alledaagse intuïties inroeien zwaar. Aanvankelijk lijkt de radicale conclusie aantrekkelijk. Diep in de nacht tussen je boeken krijg je een heel nieuwe kijk op mens en wereld. Wereld en mens te aanschouwen vanuit het licht van de eeuwigheid bezorgt je bijna een mystieke ervaring’ (p. 22). *Vanuit het licht van de eeuwigheid*? Strawson (en zeker Burms) zouden zich ergeren aan deze uitspraak – de mens is onherroepelijk verbonden met zijn eigen perspectief. *A God’s eye view* zit er niet meteen in. Of wel, maar dan betalen we een hoge prijs. Dan plaatsen we onszelf buiten de grenzen van die sociale praktijken waaraan we ons laven, en waarin we onze meest fundamentele menselijke behoeften kunnen bevredigen.

c) Sommige filosofen beweren dat mensen van nature compatibilisten zijn. We zijn van nature geneigd om wilsvrijheid los te koppelen van determinsme – ook al weten we dat alles strikt genomen vastligt, toch blijven we onszelf en anderen ter verantwoording roepen. Er zijn zelfs filosofen die empirisch willen bewijzen dat mensen van nature compatibilistisch denken over wilsvrijheid en determinisme. In het voorbije decennium is er een nieuwe stroming ontstaan in de filosofie, de experimentele filosofie, waarbij filosofen gebruik maken van empirische methoden om bepaalde intuïties te verifiëren. Men wil niet langer genoegen nemen met stelligheden als ‘iedereen weet toch dat…’ of ‘het is vanzelfsprekend dat men in deze context…’ Intuïties moeten getoetst worden – denkt iedereen (ook de slager, de poetsvrouw en de notaris) zoals de beroepsfilosoof denkt? Een van deze experimentele filosofen, de Amerikaan Eddy Nahmias, hoopt te kunnen aantonen dat mensen van nature compatibilisten zijn; anderen, zoals Shaun Nichols en Joshua Knobe, vermoeden dat een instabiele vorm van hard incompatibilisme de *default mode* is. Waarom instabiel? Omdat onze intuïties soms veranderen, vooral onder invloed van ingrijpende (im)morele gebeurtenissen. Met andere woorden: mensen zijn van nature geneigd om zichzelf en anderen te ontslaan van verantwoordelijkheid, behalve wanneer ze het slachtoffer (of bevoorrechte getuige) zijn van een zware misdaad, zoals een verkrachting. De morele verontwaardiging die ze dan voelen botst dan met hun aanvankelijke intuïties, en morele emoties blijken krachtiger te zijn dan filosofische intuïties.

* 1. **Is de vrije wil een illusie?**

De opmars van de neurowetenschappen in de voorbije jaren heeft het debat over determinisme en vrije wil opnieuw aangewakkerd. Sommige neurowetenschappers menen dat hun onderzoek aantoont dat de vrije wil een illusie is. Hebben ze goede argumenten om die radicale claim te ondersteunen? Of beter: hebben ze goede data? Die vraag staat centraal in het laatste luik van dit hoofdstuk.

Iemand die naarstig onderzoek gedaan heeft naar hoe beslissingen gevormd worden is Marcel Libet. In het verdere verloop van de cursus zal duidelijk worden wat hij precies gedaan heeft, maar eerst geven we achtergrond informatie nodig om het werk van Marcel Libet te begrijpen

Om te beginnen is het duidelijk dat Libet zich concentreert op één aspect van wilsvrijheid, namelijk broncontrole, eerder dan op het andere aspect, namelijk keuzevrijheid. Dooremalen & co citeren Libets beschrijving van broncontrole als een fenomeen waarbij ‘subjecten introspectief voelen dat zij de handeling verrichten vanuit hun eigen initiatief en dat zij vrij zijn om de handeling te beginnen of niet te beginnen zoals ze wensen’ (p. 171). *Als* broncontrole een noodzakelijke voorwaarde is voor wilsvrijheid, en *als* we kunnen aantonen dat broncontrole een illusie is, dan hebben we aangetoond dat wilsvrijheid een illusie is.

Achtergrond

Is broncontrole een noodzakelijke voorwaarde voor wilsvrijheid? Ik zei daarnet dat je twee elementen kan onderscheiden in het fenomeen van wilsvrijheid: keuzevrijheid én broncontrole. Je kan het concept van wilsvrijheid met andere woorden analyseren met behulp van twee andere concepten. Deze situatie doet denken aan Wakefields analyse van het concept ‘ziekte’ (zie les 9). Wakefield onderscheidt ook twee elementen in het fenomeen van ziekte: schade en disfunctie. Hij spreekt van noodzakelijke en voldoende voorwaarden: elk van deze voorwaarden is een noodzakelijke voorwaarde om een toestand een ziekte (of stoornis) te noemen; samen zijn ze een voldoende voorwaarde. De vraag is of je keuzevrijheid en broncontrole ook kan beschouwen als noodzakelijke en voldoende voorwaarden om een beslissing (of gedachte of handeling) vrij te noemen. Als het zo is dat keuzevrijheid en broncontrole elk noodzakelijke voorwaarden zijn, en samen voldoende voorwaarden, dan ondermijnt Libets werk het concept van wilsvrijheid. Een beslissing zonder broncontrole kan in deze zienswijze geen vrije wilsbeslissing zijn, net zomin als een toestand zonder disfunctie een ziekte kan zijn (denk bijvoorbeeld aan rouw).

Nu zijn er filosofen die betwijfelen of keuzevrijheid een noodzakelijke voorwaarde is voor wilsvrijheid. Harry Frankfurt is een goed voorbeeld, zoals ik daarnet al zei. Maar bij mijn weten zijn er geen filosofen die het belang van broncontrole in vraag stellen. Anders gezegd: de meeste filosofen lijken het eens te zijn over de stelling dat broncontrole een noodzakelijke voorwaarde is voor wilsvrijheid. Is het ook een voldoende voorwaarde? Is elke beslissing die voldoet aan de voorwaarde van broncontrole een vrije beslissing? Of zijn er nog andere elementen die we in rekening moeten brengen?

Dan de tweede premisse: is broncontrole een illusie? Is het een illusie te denken dat we ooit de bron of de oorsprong zijn van een beslissing, een gedachte of een handeling? Is het een illusie te denken dat de productie van beslissingen in handen is van een bron waarover de beslissende persoon controle heeft, zoals een eigen Ik? We hebben daarnet gezien dat libertaristen proberen om zo’n oorsprong of bron te vrijwaren, en dat ze daarbij beroep doen op argumenten die ontleend zijn aan de hardere vormen van het dualisme. Soms doen ze ook beroep op ontwikkelingen in de wetenschappen, zoals de quantumfysica. Ik citeer Verplaetse: ‘De vrije wil of het bewustzijn zou een kwantumverschijnsel zijn dat ontstaat in de zogenaamde microtubuli, zo speculeert de Britse natuurkundige Roger Penrose. Microtubuli zijn kleine eiwitstructuren in onze zenuwcellen die een belangrijke rol spelen bij celdeling en intracellulair transport. Maar die structuren vind je ook bij planten, die we geen vrije wil of bewustzijn toedichten. Bovendien ontstaan kwantumeffecten pas bij een temperatuur die het absolute nulpunt van -273,25 graden Celsius benadert. In ons brein is het veel te warm’ (p. 104-5). Het lijkt er dus niet op dat de wetenschappen het libertarisme kunnen voorthelpen.

Integendeel zelfs. Sommige wetenschappen leren ons dat al onze beslissingen verankerd zijn in causale netwerken

Zoals Dooremalen & co duidelijk maken, verfijnde Libet eerder onderzoek over hersengolfpatronen die samengaan met allerlei activiteiten. Wanneer je iemand in een experimentele setting vraagt om op een vooraf bepaald tijdstip iets te doen (met de ogen knipperen, in de handen klappen,…), en daarbij zijn hersengolfpatronen meet met behulp van een electro-encefalograaf (EEG), dan merk je dat er ongeveer een seconde *voor* de handeling ‘iets in de hersenen’ op gang komt. Men noemt dit fenomeen het *Bereitschaftspotentiaal* of gereedheidspotentiaal.

Libet liet zijn proefpersonen daarentegen zelf het tijdstip kiezen waarop ze een beslissing namen. Hij onderscheidde daarbij spontane en geplande beslissingen. Spontane beslissingen borrelen op ‘uit het niets’. Andere beslissingen worden voorbereid of gepland – proefpersonen konden zich bijvoorbeeld voornemen om op een bepaald tijdstip ‘iets te doen’. Libets onderzoek wees uit dat spontane beslissingen een gereedheidspotentiaal hadden dat tussen 400 en 700 milliseconden lag (ongeveer een halve seconde dus), terwijl de gereedschapspotentiaal van geplande beslissingen nog vroeger begon, ongeveer 700 milliseconden voor de beslissing. Vervolgens vroeg Libet de proefpersonen om aan te geven wanneer ze zich voor het eerst bewust werden van hun wil om iets te doen (het zogenaamde W-moment of Wanting-moment), en het resultaat wees uit dat het gereedschapspotentiaal bij geplande beslissingen al startte op 800 milliseconden voor de eerste bewuste wilservaring.

Libets werk

Libet was geen tafelspringer. Hij besefte zelf dat zijn werk niet meteen komaf maakte met het debat over wilsvrijheid. Zijn onderzoek wijst uit dat er onbewuste breinprocessen vooraf gaan aan bewuste beslissingen. Er gebeurt iets in het brein – iets waarvan we ons niet bewust zijn – voor we ons bewustworden van een beslissing. Maar wat is dat ‘iets’? ‘Gereedheidspotentiaal’ is een behoorlijk vaag fenomeen. Je kan het nog best vertalen als een verhoogde algemene aandacht of concentratie die voorafgaat aan een bepaalde beslissing of handeling. Ik citeer Verplaetse: ‘…de kern van Libets bevinding: bewuste beslissingen komen niet uit de luch vallen. Onbewuste processen gaan altijd aan bewuste beslissingen vooraf. De gereedheidspotentiaal is een van die processen. Hij veroorzaakt de bewuste beslissing niet, maar het is een van de talloze breinprocessen die bijdragen tot de beslissing. De gereedheidspotentiaal is geen voldoende voorwaarde voor een specifieke beslissing. Het is een onderdeel van het causale netwerk dat tot de beslissing leidt. Beslissingen zonder causaal netwerk bestaan niet, dat is de kern van Libets bevinding die dertig jaar later nog altijd standhoudt’ (p. 109).

Beslissingen komen niet uit de lucht gevallen; ik ben niet de absolute oorsprong van elk van mijn beslissingen (en van geen enkele beslissing?). En in de mate dat broncontrole een noodzakelijke voorwaarde is voor wilsvrijheid, toont Libets onderzoek dus aan dat wilsvrijheid een illusie is. De gereedheidspotentiaal die aan onze beslissingen voorafgaat trekt de ‘acausaliteit’ van het libertarisme in twijfel. Met ‘acausaliteit’ verwijs ik naar de libertarische overtuiging dat vrije beslissingen niet veroorzaakt worden (dat ze ‘uit de lucht komen vallen’).

Libet liet het niet bij zijn negatieve boodschap. Misschien is er geen sprake van acausaliteit bij positieve beslissingen, zo redeneerde hij, maar wel bij negatieve beslissingen. Positieve beslissingen zijn beslissingen om iets te doen; negatieve beslissingen zijn beslissingen om iets niet te doen – veto-beslissingen. Je kan bijvoorbeeld beslissen om een snood plan, waar je al jaren op aan het broeden bent, toch niet uit te voeren. Of je kan beslissen om die beledigende commentaar op een of ander internet-forum, die je eigenlijk al geschreven hebt, toch niet te plaatsen (iedereen zou je dankbaar zijn trouwens). Bij zulke beslissingen, merkte Libet, gebeurt er iets vreemds. Ook veto-beslissingen worden voorafgegaan door een gereedheidspotentiaal, en het EEG-patroon begint ongeveer op hetzelfde moment als positieve beslissingen, maar ongeveer 200 tot 100 milliseconden voor het afgesproken tijdstip (dat wil zeggen: het tijdstip waarop de proefpersoon volgens afspraak de geplande beweging afblaast) valt het hersengolfpatroon stil. In tegenstelling tot de oorspronkelijke positieve beslissing, valt de vetobeslissing dus wel uit de lucht. Kortom: je bent misschien niet vrij om iets te beslissen, maar wel om een beslissing te verhinderen (en, zoals Dooremalen & co aangeven, om er mee in te stemmen). Of Libets conclusie gegrond is, is niet helemaal duidelijk. Maar het lijkt in ieder geval een overhaaste conclusie – er zijn tal van hersenprocessen die aan Libets electro-encefalogram ontsnappen, en die misschien wel een rol spelen in het weigeren van een beslissing…

**FIN**