

Neurofilosofie en Compatibilisme

**Daniel C. Dennett, Antonio Damasio en de
neurofilosofische verdediging van de vrije wil**

Door: Kuba Szutkowski (s772023)
Universiteit van Tilburg

Inhoudsopgave

4	1	Inleiding
6	1.1	Probleemstelling
7	1.2	Auteurkeuze
10	2	De relevantie van de neurofilosofi
11	2.1	Determinisme en compatibilisme
14	2.2	Verschillende perspectieven van verklaren
15		
16		THE PHYSICAL STANCE
16		THE DESIGN STANCE
17		THE INTENTIONAL STANCE
19		INTROSPECTIE
20		EVOLUTIONAIR PERSPECTIEF
21		PERSONAL OR MORAL STANCE
22	2.3	Neurofilosofie, een mix van verklaringshoudingen
23		TOEPASSING EN KRITIEK OP DE HETEROFENOMENOLOGIE EN DE NEUROFILOSOFIE
27	3	Neurofilosofische onderbouwing van vrije wil
27	3.1	De handeling moet echt uit jezelf komen
27		DE MEREOLOGISCHE DROGREDEN
29		LICHAAM EN BREIN
30		HET ZELFPROCES
32		HET 3-STAPSGEWIJS OPGEBOUWDE ZELFPROCES VAN DAMASIO
37		TEGEN JEZELF COMMUNICEREN, JEZELF AANSTUREN
40	3.2	De handeling moet met een reden zijn verricht
41		DE EVOLUTIE VAN COÖPERATIE
43		HANDELEN OP GEVOEL
46	3.3	De mogelijkheid tot alternativiteit
46		DE TERREUR VAN HET ONBEWUSTE
51		DE PSYCHOLOGISCHE ONTWIKKELING VAN MORALITEIT
53	4	Conclusie en discussie
53	4.1	De bruikbaarheid van
53	4.2	Verschillen en overeenkomsten tussen Dennett en Damasio
56		KRITIEK OP DAMASIO
57		KRITIEK OP DENNETT
59	4.3	De belofte van
61	4.4	Is de vrije wil van Damasio en Dennett een wenselijke vrije wil?
61	4.5	Nawoord

“Context is everything in understanding.”
Iain McGilchrist¹

¹ McGilchrist, I. (2010, November 22), RSA. *The Divided Brain and the Making of the Western World*.
Videolecture retrieved from <http://www.thersa.org/events/video/archive/iain-mcgilchrist>

1. Inleiding

De vraag die me geïnspireerd heeft tot dit onderzoek is of de neurowetenschappelijke vooruitgang de vrije wil tot een achterhaald idee heeft gemaakt. Dat lijkt namelijk de boodschap van populair wetenschappelijke publicaties. Tot voor kort hield de empirische wetenschap zich verre van het onderwerp vrije wil omdat de discussie over dit onderwerp altijd leek samen te hangen met vragen over Cartesiaans dualisme. Een substantiële ziel is nu eenmaal iets waar de empirist geen raad mee weet.² Aangezien er door filosofen en wetenschappers in de vorige eeuw massaal afstand is gedaan van het Cartesiaans dualisme, worden nu ook bepaalde definities van de vrije wil ontleed in toetsbare hypothesen en empirisch onderzocht.

Begin jaren 80 werden de neurowetenschappen definitief onderdeel van het vrije wil debat. In een reactietijd-onderzoek liet de neurowetenschapper Benjamin Libet zien dat je hersenen (buiten je bewustzijn om) een fractie van een seconde eerder een handbeweging inzetten dan je bewust die beslissing tot handelen maakt.³ Op basis van dit onderzoek is de controversiële conclusie getrokken dat we geen controle hebben over ons handelen. Onze hersenen beslissen, niet wijzelf. Sinds de jaren 90, *The Decade of the Brain*,⁴ is er veel neurobiologische en psychologische literatuur bijgekomen.⁵ Vaak is er in die onderzoeken doorgebouwd op deze aanname en dat heeft geleid tot een aantal populair wetenschappelijke boeken waarin het bestaan van de vrije wil ontkend wordt. In de ogen van veel hedendaagse neurobiologen is de vrije wil, net als het Cartesiaans dualisme, een achterhaald idee geworden.

Dat zijn aannames die consequenties hebben voor de inrichting van onze samenleving. Uitspraken uit de neurowetenschappelijke hoek over de vrije wil zijn bijvoorbeeld relevant in de rechtspraak. Het zijn immers psychiaters en neurowetenschappers die de toerekeningsvatbaarheid van de verdachten in strafzaken moeten bepalen. Wanneer is er sprake van handelen uit vrije wil en wanneer is er (neurobiologische) onmacht? Maar wat gebeurt er als straks in de rechtspraak het idee overgenomen wordt dat de vrije wil niet bestaat? Is niemand dan meer verantwoordelijk te houden?

² Tiemeijer, W. L. (2011). *Hoe mensen keuzes Maken*. Amsterdam University Press. p. 48

³ Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, 106: pp. 623-642.

⁴ In 1990 heeft het Amerikaanse congres de jaren negentig uitgeroepen tot het "decennium van de hersenen" (The Decade of the Brain). In het daaropvolgende decennium hebben de neurowetenschappen ons begrip van de hersenen erg vergroot. Zelfs nu, een decennium later, is er een continue stroom van ontdekkingen en verhelderingen die vastgeroeste ideeën, over hoe het menselijk brein werkt, verpletteren. Nieuwe begrippen sorteren ook effect in andere disciplines als filosofie maar ook de psychologie, economie, pedagogiek en de rechtswetenschappen. Bron: <http://www.sharpbrains.com/blog/2010/02/23/brain-neuroplasticity-implications/>

⁵ Lamme, V. (2010). *Vrije Wil Bestaat Niet*. Bert Bakker; Swaab, D.F. (2010). *Wij Zijn Ons Brein*. Ruinen: Drukkerij Barriet; Wegner, D. (2002). *The Illusion of Conscious Will*, Cambridge. MA: MIT Press.

Op basis van statistische steekproeven komen Pizarro en Helzer (2010, p. 102) tot de stelling dat we vrije wil toekennen aan mensen omdat we ze verantwoordelijk willen houden. Zodat we ze kunnen straffen of belonen. Als het gaat om moreel oordelen in de rechtspraak zijn er twee modellen. Het eerste, ietwat gedateerde model is van Kohlberg en stamt uit de jaren zeventig, toen er nog een sterke hang was naar een mensbeeld dat uitgaat van vrijheid: mensen handelen op basis van redenen. Het tweede, meer recente, model gaat uit van het idee dat mensen handelen op basis van gevoelens.⁶ Er lijkt dus sprake van een jojo-effect als het gaat om morele verantwoordelijkheid. Na decennialange viering van vrijheid, lijken we nu door te slaan in het bevragen van de houdbaarheid van de menselijke vrijheid.

‘Wat is er mis met het afschaffen van aansprakelijkheid?’, zou je je af kunnen vragen. Er zijn echter voldoende aanwijzingen dat afschaffing niet goed zou zijn voor de samenleving. Recent gedragsonderzoek wijst erop dat mensen die niet in de vrije wil geloven asociaal en amoreel gedrag vertonen.⁷ Persoonlijk vind ik het een demotiverend idee dat ik niet aansprakelijk zou zijn voor een belangrijk deel van mijn doen en laten. Ik ben dan weliswaar niet aansprakelijk te houden voor mijn fouten - dat lijkt mooi - maar ook niet verantwoordelijk voor al het goede dat ik doe.

Het populaire boek *De Vrije Wil Bestaat Niet* van hersenonderzoeker Victor Lamme (2010), somt wetenschappelijk gedocumenteerde gedragingen en gedragsstoornissen op die onze onvrijheid zouden aantonen. De tophersenonderzoeker Dick Swaab noemt de vrije wil een plezierige illusie. De bekroonde neurobioloog Marian Joëls zei onlangs iets vergelijkbaars: “Technisch gezien bestaat de vrije wil niet.”⁸ Zo zijn er talloze prominente neurowetenschappers te vinden die vanuit hun discipline een uitspraak doen over een onderwerp dat traditioneel aan filosofen (en theologen) overgelaten werd.

Deze neurowetenschappers nemen daarbij wel vaak klassiek filosofische definities van de vrije wil als uitgangspunt, bij voorkeur de Libertijnse definitie van vrije wil.⁹ Maar een neurowetenschapper is per definitie een naturalist en een naturalist hangt over het algemeen een deterministisch wereldbeeld aan of dat nu metafysisch determinisme is, of genetisch of neurobiologisch determinisme. In een deterministisch wereldbeeld is in elk geval geen plaats voor de Libertijnse *mogelijkheid tot alternativiteit*. Fysische processen in je hersenen bepalen immers dat je kan maar één ding kan doen; je hebt geen *alternatieve mogelijkheden*. Alleen dit al is een redelijk eenvoudig argument tegen het bestaan van de vrije wil.

Daartegenover staat er een groep filosofen en neurowetenschappers die vanuit neurowetenschappelijke hoek argumenten aandraagt om het bestaan van de menselijke vrije wil te verdedigen. Zij worden compatibilisten genoemd omdat zij menen dat de vrije wil en determinisme samen kunnen gaan, compatibel zijn. Tot de compatibilisten behoren Daniel C. Dennett en Antonio Damasio, de protagonisten in deze scriptie.

⁶ Tiemeijer, W. L. (2011). *Hoe mensen keuzes Maken*. Amsterdam University Press. p. 52

⁷ Tiemeijer, W. L. (2011). pp. 56-57

⁸ Universiteit van Utrecht, alumnidag: *375 jaar Lichaam en Geest* (26 maart 2011)

⁹ Swaab, D.F. (2010). Hoofdstuk XVIII. ‘De Vrije Wil: Een Plezierige Illusie’ in *Wij Zijn Ons Brein*. p. 291.

1.1. Probleemstelling

Er wordt nu al enige decennia vanuit een interdisciplinair kader naar het vrije wil vraagstuk gekeken. Deze interdisciplinaire benadering wordt ook wel de neurofilosofie genoemd.

En de vraag is nu: kan de neurofilosofie uitsluitend geven over het bestaan van de vrije wil?

Is de vrije wil een “plezierige illusie”, zoals Swaab schrijft? Of kan de neurofilosofische benadering ook goede argumenten voor het bestaan van de vrije wil aandragen?

In deze scriptie onderzoek ik de houdbaarheid van de neurofilosofische benadering in het vrije wil debat. Daarbij zet ik de visies van twee prominente vrije wil verdedigers, Daniel C. Dennett en Antonio Damasio, af tegen de gangbare neurowetenschappelijke retoriek.

De probleemstelling onderzoek ik aan de hand van een aantal deelvragen.

- In hoofdstuk 2 onderzoek ik de vraag of het interdisciplinaire kader van de neurofilosofie bruikbaar is in de vrije wil discussie. Is het misschien een puur filosofisch vraagstuk? Moeten we de neurobiologische benadering in zijn geheel uitsluiten van dit debat? Ik bespreek Dennett's heterofenomenologische methode en zijn filosofische rechtvaardiging van het gebruik van de neurofilosofische methode.
- In hoofdstuk 3 vraag ik me af welke oplossingen vanuit het neurofilosofisch kader worden aangedragen voor traditionele problemen in het vrije wil vraagstuk. Ik concentreer me hierbij op Damasio's gebruik van neurowetenschappelijke kennis bij het behandelen van filosofische problemen met betrekking tot de vrije wil.
- In hoofdstuk 4 vat ik het voorgaande samen en vraag me af of het neurofilosofische kader bruikbaar is gebleken met betrekking tot klassieke en actuele problemen in de vrije wil discussie. Is er een weg ingeslagen die ons verder helpt in het vrije wil vraagstuk? Of beter nog: valt vanuit dit kader de vraag naar het bestaan van de vrije wil te beantwoorden?

In Hoofdstuk 2 zal ik onderzoeken of Dennett's ideeën geloofwaardig zijn en de neurofilosofie tot bruikbaar interdisciplinair kader maken voor het denken over de vrije wil. Dit wordt namelijk voetstoots aangenomen door veel neurowetenschappers (zoals Damasio, Swaab, Libet e.a.). Aan de hand van Dennett's verklaringshoudingentheorie en zijn kritiek op de metafysische benadering, probeer ik het betrekken van de neurowetenschappen in de vrije wil discussie filosofisch te rechtvaardigen. In dit hoofdstuk wordt tevens de eerste stap gezet in de argumentatie die de bruikbaarheid van de neurofilosofie voor de compatibilist aan toont.

In Hoofdstuk 3 bespreek ik neurofilosofische argumenten van Damasio en Dennett ter verdediging van de vrije wil. Incidenteel verwijs ik ook naar Dennett's positie in bepaalde vraagstukken. Ik laat zien hoe Damasio het neurofilosofische kader gebruikt bij het beantwoorden van klassieke problemen in het vrije wil debat. Die klassieke problemen worden ook door Dick Swaab erkend en behandeld. In de filosofie bestaat geen overeenstemming over een definitie van de vrije wil, maar Swaab ziet drie (hoofd)voorwaarden waaraan voldaan moet worden om te kunnen zeggen dat de vrije wil bestaat. Ook in het werk van Damasio en Dennett worden deze voorwaarden geadresseerd.

Ik vraag me af of en hoe mensen volgens hen aan deze voorwaarden kunnen voldoen. Wat is de rol van de neurofilosofie in het behandelen van deze drie klassieke problemen?

(1) een handeling is pas vrij als je hem ook niet had kunnen doen (de Libertijnse voorwaarde van 'alternative possibilities')

(2) de handeling moet met een reden zijn verricht

(3) de handeling moet echt uit jezelf komen.

In hoofdstuk 4 zal ik verdedigen dat het juist is om vanuit een neurofilosofisch kader over vrije wil te denken. De interdisciplinaire benadering kan tot nieuwe inzichten leiden. Ook wetenschappelijke ontdekkingen op het gebied van de biologie, sociologie en zelfs de wiskunde kunnen bijdragen aan een beter begrip van menselijke vrijheid. De neurofilosofie is net zo aantrekkelijk voor de incompatibilist als voor de compatibilist. De neurofilosofie is vooral bruikbaar in het ontdekken van de grenzen van de vrije wil en de manieren waarop vrijheid beperkt dan wel vergroot kan worden.

1.2. Auteurkeuze

Tijdens een alumnidag van de Universiteit van Utrecht¹⁰ doet wetenschapsfilosoof Herman Philipse het boek van neurobioloog Victor Lamme *De Vrije Wil Bestaat Niet* (2010) af als leuke bedliteratuur. "Het boek brengt ons aardig *up-to-speed* over trends in de neurofysiologie, maar een onderbouwde definitie van de vrije wil is, net als een argumentatie structuur die het bestaan daarvan zou ontkrachten, niet te vinden", aldus Philipse.

Victor Lamme's werk op deze manier afdoen vind ik echter te ver gaan. Ik zie in zijn benadering juist kansen voor de filosofie van de geest. Empirische kennis van hersenen en gedrag kan leiden tot nieuwe inzichten in het filosofische vraagstuk naar de vrije wil. . Na de kennis tot ons te hebben genomen en te hebben doorgedacht kunnen we immers alsnog beslissen of zo'n nieuw concept past bij onze beleving van de werkelijkheid. In andere woorden: het kan zinvol blijken na te denken over de vrije wil vanuit een interdisciplinair kader zoals de neurofilosofie die is.

Er wordt veel gepubliceerd op het gebied van "de neurofilosofie", daarom heb ik in deze thesis een keus moeten maken voor een beperkt aantal teksten, een beperkt aantal auteurs. In mijn optiek zijn Daniel C. Dennett en Antonio Damasio twee centrale spelers. Zij hangen beide een naturalistisch wereldbeeld aan en verwerpen het substantiedualisme. Tegelijkertijd verdedigen ze het bestaan van de vrije wil. Met andere woorden: het zijn beide compatibilisten.¹¹ Ik heb voor hen gekozen op basis van de volgende criteria:

- de auteur moet neurowetenschappelijke vooruitgang volgen
- de auteur moet thuis zijn in de *Philosophy of Mind*-discipline
- de auteur moet veel geciteerd worden in dit interdisciplinaire debat
- de auteur moet onderbouwde syntheses maken op basis van de status quo in dit debat

¹⁰ Philipse, H. "Het dualisme tussen lichaam en geest revisited." 375 jaar Lichaam en Geest [universiteitsdag]. Universiteit van Utrecht. 26 maart 2011.

¹¹ Compatibilisme is het geloof dat vrije wil en determinisme compatibele ideeën zijn

Daarnaast moeten de gekozen auteurs representatief zijn voor de neurofilosofische methode. De definitie van neurofilosofie¹² omschrijft twee benaderingen.

Neurofilosofie is de interdisciplinaire studie van neurowetenschap en filosofie. Er zijn hier binnen globaal twee benaderingen te onderscheiden. De eerste benadering houdt in dat men oude filosofische vraagstukken zoals dualisme, materialisme, reductionisme en lichaam-geestinteractie benadert vanuit het perspectief van de neurowetenschappen. De tweede benadering tracht resultaten van neurowetenschappelijk onderzoek te verklaren vanuit de denkwijze en methoden van de wetenschapsfilosofie.¹³

Naar mijn inzicht vertegenwoordigen Damasio en Dennett deze twee neurofilosofische benaderingen. Antonio Damasio is neurobioloog en vertegenwoordigt de eerste benadering. Hij doet uitspraken over filosofische onderwerpen als (zelf)bewustzijn, de verhoudingen tussen rationaliteit, emotie en vrije wil. Dit doet hij nagenoeg uitsluitend uitgaand van neuropsychologische feiten.

De filosoof Daniel C. Dennett vertegenwoordigt de tweede benadering. Hij viel mij tijdens mijn studie toegepaste cognitieve psychologie al op. In het lijvige introductieboek *Cognitive Neuroscience: The Biology of the Mind*, van Gazzaniga, Ivry en Mangun (2002) stonden interviews met een twintigtal sleutelfiguren uit de neurowetenschap. De enige filosoof die daar tussen stond was Dennett. Hij is thuis in de wetenschapsfilosofie, metafysica en de filosofie van de geest. Op grond van deze intellectuele bagage doet hij uitspraken over kwesties in de neurowetenschappen, de mathematica en de (meta)fysica.

Dennett wordt vaak in het kamp van de vrije wil ontkenners gezet; ten onrechte, dat vindt hij zelf ook.¹⁴ Hij gelooft namelijk weldegelijk in een vorm van vrije wil. Antonio Damasio is eveneens een bepleiter van een empirisch onderbouwd concept van de vrije wil. Zij zijn het op veel fronten met elkaar eens. Door Damasio naast Dennett te zetten hoop ik het verband tussen hun ideeën te doorgronden.

Deze scriptie heeft de structuur van een wetenschappelijk artikel. Eerst worden de probleemstelling (hypothese) en vraagstelling geformuleerd, vervolgens wordt de onderzoeksmethode uitgelegd en verantwoord (in dit geval het neurofilosofisch denkkader), dan wordt er een opsomming van de onderzoeksresultaten gegeven, en tot slot worden conclusies getrokken over de houdbaarheid van de hypothese.

Definities van de vrije wil zijn soms schimmig en vaak tegengesteld. Dat komt naar mijn idee doordat verschillende perspectieven van begrijpen en verklaren door elkaar worden gebruikt. Zo ontstaan misverstanden en tegenstellingen die er in werkelijkheid niet zijn. Dat is een handicap in het hedendaagse debat over vrije wil.

¹² Meer over neurofilosofie hierna

¹³ Prominenten in de neurofilosofie zijn: Paul en Patricia Churchland, Antonio Damasio en in Nederland Johan de Boer. Neurofilosofie. (2011, maart 4). *Wikipedia, de vrije encyclopedie*. Opgehaald 14:49, maart 4, 2011 van <http://nl.wikipedia.org/w/index.php?title=Neurofilosofie&oldid=24676480>.

¹⁴ Dennett, D. (2003) *Freedom Evolves*. London: Penguin Books. p. 15

Een voorbeeld van een schijnbare tegenstelling in de neurofilosofische arena zijn de conclusies die de psycholoog Daniel Wegner trekt in het veelbesproken boek *The Illusion of Conscious Will* (2002).¹⁵ Wegner argumenteert dat wij werkelijk beschikken over verantwoordelijkheid voor onze daden, maar hij noemt dit geen vrije wil. Wegner ziet vrije wil als een waarheidsgetrouwe illusie. Daniel Dennett snapt hier niets van. Waarheidsgetrouwe illusie? Dennett vindt deze uitspraak grenzen aan tegenstrijdigheid en bovenal onnodig verwarrend. Hij adviseert Wegner gewoon te zeggen dat de vrije wil echt is en bestaat (ook al is het een kwetsbaar fenomeen met notabele beperkingen).¹⁶

In het volgende hoofdstuk zal ik ingaan op de verschillende perspectieven van verklaren en begrijpen. Een besef van het bestaan van verschillende verklaringshoudingen/ begripsniveaus is namelijk cruciaal in de vrije wil discussie. Dit besef is bovendien bruikbaar in het ontmaskeren van schijnbare tegenstellingen. Het is belangrijk om de relevantie van verschillende verklaringshoudingen (*stances*) in te zien en te betrekken in het vrije wil debat. De heterofenomenologische methode van Dennett is het verklaren van geestfenomenen door jezelf op een bepaalde manier te bedienen van de diverse verklaringshoudingen. In mijn optiek doet de neurofilosofie dat, of zou ze het moeten doen. Dennett en Damasio doen het in ieder geval.

In het kader van deze thesis is de heterofenomenologie (met de verschillende verklaringshoudingen) een belangrijke methode in het laten zien dat de neurofilosofie ten opzichte van Peter van Inwagen's metafysische methode bruikbaar is in het denken over de vrije wil. De heterofenomenologie is bovendien een methode waar Damasio en Dennett gebruik van maken die hen helpt argumenten te vinden die het bestaan van de vrije wil onderbouwen.

¹⁵ Wegner, D. (2002) *The Illusion of Conscious Will*, Cambridge MA: The MIT Press

¹⁶ Dennett, D. (2003) p. 224

2. De relevantie van de neurofilosofie

In deze scriptie concentreer ik me op de neurofilosofische bijdrage aan het vrije wil debat. Zoals in hoofdstuk 1 al geformuleerd: de neurofilosofie kent twee denkrichtingen. Het is enerzijds een discipline dat neurowetenschappelijke kennis gebruikt om uitspraken te doen over filosofische concepten maar het is ook een discipline dat wetenschapsfilosofisch reflecteert op de methoden van de neurowetenschap.

Antonio Damasio laat in zijn diverse boeken zien hoe neurowetenschappelijke kennis bij kan dragen aan een filosofisch inzicht in het geestfenomenen als bewustzijn, rationaliteit en emotionaliteit en het zelf-proces. Dennett reflecteert middels een filosofie van de wetenschap benadering op neurowetenschappelijke aannames. Ik neem in hoofdstuk 3 de vrijheid om hun argumenten te lenen om aan te tonen dat het compatibilistisch is om te zeggen dat we uit onszelf kunnen handelen (en niet alleen door externe factoren worden gestuurd), dat we handelen naar redenen (en niet alleen naar impulsen) en dat we beschikken over de mogelijkheid tot alternativiteit. Dit zijn de basisvoorwaarden, zoals ook Dick Swaab ze formuleert, die voldaan moeten worden om te kunnen zeggen dat de mens (de potentie tot) vrije wil heeft.

Maar om te kunnen zeggen dat de neurofilosofie inzetbaar is bij het bespreken van het vrije wil vraagstuk, moet eerst worden bewezen dat neuropsychologische kennis relevant is. De discussie over het al dan niet bestaan van de vrije wil wordt immers vaak gesmoord door het incompatibilistische argument van Peter van Inwagen, namelijk dat vrije wil incompatibel is met metafysisch determinisme. Los van de vraag of de mens vrije wil heeft, is er de vraag of vrije wil in een deterministische wereld überhaupt denkbaar is. Van Inwagen (en vele anderen) menen van niet. Dennett is zich ervan bewust dat je dit metafysische probleem moet behandelen voordat je het over het brein kan hebben. Hij onderkent dat mensen het idee van de vrije wil als een wetmatigheid ervaren, maar vindt dat de fysica en metafysica ons er nagenoeg niets over leren.

We count on people 'having free will' the same way we count on them falling when pushed off cliffs and needing food and water to live, but it is neither a metaphysical background condition nor a fundamental physical condition.¹⁷

Bij het bespreken van geestfenomenen gaan neurowetenschappers als Dick Swaab, Daniel Wegner maar ook Antonio Damasio voorbij aan deze discussie over de positie van de vrije wil ten opzichte van de fysische wetten en/of metafysische voorwaarden. De neurofysiologie bevraagt de populair wetenschappelijke natuurwetten niet en gaat niet in op metafysische vraagstukken als: 'is het universum deterministisch of indeterministisch ingericht?'. Ze gaat voorbij aan de metafysische voorwaarden van vrije wil en aan wat Dennett *The physical stance* noemt, namelijk het begrijpen van het vrije wil concept vanuit het perspectief van de fysica. Zijn de ideeën van Swaab en Damasio daarmee irrelevant? Dennett meent van niet. *The physical stance* kan indrukwekkende verklaringen,

¹⁷ Dennett. D. (2003). p. 10.

beschrijvingen en voorspellingen opleveren over fenomenen in de objectief bestudeerbare werkelijkheid, maar voor het verklaren van geestfenomenen is de bruikbaarheid zeer beperkt, zegt Dennett.¹⁸

In de komende paragrafen ga ik aantonen dat filosoferen over de vrije wil vanuit een metafysisch perspectief zoals Van Inwagen dat doet, beperkt zinvol is. Daarna zal ik aan de hand van Dennett's omschrijving van een aantal alternatieve verklaringshoudingen (*stances*) voor de fysische benadering laten zien dat het zinvol is om voorbij te gaan aan de fysica en metafysica in het vrije wil debat. Het bewust schakelen tussen die alternatieve verklaringshoudingen is bovendien een handige methode voor de neurofilosoof. Dennett noemt dit de heterofenomenologische methode. Antonio Damasio wordt gezien als een sleutelfiguur in de neurofilosofie en in zijn werk is duidelijk te zien dat hij zich bedient van deze heterofenomenologische methode.¹⁹ Maar nu eerst een reflectie op Van Inwagen's determinisme argument tegen de vrije wil.

2.1. Determinisme en compatibilisme

Menig filosoof ziet de metafysica als het belangrijkste uitgangspunt in het onderzoeken van het vrije wil vraagstuk. Zo ook de Amerikaanse filosoof Peter van Inwagen. Hij meent dat de natuurwetten ons laten zien dat de werkelijkheid deterministisch van aard is. De vrije wil is volgens Van Inwagen incompatibel met determinisme. In de natuurwetten ziet hij een argument om het bestaan van de vrije wil te ontkennen.

Er moet wel aangegeven worden dat metafysica en fysica niet hetzelfde zijn al is er overlap in termen en methodes. Hard geformuleerd is de fysica een beschrijvende wetenschap en niet een voorspellende wetenschap zoals de natuurfilosofie, onderdeel van de metafysica. Toch wordt ook in de fysica gezocht naar de samenhang tussen verschijnselen en daarin heeft het ook veel raakvlakken met de metafysica.²⁰ Peter van Inwagen roept ons op om natuurwetten en natuurfilosofie op één hoop te gooien omdat de meeste fysici dat ook zouden doen.²¹

Het is overigens niet vanzelfsprekend dat ons universum deterministisch is.²² Desondanks beslist Dennett mee te gaan met Van Inwagen in het idee dat determinisme waar is. Het is het meest populaire standpunt in de wetenschap. Zelfs nieuwe theorieën over quantum indeterminisme hebben het idee niet weg kunnen nemen dat gevolgen altijd een oorzaak hebben. Op praktisch niveau is causaliteit echt, net als determinisme. Dennett bekritiseert echter wel het klassieke idee dat metafysisch determinisme de 'mogelijkheid tot alternativiteit' uitsluit.

¹⁸ Dennett. D. (2003). pp. 39-40

¹⁹ Damasio. A. (2010). p. 15

²⁰ Natuurkunde. (2012, december 11). *Wikipedia, de vrije encyclopedie*. Opgehaald 09:38, december 11, 2012 van //nl.wikipedia.org/w/index.php?title=Natuurkunde&oldid=33953082

²¹ Zie: Van Inwagen, P. (1983), *An Essay on Free Will*, Oxford: Clarendon Press, p. 3

²² Er zijn aanwijzingen dat op quantum niveau de wereld indeterministisch is. Een theorie terug te vinden in de quantum fysica. Zie: Quantum indeterminacy. (2012, November 10). In *Wikipedia, The Free Encyclopedia*. Retrieved 14:27, December 14, 2012, from http://en.wikipedia.org/w/index.php?title=Quantum_indeterminacy&oldid=522265459

Peter van Inwagen definieert determinisme als volgt: “[*Determinism is the thesis that*] there is at any instant exactly one physically possible future.”²³ Hij ziet determinisme als tegenstrijdig met het idee van de vrije wil. Die contradictie wordt ervaren als men uitgaat van de Libertijnse definitie van vrije wil waarbij het bestaan van de vrije wil wordt bepaald door de vraag of er een ‘mogelijkheid tot alternativiteit’ bestaat:

Iemand is vrij met betrekking tot H wanneer die persoon een keuze heeft met betrekking tot H.

In andere woorden: een beslissing/handeling kan alleen vrij zijn als op tijdstip “t” waarop de beslissing/handeling genomen wordt de beslissing ook niet genomen had kunnen worden. Als determinisme waar is dan is deze mogelijkheid tot alternativiteit er niet, aldus Van Inwagen.²⁴

Dennett verzet zich tegen dit idee. Alternatieve mogelijkheden bestaan in een deterministische werkelijkheid, meent hij. Dennett onderbouwt dit door te zeggen dat:

(A) determinisme is niet hetzelfde als: *onvermijdelijkheid*; je hebt mogelijkheden

(B) indeterminisme is niet noodzakelijk voor vrije wil

Volgens Dennett bestaat de mogelijkheid tot alternativiteit in een deterministische wereld omdat daarin ook zoiets kan bestaan als ‘vermijden’. Dennett bekritiseert het feit dat nooit ingegaan wordt op de ambivalentie van het concept ‘mogelijkheid tot alternativiteit’. De mogelijkheid tot alternativiteit stelt dat het mogelijk moet zijn uitkomsten te veranderen om te kunnen zeggen dat vrijheid bestaat. Dit is een tegenstrijdige eis. Dennett stelt een aanvulling voor die dit criterium van zijn tegenstrijdigheid verlost. Dennett pleit dat de enige manier waarop je de zin: “veranderen van een uitkomst” kan begrijpen is door het te herformuleren in “veranderen van een geanticipeerde uitkomst”. De werkelijke uitkomst kan niet veranderd worden. Ook niet in een indeterministische wereld. Alleen een geanticipeerde uitkomst kan vermeden worden.

Determinisme maakt het veranderen van (geanticipeerde) uitkomsten juist makkelijker. In een deterministische wereld is namelijk anticiperen (succesvol voorspellen) mogelijk.

Volgens Dennett is determinisme zelfs compatibel met het idee van *events without a cause*. Een voorbeeld. Het heeft geen oorzaak dat het kop is na het opgooien van een muntje. Ten eerste blijft het praktisch onvoorspelbaar of het kop of munt wordt. We kunnen namelijk niet één oorzaak aanwijzen die de uitkomst heeft bepaald. Er is ook niet één situatie noodzakelijk voor de uitkomst. Immers, het is een samenspel van verschillende elementen (worp, wind, luchtvochtigheid etc.). Een ander samenspel van diezelfde elementen kunnen evengoed tot het tegenovergestelde resultaat leiden. We moeten bovendien erkennen dat de uitkomst niet slechts door één factor veroorzaakt kan worden, maar meerdere factoren de oorzaak kunnen zijn. Sterker nog, het kunnen verschillende combinaties van dezelfde factoren zijn om te resulteren in eenzelfde uitkomst of in een andere uitkomst. Een uitkomst zonder één oorzaak.

²³ Van Inwagen, P. (1983), *An Essay on Free Will*, Oxford: Clarendon Press, p. 3

²⁴ Van Inwagen, P. (1975) The Incompatibility of Free Will and Determinism, in *Philosophical Studies* 27, pp. 185-199

Om over vrije wil te beschikken, moet je over een aantal mechanismen beschikken. Formuleren wat die mechanismen zijn doe je niet door naar de effecten van determinisme te kijken in *the physical stance*. Relevante mechanismen leg je bloot door *the design stance* aan te nemen, zoals de neurowetenschapper doet in het bestuderen van (neuro)biologische systemen. In de volgende paragraaf meer over de verklaringshouding: *the design stance*.

Dennett laat zien dat strikt metafysisch naar het vrije wil onderwerp kijken ons niet zo veel leert. Bovendien is indeterminisme niet het nodige alternatief. Belangrijk is om te beseffen dat vrije wil beter te bestuderen valt door alternatieve verklaringshoudingen aan te nemen dan Van Inwagen's (meta)fysische perspectief. Deze alternatieve perspectieven sluiten de neurowetenschappen en gedragswetenschappen in als bron van relevante kennis met betrekking tot het vrije wil vraagstuk.

Door kennis te nemen van de verschillende begripsniveaus/verklaringshoudingen voorkom je schijnbare tegenstellingen. Iets wat waar is, als je uitgaat van neurofysiologische kennis, kan twijfelachtig zijn als je uitgaat van de fysica.

Een voorbeeld. De wetten van Newton zijn vandaag de dag nog steeds bruikbaar in de klassieke mechanica bij het voorspellen en verklaren van fysieke beweging. Ook in de neurowetenschap wordt uitgegaan van causale mechanische principes als het gaat om hersenactiviteit.²⁵ Onder de meeste omstandigheden zijn de wetten van Newton waar, echt en betrouwbaar. Er is echter ontdekt dat als het gaat om beweging met een snelheid die de lichtsnelheid benadert, de wetten van Newton niet specifiek genoeg zijn.²⁶ Sterker nog: in die situatie zijn de wetten van Newton onhoudbaar. Dit betekent dat onder bepaalde omstandigheden gepleit kan worden voor de wetten van Newton en onder andere omstandigheden tegen. Zo is er sprake van een schijnbare tegenstelling: de waarheid of onwaarheid van een stelling is afhankelijk van de context waarin hij bevestigd wordt.

Uit dit voorbeeld leren we dat we ons bewust moeten zijn over de omstandigheden waarin we fenomenen als de vrije wil onderzoeken. Een bewijs dat voor of tegen de vrije wil wordt gevonden onder bepaalde omstandigheden kan mogelijk verworpen worden onder andere omstandigheden.

De neurofilosofie bij uitstek moet rekening houden met dergelijke verklaringshoudingen. De fysische houding is beperkt maar eveneens uitsluitend uitgaan van neuroanatomische kennis is een beperkte strategie. Een eenzijdige fysische of neurofysiologische behandeling van complexe fenomenen als bewustzijn en vrije wil levert contradicties op. De vrije wil is iets waarop gereflecteerd kan worden vanuit een subjectief, een intersubjectief en objectief perspectief. Vrije wil heeft ook verschillende betekenissen in deze verschillende perspectieven. Eenzijdig kijken naar neuroanatomie of hersenactiviteit, waar Paul en Patricia Churchland voor pleiten, beperkt je kennisvoortgang en kennisniveau. De heterofenomenologie van Dennett biedt de neurofilosoof meer ingangen om een compleet maar nog steeds materialistische omschrijving van de vrije wil te realiseren. Hieronder laat

²⁵ Schwartz, J. M., Stapp, H. P., & Beauregard, M. (2005). Quantum physics in neuroscience and psychology: a neurophysical model of mind–brain interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1458), p. 1310.

²⁶ Newton's laws of motion. (2013, June 30). In *Wikipedia, The Free Encyclopedia*. Retrieved 20:31, July 1, 2013, from http://en.wikipedia.org/w/index.php?title=Newton%27s_laws_of_motion&oldid=562207396

ik zien waarom die verschillende verklaringshoudingen nodig zijn en hoe die tot elkaar verhoudt moeten worden.

2.2. Verschillende perspectieven van verklaren

In *Brainstorms* (1981)²⁷ introduceert Dennett al een aantal “houdingen van verklaren” (*stances*) bij het bestuderen van gedrag. Later heeft hij zijn verklaringshoudingen theorie verder uitgediept en aangevuld. Dennett noemt het switchen dus deze verschillende verklaringshoudingen de heterofenomenologische methode.

De neurofilosofie is een relatief jonge benadering van geestproblemen. Damasio bedrijft de neurofilosofie en maakt daarbij gebruik van de heterofenomenologische methode, zo merkt Dennett op.²⁸ En dat is ook mijn beleving na het vergelijken van de heterofenomenologische methode van Dennett met Damasio’s methodiek in het boek *Self Comes to Mind*.

De bruikbaarheid van de verschillende verklaringshoudingen bij het onderzoeken van fenomenen van de geest – zoals de vrije wil - loopt echter uiteen. Hieronder geef ik eerst een opsomming van de verklaringshoudingen die Dennett onderscheidt. Daarna bespreek ik aan de hand van Dennett’s argumentatie waarom de verklaringshoudingen ingezet moeten worden zoals dat gebruikelijk is voor de heterofenomenologie. Dit om uiteindelijk een uitspraak te kunnen doen over de toepasbaarheid van de neurofilosofie zoals Damasio die bedrijft.²⁹ De verschillende verklaringshoudingen zijn:

(1) *the physical stance* (het (meta)fysisch perspectief)

Geestfenomenen worden verklaard vanuit de natuurwetten. Gedrag van een organisme wordt uitgelegd in termen van natuurfenomenen, zoals materie, straling, energie, kracht etc.

(2) *the design stance* (het design perspectief)

Gedrag van een organisme wordt verklaard in teleologische termen: hoe werken de componenten van het organisme samen zodat het organisme doet wat het doet. Het gedrag wordt uitgelegd in termen van de functie die het heeft in het systeem van het organisme als geheel.

(3) *the intentional stance* (het intentionele perspectief)

Gedrag van een organisme wordt verklaard vanuit de *redenen* en *motieven* die het organisme heeft voor dat gedrag.

(4) *het evolutionair perspectief*

Gedrag van een organisme wordt verklaard vanuit het evolutionaire nut. Gedrag wordt verklaard in termen van overlevingskansen.

(5) *introspectie*

Introspectie, ook wel zelfreflectie genoemd, is een activiteit waarbij getracht wordt om

²⁷ Dennett, D. (1981), *Brainstorms*. Cambridge, MA: MIT Press. pp. 3-22

²⁸ Dennett, D. (2005). *Sweet dreams: Philosophical obstacles to a science of consciousness*. The MIT Press. p. 146

²⁹ Antonio Damasio, als sleutelfiguur in de neurofilosofie, maakt gebruik van de heterofenomenologie.

geestfenomenen en gedrag te verklaren door de eigen subjectieve gedachten, gevoelens en herinneringen te overdenken en te beschrijven.³⁰

(6) *moral or personal stance* (het morele perspectief)

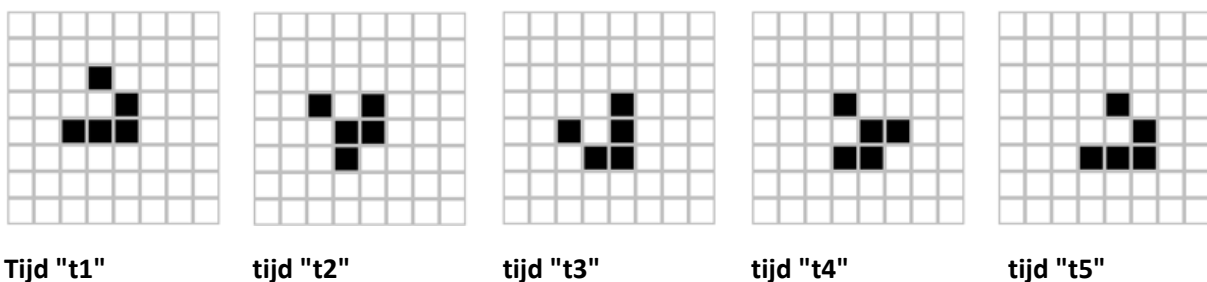
Gedrag wordt verklaard vanuit de voor een individu of groep geldende moraal.

THE PHYSICAL STANCE

Hierboven heb ik al kort uitgelegd wat de beperkingen van de metafysica zijn in het bespreken van het vrije wil vraagstuk. Om de beperktheid aan te tonen van de fysische benadering in het verklaren en voorspellen van gedrag van organismen kan een gesimplificeerd model van de werkelijkheid dienstbaar zijn, zegt Dennett;³¹ een soort laboratorium opstelling van de werkelijkheid waarin je onder gecontroleerde omstandigheden de effecten van determinisme op gedrag kunt onderzoeken.

Hij stelt voor gebruik te maken van J.H. Conway's *The Game of Life* (of in het kort: *Life*).³² In *Life* zijn ruimtes gedigitaliseerd. Het is een tweedimensionale versie van de werkelijkheid. Er zijn pixels in *Life* die aan of uit kunnen (zwart of wit zijn). *Life* kan een afmeting hebben van 8 bij 8 (zie figuur 1) maar er is later ook met 'universa' geëxperimenteerd die een veel grotere resolutie hebben. Er is geen willekeur in dit universum, het is immers deterministisch. Er zijn regels. *Life* heeft een aantal (natuur)wetten, de zogenaamde transitieregels. Het zijn wetten die bepalen hoe dit specifieke universum eruit ziet in de volgende tijdseenheid ten opzichte van de huidige tijdseenheid. Door die regels ontstaan er veranderingen in *Life*.

De transitieregels van *Life* luiden als volgt: als een nieuwe tijdseenheid aanbreekt doe het volgende: tel bij elke pixel op je speelveld hoeveel van de aangrenzende acht pixels AAN zijn. Als die hoeveelheid precies twee is dan blijft de pixel ongewijzigd (AAN of UIT). Als die hoeveelheid precies drie is dan gaat de pixel AAN. Bij alle andere hoeveelheden gaat de pixel UIT of blijft die UIT.



Figuur 1: levensverloop van een organisme bekend als een *slider*, in Conway's Game of Life. Door slaafs de wetten van *Life* te volgen doet de slider in vijf tijdseenheden een stapje naar rechtsonder.

³⁰ In *Brainstorms* (1981) definieert Dennett ook een *moral or personal stance*. Dit verklaringmodel is wel relevant in het vrije wil vraagstuk maar niet relevant in het verantwoorden van de neurobiologie als relevant perspectief.

³¹ Dennett, D. (2003) pp. 25-29

³² Dennett, D. (2003) p. 36

Klinkt als een simpel universum, toch? Niets is minder waar. Zelfs onder deze vereenvoudigde omstandigheden blijkt *the physical stance* beperkt bruikbaar. Een dergelijk universum is enorm complex. Haast te complex om praktisch mee te kunnen werken in het nadenken over gedrag. Dit universum heeft bij één tijdseenheid al 18.446.744.073.709.551.616 ($=2^{64}$) mogelijke staten waarin het kan verkeren. Bij tien tijdseenheden heb je 18.446.744.073.709.551.616¹⁰ mogelijke sequenties waar dit universum uit kan bestaan. Heel, heel veel staten om te bestuderen dus. En dan is dit nog een buitengewoon simpel universum dat nauwelijks te vergelijken is met de complexiteit van de 'echte' wereld.

Zijn belangrijkste kritiek op het gebruik van *the physical stance* in de vrije wil discussie is dan ook dat we moeten beseffen hoe snel zaken onvoorstelbaar complex kunnen worden als je gedragingen van levende wezens vanuit een (sub)atomair niveau wilt bestuderen, verklaren en/of voorspellen.

Met alle respect geef ik hier een aanvulling op de theorie van Dennet. Het is onjuist om te beweren dat geen enkele vorm van gedrag te bestuderen valt vanuit *the physical stance*, onder kunstmatig vereenvoudigde omstandigheden. Dat doen fysici immers voortdurend.³³ Het gaat dan echter vaak om het gedrag van die allerkleinste deeltjes en niet om het gedrag van organismen waar die deeltjes onderdeel van uitmaken. Ik wil echter wel met Dennett mee gaan in zijn stelling dat gedrag van een cel (vele atomen), laat staan van een complex organisme (dan wel de 2D-variant daarop), praktisch en theoretisch onmogelijk te volgen is door alleen naar het atomaire (pixel) niveau te kijken.

Dennett stelt voor om een alternatieve verklaringshouding aan te nemen en zo alsnog de waarheid van (on)vermijdelijkheid en (alternatieve) mogelijkheden te onderzoeken in *Life*. Hij zegt dat je het gedrag ook kunt proberen te verklaren door naar de vormgeving van het organisme te kijken. De verklaringshouding die je dan aanneemt is het design perspectief (*taking a design stance*).

THE DESIGN STANCE

Om te ontdekken of een organisme een vermogen tot vermijden bezit (een alternatieve mogelijkheid heeft) moeten we kijken naar zijn *design*, meent Dennett.³⁴ Door simpele entiteiten in *Life* te bestuderen op vormgevingsniveau leggen we mechanismen bloot die minimaal noodzakelijk zijn voor vrije wil.

In John Conway's *Game of Life* kun je voor *Hacker God* spelen. Je kan de staat van *Life* "hacken" en zo een God zijn. Als een soort designer bepaal je dan welke pixels bij aanvang van de simulatie (het startpunt van dat universum) aan of uit zijn. Als er sprake is van een groepering pixels dan wordt dit een organisme, of *creatie*, genoemd. Als de simulatie start en de klok gaat lopen, verandert *Life* bij elke volgende tijdseenheid volgens de transitierregels. In figuur 1 zie je hoe een creatie, bekend als een *slider*, een stap naar rechtsonder zet op het grid van 8 bij 8. Door *Life* te spelen, leer je hoe moeilijk het is een creatie te ontwerpen die kan overleven zodra de ruimte groter wordt en andere creaties zich rondom je creatie bevinden (en in collisioneers liggen). Het bewegen zelf kan al fataal

³³ Opvallend genoeg wordt bij het beschrijven van gedragingen van atomen ook over eigenschappen of vermogens van die atomen gesproken.

³⁴ Dennett, D. (2003) p. 56

zijn voor je creatie, laat staan het botsen met andere creaties. Fataal in de zin dat het je creatie zijn herkenbaarheid of absolute bestaan verliest (alle pixels gaan op wit).

Life is een perfect deterministische wereld en toch is de consistentie van een *slider* beter te verklaren door naar zijn vormgeving te kijken dan sec naar de transitierregels (de natuurwetten van *Life*). De transitierregels zijn voor alle creaties hetzelfde, maar de uitwerking is verschillend door de verschillen in het *design* van de verschillende creaties. Sommige creaties kunnen beter overleven dan andere in *Life*. Zo is een *slider* een creatie die zijn identiteit behoudt door de tijd heen. Een *slider* kan dus goed overleven totdat die bijvoorbeeld in een botsing komt met een andere creatie.

Neurowetenschappers hanteren over het algemeen ook *the design stance*. Om extern of intern gedrag bij een mens te verklaren wordt er bijvoorbeeld gekeken naar de functionele architectuur en samenhang van relevante hersendelen. Er wordt gekeken naar de vormgeving van de hersenen niet naar de natuurwetten, die worden als een constante geaccepteerd. Soms is echter menselijk gedrag te complex om uit leggen in termen als architectuur en design. Denk daarbij bijvoorbeeld aan sociaal gedrag. Een betere verklaringshouding is dan *the intentional stance*.

THE INTENTIONAL STANCE

Je kunt het gedrag in *Life* vaak verklaren door te kijken naar de vormgeving (zoals hierboven is aangetoond), maar je kunt het soms zelfs nog beter verklaren door het organisme op te vatten als een creatie met intenties (*taking the intentional stance*).

Als Bart een nieuwe collega ontmoet op het werk, dan geeft hij die persoon een hand. Waarom doet hij dat? Je kunt het gedrag van Bart proberen uit te leggen aan de hand van de natuurwetten. Dan neem je *the physical stance* aan. Het is onwaarschijnlijk dat je dan tot een bruikbare verklaring komt. De vraag is zelfs of we op (sub)atomair niveau Bart kunnen onderscheiden van zijn kleding, van de lucht en van de nieuwe collega.

Je kunt het gedrag ook proberen te verklaren door naar de hersenen van Bart te kijken. Je neemt daarbij *the design stance* aan. Mogelijk vind je wat aanwijzingen over de oorzaken van het gedrag als je de activiteiten in Barts hersenen gadeslaat. Bepaalde hersengebieden worden actief als Bart de nieuwe persoon ontdekt en andere hersengebieden leven op als hij zijn lichaam (moet) gaat bewegen, etc.

Barts gedrag is het beste uit te leggen vanuit het *intentionele perspectief (the intentional stance)*. Bart geeft een hand omdat hij vindt dat je nieuwe mensen op het werk altijd een hand moet geven. Dat is een gangbare mening, maar deze geeft ook de meest volledige verklaring. In dit geval vind je de beste verklaring van dit gedrag door een (rationeel) intentioneel perspectief te kiezen.

Ander gedrag van Bart is weer beter te verklaren vanuit *the design stance*. Als een dokter met een hamertje op zijn knie tikt waardoor Bart's onderbeen omhoog schiet, is *the design stance* relevanter, evenals wanneer Bart's hersenen bepaalde hormonen vrijgeven omdat hij honger heeft. Gedragingen van Bart zullen echter zelden vanuit *the physical stance* te verklaren zijn.

The Game of Life bewijst dat een creatie weerbaarder wordt als hij zo is ontworpen dat hij toekomstig onheil kan anticiperen.³⁵ Hoe verder je gaat in het verfijnen van de vormgeving van je creatie (en daarmee de creatie weerbaarder maakt tegen de omstandigheden in *Life*), des te complexer deze creatie wordt. Je zal merken dat het gedrag van je creatie op een bepaald moment makkelijker te verklaren is vanuit zijn intenties dan vanuit zijn ontwerp. Je kiest dan niet langer voor *the hacker God's perspective* in het verklaren van het gedrag, zoals bij *the design stance*, maar voor *the perspective of the hacker God's creation*. Met andere woorden: je kent je creatie een vorm van intentionaliteit toe en verklaart zijn gedrag vanuit zijn bedoelingen. Dennett:

Enriching *the design stance* by speaking of configurations as if they “know” or “believe” something, and “want” to accomplish some end or other is moving up from the simple *design stance* to what I call *the intentional stance*. [our creations] have been reconceptualized as rational agents or intentional systems.³⁶

Conway's team hebben in *Life* buitengewoon “slimme” creaties ontworpen. Voorbeelden zijn een zelfreproducerende *Universal Turing Machine* en een computer die een veelheid aan berekeningen kan doen. De creaties zijn dermate geavanceerd geworden dat hun gedrag niet langer uitgelegd kan worden in termen van *design*. Dat zou een onnavolgbare uitleg opleveren met opnieuw een enorme mate van complexiteit. In deze gevallen blijkt het efficiënter om het gedrag uit te leggen in termen van bedoelingen (*the intentional stance*). Neuro- en gedragswetenschappers worden met hetzelfde probleem geconfronteerd. Sommige gedragingen, zoals complexe morele beslissingssequenties, zijn moeilijk te bestuderen zonder daarbij rekening te houden met de intenties van de persoon.

Jürgen Habermas bekritiseert de poging om neurowetenschap te verbinden met de discussie over agentschap (een discussie die gerelateerd is aan de ‘hebben we vrije wil?’-discussie).³⁷ Habermas verdedigt dat de winnaar in een debat slechts gekozen kan worden op basis van logische deductie en niet door naar de causale uitkomst van processen in de hersenen te kijken.³⁸

Ten onrechte ziet Habermas hierin een reden om de neurowetenschap te weren van het debat over agentschap. John R. Searle merkt op dat Habermas weliswaar gelijk heeft als hij stelt dat een debat beslist moet worden door de geponeerde argumenten logisch te wegen en te vergelijken, maar om tot een beter begrip te komen over de mate van agentschap kan het neurofysiologische (*design*) perspectief heel goed naast het intentionele perspectief worden gelegd. Searle:

[...] just as the speech act has a level of description where it is a sound, it also has a description where it has semantic properties; so the thought has a level of description where

³⁵ Dennett verwijst hier naar de fysisch Jorje Wagensberg (2000) die, zonder *The Game of Life* te betrekken, op basis van definities van: informatie, onzekerheid en complexiteit, berekent dat *anticipatie* de beste overlevingsstrategie is van *biota* (= *vitae of leven*). Waarbij overleven als *persistance* of identity wordt gezien en afhankelijk is van de mate van autonomie ten opzichte van een onzeker milieu.

³⁶ Dennett. D. (2003). p 45.

³⁷ Habermas, J. (2007). The Language Game of Responsible Agency and the Problem of Free Will: How can epistemic dualism be reconciled with ontological monism?. *Philosophical Explorations*, 10(1), 13-50.

³⁸ Habermas, J. (2007). p 25.

it consists of neurobiological processes, it also has another level of description where it has intentionalistic and indeed semantic properties.³⁹

De uitspraak van Searle doet me enigszins twijfelen over de houdbaarheid van de neurofilosofie in het onderzoeken van het vrije wil vraagstuk. De neurofilosofie is dus slechts bruikbaar als ook *the intentional stance* gerekend wordt tot die methode. Er wordt door Searle gesuggereerd dat gedrag ook verklaard moet worden in termen van intenties en niet alleen in termen van vormgeving en functie. Deze observatie van Searle pleit voor de heterofenomenologie als basis van de neurofilosofie. Wat blijkt, Searle maakt het in hetzelfde artikel aannemelijk dat *the intentional stance* ook al tot de toolbox behoort van de neurowetenschapper. Dus dan kunnen we hem toch zeker niet weren uit de neurofilosofie. Searle zegt:

For a long time most neurobiologists were reluctant to recognize the irreducibly mental, intentionalistic and conscious aspects of brain functioning. Like Habermas they thought [...] that consciousness, rationality etc. could not be construed as scientific problems. That is now very much changing. Consciousness has now become a major topic in neuroscience, and with consciousness certain other problems, such as the sense of selfhood and free action.⁴⁰

Denken in termen van intentionaliteit is binnen de filosofie niet vreemd maar dit wordt nu dus ook gedaan in de neurowetenschap. Er is dus binnen de neurowetenschappen een verschuiving gaande richting de filosofie. De neurofilosofische methode is echter nog steeds significant anders dan de neurowetenschappelijke methode. Swaab lijkt vooral de neurowetenschappelijke methode toe te passen, waarin nog een 'substantie'-scheiding sluimert zoals we die kennen van het Cartesiaanse dualisme, terwijl Damasio die scheiding definitief achter zich lijkt te hebben gelaten en de neurofilosofische benadering heeft omarmd. Meer hierover in paragraaf 3.1.

INTROSPECTIE

Een populaire manier om vrije wil, en andere geestelijke fenomenen, te bestuderen is introspectie. Daarbij wordt vaak de uniciteit van de zogenaamde qualia, de privé-ervaringen, als argument gebruikt tegen het toelaten van de neurowetenschap in het debat over o.a. bewustzijn en vrije wil. Introspectie is echter al op verschillende manieren ontmaskerd als beperkt inzetbaar in de gedragswetenschappen,⁴¹ maar nu ook in de vrije wil discussie.⁴²

Je subjectieve waarnemingen van je agentschap zijn dus onbetrouwbaar. Ze zijn echter wel illustratief. Dennett hamert er geregeld op dat we in de discussie over de vrije wil niet uit het oog

³⁹ Searle, J. R. (2007). Neuroscience, intentionality and free will: Reply to Habermas. *Philosophical Explorations*, 10(1), 71.

⁴⁰ Searle, J. R. (2007). 71-72.

⁴¹ Psychologische bewegingen, zoals het functionalisme en het behaviorisme, hebben empirische bezwaren gevonden m.b.t. introspectie. Schultz, D. P.; Schultz, S. E. (2012). *A history of modern psychology* (10th ed.). Belmont, CA: Wadsworth, Cengage Learning. pp. 67–77, 88–100.

⁴² Horgan, T., & Timmons, M. (2011). Introspection and the phenomenology of free will: Problems and prospects. *Journal of Consciousness Studies*, 18(1), 180-205.

moeten verliezen wat onze intuïtieve beleving van vrije wil is. Een theorie over vrije wil moet rekening houden met de collectieve maar ook subjectieve beleving ervan. Wat heb je aan een vrije wil concept waarin je jezelf niet herkent en dat je ook niet kan gebruiken om aan te geven dat je zelf aansprakelijk bent of dat anderen het zijn? Hiermee is echter niet gezegd dat introspectie het uitgangspunt van gedragsverklaring moet zijn, want daarvoor zijn onze subjectieve ervaringen te vluchtig. Introspectie moet gezien worden als een controle perspectief.

EVOLUTIONAIR PERSPECTIEF

Dennett en Damasio zien allebei heil in het betrekken van een evolutionair perspectief om het bestaan van de menselijke vrijheid en andere geestfenomenen inzichtelijk te maken. Het evolutionair perspectief is een noodzakelijke aanvulling op de hierboven omschreven verklaringshoudingen. De eerder omschreven perspectieven zijn bruikbaar en essentieel in het vormen van begrip, maar het evolutionaire perspectief is nodig om aan te tonen dat de perspectieven met elkaar samenhangen.

Het bestaan van een *Hacker God*, als de schepper van onze complexe werkelijkheid, is aanzienlijk minder waarschijnlijk dan schepping als gevolg van evolutie. Evolutie is praktisch niet te simuleren; niet in Conway's *Life*, laat staan in nog complexere modellen als Quine's *3D Democritisch Universum*.

Wat Conway in *Life* nog wel voor elkaar kreeg is de zelfreproductie van een Universal Turing Machine. Voor evolutie is er echter meer nodig dan zelfreproductie, namelijk: mutatie. Om een vorm van mutatie te bewerkstelligen in *Life* is volgens berekeningen een virtueel onrealiseerbare hoge resolutie vereist; de complexiteit is te groot om deze modelachtige simulatie draaiende te houden. De eenvoud van het model sluit bepaald geavanceerd gedrag uit, zo ook de creativiteit die evolutie is. Dennett noemt dit het "self-limiting equilibrium" van het model.⁴³

Ondanks dat evolutie moeilijk te simuleren valt in modellen is het wel een goed onderzocht fenomeen in de echte werkelijkheid; buiten de modellenwereld van Conway. Het aannemen van een evolutionair perspectief is behulpzaam in het verbinden van de vondsten die voortkomen uit de andere verklaringshoudingen.

Een voorbeeld. Bij een persoon die pijn ziet bij een ander, kan de amygdala, verantwoordelijk voor de angstmotie, emotiekaarten activeren en tot actie aanzetten. Het is gevonden dat het eenzelfde patroon van hersenactiviteit is als we die zien bij de andere persoon die de pijn daadwerkelijk ondergaat.⁴⁴ Wij spiegelen dus de emoties en gedragingen van anderen in onze hersenen.

De persoon die de pijn bij de ander waarneemt gaat helpen vanwege een bedoeling, zo luidt de verklaring vanuit het intentioneel perspectief. Tegelijk is het helpgedrag te verklaren door te wijzen op deze spiegelingen in hersenactiviteit: de helpende persoon komt in beweging om de eigen hersenactiviteit, die de pijn simuleert, te stoppen. Twee verschillende verklaringen voor het helpgedrag. Is *the design perspective* superieur aan *the intentional perspective*, of andersom? Allebei niet. Bij helpen is er meer gaande dan alleen het afspelen van hersenprogramma's maar het is ook

⁴³ Dennett, D. (2003). p. 50

⁴⁴ Damasio, A. (2010) p. 103

niet alleen intentionaliteit. De verklaringen vullen elkaar aan en geven een breder begrip. Door het aannemen van de evolutionaire verklaringshouding kan ontdekt worden hoe deze verschillende verklaringen zich tot elkaar verhouden. Moreel gedrag is een fenomeen dat vaak beter te verklaren valt op een intentioneel niveau (vanwege de complexiteit) maar het bestaan ervan kan worden uitgelegd door te verwijzen naar de evolutie van de hersenen. Evolutie geeft een verklaring voor het ontstaan van moreel gedrag door de overlevingsvoordelen als ook de neuroanatomische ontwikkelingen te duiden.

Om een complete verklaring te hebben voor moreel gedrag, moeten we nog een andere verklaringshouding betrekken. Dennett noemt dit de *moral stance*.

PERSONAL OR MORAL STANCE

In *Brainstorms* (1981) definieert Dennett ook een *moral* (ook wel: *personal*) *stance*. Deze verklaringshouding is relevant in het vrije wil vraagstuk, zo blijkt althans uit de slothoofdstukken van het hier behandelde boek van Dennett: *Freedom Evolves* en uit Damasio's boek: *Self Comes to Mind*. Dennett en Damasio pleiten beide voor het betrekken van een sociocultureel (moreel) perspectief bij het spreken over de menselijke vrijheid. Zij zien de mens namelijk als een biologische machine die potentieel vrij is omdat hij in staat is socioculturele instituties te creëren die hem vrij maken van bepaalde dwingende omstandigheden, vooral van zijn eigen driften. Denk bijvoorbeeld aan instituties die veiligheid waarborgen zoals de democratie en de rechtstaat, aan instituties van vooruitgang zoals de wetenschap en de media en aan instituties van welvaart zoals het monetaire stelsel en de markt. De mens kan vrijer worden doordat deze instituties hem kunnen helpen zijn door externe dwang of driften ingegeven onvrije gedrag te leren beheersen. Dennett en Damasio geloven dat vrijheid geen absoluut gegeven is maar een vloeibaar en vormbaar iets. Een mens is niet vrij of onvrij maar meer of minder vrij.

The moral stance staat enigszins los van de eerder omschreven perspectieven, die kunnen blootleggen in welke mate de mens geschikt is voor vrijheid en bijdragen aan een antwoord op de vraag of de mens in potentie vrije wil heeft. *The moral stance* geeft aan dat de mate van vrijheid per mens, per cultuur, per moment en per situatie verschillend is.

Door het morele perspectief aan te nemen, kunnen we ontdekken in welke mate menselijk gedrag van vrijheid getuigt. Dat is meteen het grote verschil met *the intentional stance*. Het gedrag van een mens kan weliswaar verklaard worden door de reden te vinden voor het gedrag (de intentie die de persoon heeft), maar daarmee is nog niet aangetoond dat dit gedrag uit vrijheid is ontstaan. Hier is *the moral stance* voor nodig. Dit perspectief kan (theoretisch) aantonen of de reden door de natuurlijke wil (de natuurlijke driften/behoefte) van de desbetreffende mens is ingegeven, of voortkomt uit door de samenleving gemaakte redenen.

Een voorbeeld. Het eten van een vette snack wordt in veel gevallen ingegeven door de natuurlijke wil. Het laten staan van die ongezonde hap is een teken van vrije wil, ingegeven door kennis uit de samenleving. Beslissen en handelen naar de kennis die je haalt uit de instituties (kennisinstellingen,

morele instellingen, politieke instellingen etc.) van de samenleving, wordt door Dennett, en ook door Damasio, gezien als beslissen en handelen uit vrije wil.

Naar mijn idee is *the moral stance* niet relevant bij het beantwoorden van de vraag of de mens vrije wil heeft; zo althans interpreteer ik de woorden van Dennett en Damasio. Het perspectief is alleen relevant bij het beantwoorden van de vraag in 'welke mate' een bepaalde mens in een bepaalde situatie vrije wil heeft. Daarom wil ik niet dieper ingaan op de houdbaarheid en bruikbaarheid van het morele perspectief. Damasio spreekt over 'socioculturele homeostase' om aan te tonen hoe cultuur onze vrijheid vergroot, maar dat is juist het meest bevraagde⁴⁵ en minst uitgewerkte deel van zijn denken over (zelf)bewustzijn en vrijheid. Bovendien is het morele perspectief het meest controversiële van Dennett's verklaringshoudingen.⁴⁶

2.3. Neurofilosofie, een mix van verklaringshoudingen

De combinatie van het design perspectief en het intentionele perspectief levert wetenschappelijk de meest waardevolle informatie op bij het verklaren van geestesfenomenen, aldus Dennett.⁴⁷ Het fysische perspectief is beperkt bruikbaar. Achteraf, na kritische vragen van andere filosofen, heeft Dennett, net als Damasio, ook de bruikbaarheid van introspectie erkent. Introspectie helpt onze 'objectief' verworven concepten te toetsen aan onze beleving. "Kloppen onze objectieve concepten met hoe we agenschap ervaren?", is een legitieme vraag volgens Dennett. Dus introspectie kan incidenteel ingezet worden als een controle perspectief.

Het evolutionair perspectief is een noodzakelijke aanvulling. Dit perspectief geeft een antwoord op de vraag waarom gedrag, beleving (subjectieve ervaring) en neuro-architectuur samenhangen op een manier die eigen is aan de mens. Damasio zegt:

Evidence from these three perspectives, even when intelligently aligned, is usually not enough to generate a smooth transition across the three kinds of phenomena— introspective, first-person inspection; external behaviors [*intentional stance*]; and brain events [*design stance*]. [...] A fourth perspective is needed, one that requires a radical change in the way the history of conscious minds is viewed and told. [mijn parenthesen]⁴⁸

De neurofilosofie zou zich dus, net als in de heterofenomenologie, moeten bedienen van het heen en weer schakelen tussen vier perspectieven: het design, intentioneel, introspectie en evolutionair perspectief. Het morele perspectief is een houding die aangenomen kan worden als nagedacht moet worden over beperkingen van vrijheid in een sociaal-culturele dimensie.

⁴⁵ Zie bijvoorbeeld: Rose, S. (2011) *Self Comes to Mind: Constructing the Conscious Brain* by Antonio Damasio – review, *The Guardian*. Retrieved from: <http://www.theguardian.com/books/2011/feb/12/self-comes-mind-damasio-review>

⁴⁶ Zie bijvoorbeeld: Haugeland, J. (1980), untitled review. in *Philosophy of Science*, Vol. 47, No. 2, pp. 326-327

⁴⁷ Dennett, D. (1981), *Brainstorms*. Cambridge, MA: MIT Press.

⁴⁸ Damasio. A. (2010). p. 15

TOEPASSING EN KRITIEK OP DE HETEROFENOMENOLOGIE EN DE NEUROFILOSOFIE

Het betrekken van de heterofenomenologie als centraal instrument in de neurofilosofie is controversieel. Sterker nog, de concreetheid van de heterofenomenologie zelf is voor sommige filosofen al controversieel. Dat geestfenomenen op verschillende begripsniveau's bestudeerd moeten worden is dat niet. En dat die begripsniveau's ook tot elkaar verhoudt moeten worden en niet los van elkaar gezien moeten worden is voor neurofilosofie pioniers als Patricia Churchland evident.⁴⁹ Het is belangrijk om te realiseren dat Patricia Churchland meer een reductionistisch perspectief heeft op wat neurofilosofie moet zijn. Zij stelt voor dat in de toekomst er niet langer gesproken zal worden in termen van behoeftes en beslissingen (intenties) maar in termen van hersenanatomie en hersenactiviteiten. De haalbaarheid van dit project is gering.⁵⁰ Dat is ook gebleken hierboven uit mijn behandeling van Dennett's heterofenomenologische methode; *the design stance* heeft onvoldoende verklarende kracht als het om complex gedrag gaat.

Hoe Dennett en Damasio met de neurofilosofie te werk gaan en hoe Patricia en Paul Churchland dat doen is in mijn optiek voor deze scriptie een verwaarloosbaar verschil. Het zijn in de kern beide materialistische posities. Churchland's project is een in mijn optiek meer een theoretische ambitie en Dennett's project een praktische ambitie. Over de voor en nadelen van de twee methoden wil ik hier niet in detail schrijven. Zeer wel mogelijk dat bij het ene probleem de ene methode meer oplevert en bij een ander probleem de andere methode superieur is. Het is voor mij genoeg om te weten dat Damasio zich bediend van de heterofenomenologische methode in zijn neurofilosofisch onderzoek.

Dennett roept op om naar onze (neuro)biologie te kijken. Hij verwacht hierin een structuur te vinden die onze vrijheid mogelijk maakt. En wat Dennett verkondigt wordt door Antonio Damasio daadwerkelijk toegepast. Damasio onderzoekt in eerste instantie het zenuwstelsel op microscopisch niveau. Maar hoe gedetailleerd hij dit ook bestudeerd het brengt hem er niet toe de hersenen te bespreken in termen van natuurprincipes. Damasio neemt zelden of niet de zogenaamde *physical stance* aan. Alleen de drang tot overleven die in de genen van alle levende organismen is geprogrammeerd, ziet Damasio als een natuurprincipe. Hij is er kort en bondig over; hij weet niet waar dit principe oorspronkelijk vandaan komt, maar voor het bespreken van gedrag vindt hij het voldoende gedetailleerd om naar de genen te kijken.

Los hiervan bespreekt Damasio zaken als het menselijk bewustzijn, agentschap en dergelijke afwisselend vanuit de *design stance*, de *intentional stance*, het evolutionair perspectief en incidenteel controleert hij zijn theorieën door middel van introspectie. Geen wonder dat Dennett bij Damasio eenzelfde manier van verklaren en begrijpen van geestvraagstukken herkent als die hij zelf praktiseert.⁵¹ Mijn conclusie is dat beide gebruik maken van de heterofenomenologische methode. Ook vind ik dat Dennett het aannemelijk heeft gemaakt dat determinisme het bestaan van de vrije

⁴⁹ Churchland, P. S. (1989). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT press. pp. 383-384

⁵⁰ Schaffner, K. F. (2009). The person and philosophy of science and medicine. *International journal of integrated care*, 10(Suppl), e017-e017.

⁵¹ Dennett, D. (2005). *Sweet dreams: Philosophical obstacles to a science of consciousness*. MIT press. pp. 42-44

wil niet tegenspreekt en dat de neurofilosofie een bruikbare bijdrage kan leveren aan het debat over de vrije wil.

Ik zie echter ook problemen. Dennett is van mening dat metafysica geen plek heeft in het vrije wil debat. Maar is niet juist de metafysica een belangrijk onderdeel van Dennett's verdediging van de neurofilosofische methode? Dennett schuift de *physical stance* te gemakkelijk opzij. Dat doet hij door kritiek te leveren op het verkeerd gebruik van deze verklaringshouding, terwijl hij zelf laat zien dat je middels de *physical stance* ook een constructieve bijdrage aan het vrije wil debat kan leveren. Maar deze bijdrage van Dennett is dan metafysisch van aard en niet fysisch. De metafysica blijft belangrijk in het discussiëren over de definities van (en spanningen tussen) (in)determinisme en vrijheid. Zelfs het definiëren van vrije wil kan in een metafysische reflectie gerealiseerd worden.

Dat brengt meteen een volgend probleem naar boven. Het is soms problematisch hoe Dennett de ene keer over metafysica en dan weer over fysica spreekt. Het onderscheidt wordt onvoldoende gemaakt in Dennett's boek: *Freedom Evolves*. Hij beschrijft de fysische verklaringshouding als beperkt als het gaat om denken over de vrije wil. Populaire metafysische gedachte-experimenten doet hij van de hand als ruis maar gebruikt zelf ook metafysische argumenten voor zijn compatibilistische positie. Hij spreekt dan bijvoorbeeld over causaliteit en noodzakelijkheid. Ondanks dat Dennett blijft hameren op een pragmatische basis voor onze definitie van vrije wil laat hij onvoldoende zien wat de plek van de metafysica is en/of deze net als de *physical stance* uitgesloten moet worden.

Toch maak ik in het vervolg van deze scriptie geen gebruik meer van *the physical stance*. Ik ben er namelijk wel van overtuigd dat alleen middels de neurofilosofie en de heterofenomenologische methode (als een combinatie van verklaringshoudingen zoals opgesomd in paragraaf 2.3.) ontrafeld kan worden of de mens (als geavanceerde complexe creatie) voldoet aan de traditionele voorwaarden voor vrije wil zoals ook door Swaab gehanteerd. Ik wil bovendien vooral kijken naar de bruikbaarheid van de neurofilosofische argumenten van Damasio met betrekking tot de vrije wil.

De verschillende perspectieven (*stances*) lijken soms moeilijk van elkaar te onderscheiden. Kan 'het groeten van de buurman' het beste verklaard worden vanuit het intentionele perspectief of vanuit het morele perspectief? Of toch vanuit het design perspectief? Alle drie zijn toepasbaar. De verschillende perspectieven zullen (hopelijk) verschillende aspecten van hetzelfde fenomeen blootleggen en zo een meer compleet begrip opleveren.

Dennett suggereert dat er een hiërarchie of continue schaal bestaat bij de verschillende verklaringshoudingen. Zijn voorbeelden van *the Game of Life* geven aan dat als een creatie complexer is, we een 'hogere' verklaringshouding moeten aannemen, bijvoorbeeld overstappen van het design perspectief naar het intentionele perspectief. Zo zullen we complexer gedrag 'beter' begrijpen. Dit vind ik minder sterk. Ik ben er namelijk van overtuigd dat een 'lagere' verklaringshouding wel degelijk een bruikbaar inzicht kan verschaffen in het begrijpen van dat gedrag; bijvoorbeeld *the design stance* in het geval van moreel gedrag. Zo is ook het betrekken van fysische informatie een bijdrage aan de begripsvorming van geestfenomenen. Daarom vind ik Dennett's kritiek op *the physical stance* ietwat overtrokken. Wel ben ik het met Dennett eens dat de bruikbaarheid beperkt is van deze zogenaamde 'lagere' verklaringshoudingen als het te verklaren gedrag complex is.

Daarbij komt dat de manier waarop Dennett gebruik maakt van het woord 'intentionaliteit' controversieel is. Volgens Dennett is "*rationaliteit* de moeder van intentie".⁵² Hij omschrijft rationaliteit als de "...optimal design relative to a goal and a set of constraints ."⁵³ Als je een hap neemt uit een appel, dan is dat een voorbeeld van rationeel gedrag: optimal design. Je hebt het doel je honger te stillen en doet dat op een manier die rekening houdt met de beperking van de hoeveelheid mogelijkheden waarop je dat gerealiseerd kan krijgen (je kunt je arm maar één kant opbuigen, je krijgt een appel niet binnen door 'm tegen je buik te drukken, je moet de appel bij je mond zien te krijgen etc.).

In een recensie van het boek *Brainstorms* bekritiseert de filosoof John Haugeland Dennett's definitie van rationaliteit. Rationaliteit wordt in zijn optiek gereduceerd tot een muizenval die een optimale bouw kent (design) om een muis te vangen (het doel).⁵⁴ Haugeland realiseert zich niet dat dit precies Dennett's bedoeling is. Dennett verwoordt dit zelfs nog sterker:

Yes, trees can "decide" that spring has come and it is time to push out the blossoms, and clams can "decide" to clam up tight when they feel an alarming bump on their shells, but these options are so rudimentary, so close to being simple switches, that they are decisions by courtesy only. But even a simple switch, turned on and off by some environmental change, marks a *degree of freedom*.⁵⁵

Hersenen zijn in feite een netwerk van verbonden 'aan en uit' -knoppen (switches) die via input en output, en kennis van het verwerkte verleden, voorspellingen doen over de toekomst.

Voorspellingen om naar te handelen.

Rationaliteit is een vermogen dat voortkomt uit een optimale machinerie. Het gedrag van de muizenval is te verklaren vanuit een intentioneel perspectief. Haugeland voelt hier weinig voor – en niet ten onrechte -, maar hij maakt een fout door te suggereren dat het *niet* kan. Het gedrag van de muizenval kan weldegelijk in termen van intentionaliteit omschreven worden, ongeveer als volgt:

- de muizenval wil een muis doden en zijn reden is de beroering van zijn bewegingssensor.

Dit klinkt absurd, maar het is geen foute verklaring; hij is alleen niet praktisch. De muizenval heeft immers een buitengewoon beperkte wil; hij heeft niet veel meer te willen dan een muis vangen. Daarom geniet de *design stance* de voorkeur als gedragsverklarend model. Het geeft ons betere inzichten in de mechanismen van de val en dat ervaren we als handiger. Toch is hiermee niet gezegd dat het intentionele verklaringmodel alleen kan gaan over de hogere orde levende wezens. Het enige wat is aangetoond, is dat in deze situatie het design perspectief de voorkeur geniet. Het fysische perspectief zou hier evenmin verkeerd zijn, alleen maar onbruikbaar vanwege de complexiteit op macroniveau.

⁵² Dennett, D.C. (1981. Eerste druk was in 1978). *Brainstorms: philosophical essays on mind and psychology*. MIT Press. p. 9

⁵³ Dennett, D.C. (1981). p. 5

⁵⁴ Haugeland, J. (1980), untitled review. in *Philosophy of Science*, Vol. 47, No. 2, pp. 326-327

⁵⁵ Dennett, D.C. (2003) p. 162

Besef van de verschillende verklaringshoudingen opent de deur om vanuit de neurofilosofie te kijken naar het vrije wil vraagstuk. Het 'echte werk' moet nog beginnen. Nu de neurofilosofie gerechtvaardigd is, rijst de vraag of dit discipline ook inzichten oplevert die bruikbaar zijn voor het aantonen van de vrije wil bij de mens? Heeft de mens de juiste capaciteiten voor vrije wil? Als we de Peter van Inwagen's metafysica links laten liggen en het genotype en fenotype van de mens als uitgangspunten nemen, kunnen we dan zeggen dat menselijke vrijheid bestaat? Of zijn de argumenten van Swaab, Lamme, Wegner en anderen afdoende om de vrije wil als een illusie af te doen? Op deze vragen ga ik in het volgende hoofdstuk verder in.

3. Neurofilosofische onderbouwing van vrije wil

In het vorige hoofdstuk zagen we hoe de neurofilosofische benadering gerechtvaardigd werd. Door gebruik te maken van de heterofenomenologische methode, een combinatie van *the design stance*, *intentional stance* en het *evolutionair perspectief*, kunnen we onze ideeën over geestvraagstukken verdiepen. Daar lijken Dennett en Damasio het over eens. Beiden komen ze tot de conclusie dat de vrije wil bestaat. Het zijn beiden compatibilisten, maar zij putten niet geheel uit dezelfde bronnen voor hun argument. Dennett richt zich, in het boek *Freedom Evolves*, voornamelijk op het uitleggen dat metafysisch determinisme compatibel is met een vrije wil. Terwijl Damasio, in het boek *Self Comes to Mind*, voornamelijk kijkt naar wat hersenonderzoek ons kan leren over bewustzijn, het zelfproces, agentschap en menselijke vrijheid.

In dit hoofdstuk zullen we ons concentreren op de toepassing van de neurofilosofie zoals we dat zien in het werk van Antonio Damasio.

Voor Damasio is de vrije wil (nog) geen centraal thema geweest in zijn diverse boeken, maar hij bespreekt mijns inziens wel de belangrijkste problemen die relevant zijn in de vrije wil discussie. Damasio richt zich met name op de vraagstukken rondom (zelf)bewustzijn, agentschap en zelfcontrole. Ik neem in dit hoofdstuk de vrijheid om zijn argumenten te lenen voor een neurofilosofische onderbouwing van de vrije wil.

In zijn argumenten zijn oplossingen te vinden voor drie klassieke voorwaarden van vrije wil die ook Swaab centraal stelt. Om te kunnen zeggen dat een handeling uit vrije wil is gedaan, moet voldaan worden aan de volgende criteria:

- (1) een handeling is pas vrij als je hem ook niet had kunnen doen (de Libertijnse voorwaarde van *alternative possibilities*),
- (2) de handeling moet met een reden zijn verricht, en
- (3) de handeling moet echt uit jezelf komen.

In welke mate zien we deze criteria voldaan worden door neurofilosofische argumenten van Damasio? En in welke mate zijn deze argumenten houdbaar? In het komende hoofdstuk gaan we hiernaar kijken. We beginnen met voorwaarde 3 en werken dan terug naar voorwaarde 1.

3.1. De handeling moet echt uit jezelf komen

DE MEREOLOGISCHE DROGREDEN

Door *the physical stance* achter ons te laten, hebben we nog niet volledig het gevaar van determinisme verholpen. Op het begripsniveau van *the design* en *intentional stance* ligt biologisch en omgevingsdeterminisme op de loer. Dennett stelt dat je als mens ook deze vormen van

determinisme dankzij de wil, cultuur en educatie kan leren herkennen en ontwijken. Ontwijken is denkbaar, zelfs in een gedetermineerde wereld, zo hebben we in het voorgaande hoofdstuk geleerd.

Als samenleving kunnen wij invloed uitoefenen op de omgeving die ons vormt en binnenkort zelfs ook op de genen. In andere woorden: onbeïnvloedbaar zijn die genen van Swaab niet. Je moet niet de vraag stellen of determinisme bestaat, maar wat je allemaal kan veranderen in die gedetermineerde wereld. Dat is de relevante vraag met betrekking tot de vrije wil discussie, aldus Dennett.⁵⁶ Over genetisch determinisme en omgevingsdeterminisme is Dennett kort: niemand gelooft slechts in het één of het ander. De werkelijkheid is een combinatie van beide aangevuld met (pseudo-) willekeur. Determinisme is niets om bang voor te zijn, maar iets om te omarmen.

Knowledge of the roles of our genes, and the genes of the other species around us, is not an enemy of human freedom, but one of its best friends. [...] Culture provides the vantage point from which we can see how to change the trajectories into the future that have been laid down by the blind explorations of our genes.⁵⁷

Dick Swaab verklaart zich een 'neurocalvinist'⁵⁸, omdat het fenotype van een persoon, het uiterlijke voorkomen, voor het grootste gedeelte door genen en vroege jaren gevormd wordt. Je bent je genen, zegt Swaab. Het kleine beetje dat vormbaar is, wordt in je vroege jaren gerealiseerd (je taal, je levensbeschouwelijke positie, etc.). Je bent dan nog te jong om zelf invloed uit te oefenen op wie je wordt. Hij onderbouwt deze stelling door te verwijzen naar verschillende onderzoeken in de genetica en ontwikkelingspsychologie. Een sterk pleidooi voor genetisch determinisme.

Dick Swaab citeert Joseph Price: "[De vrije wil is] de mogelijkheid om te besluiten iets wel of niet te doen zonder interne of externe beperkingen die deze keuze bepalen."

Swaab suggereert dat ruimte voor wilskracht verwaarloosbaar klein is bij een volwassen persoon. Is dat terecht? Hierover is een wetenschappelijke strijd te leveren door allerlei onderzoeken naast elkaar te leggen die deze ruimte van wilskracht groter of kleiner maken. In mijn optiek is het vooral belangrijker te kijken naar het citaat van Price. Het meest opvallende probleem in het Joseph Price-citaat is aan te duiden door de volgende vraag te stellen: "waardoor zou je nog meer beïnvloed kunnen worden dan door interne en externe zaken?" Slechts Cartesiaans dualisme lijkt dan nog uitkomst te bieden. Een geest die losstaat van het materiële lichaam/breïn. Interne en externe factoren zijn immers alles.

Ik neem de vrijheid Swaab's uitspraak te interpreteren. Bij externe zaken kan ik me iets voorstellen, dat kunnen weersomstandigheden zijn, maar ook sociale en politieke krachten. Bij interne beperkingen neem ik bij deze aan dat Swaab hiermee *onbewuste* geest- en lichaamsprocessen (die bepaald zijn door de genen of opvoeding) bedoelt die ons handelen sturen. Zoals Libet in zijn onderzoek naar de *readiness potential* zou hebben laten zien. Het lijkt erop neer te komen dat als ons gedrag door buitenliggende factoren (zoals het hamertje op je knie van de dokter) of door onze

⁵⁶ Dennett, D. (2003) p. 160

⁵⁷ Dennett, D. (2003) pp. 162-165

⁵⁸ Dit begrip verwijst naar de predestinatieleer die het leven van calvinisten bepaalde

hersenen bepaald wordt, dat er dan sprake is van onvrijwilligheid (met hooguit de toevoeging van een illusie van vrijheid). Daarmee zegt Swaab in feite dat vrije wil alleen kan bestaan als wij niet onze hersenen zijn.

Swaab maakt zich hiermee schuldig aan het denken in termen van lichaam versus geest. Als Swaab het Cartesiaans substantiedualisme verwerpt, dan wordt het ook eens tijd dat hij, op z'n Amerikaans *puts his money where his mouth is*. In andere woorden: hij moet ons lichaam - en daarmee ook onze hersenen - gaan zien als hetgeen wij zijn. Als ons lichaam, onze hersenen ons iets laten doen, kun je evengoed zeggen dat ik mezelf iets laat doen. Naast je genen en je omgeving is er ook nog een derde belangrijke bron van invloed op je gedrag: de invloed die je zelf op jezelf hebt. In andere woorden: de invloed die je eigen hersenen op je eigen hersenen hebben. Het gebruik van het Joseph Price-citaat laat zien hoe problematisch en sturend de ideeën en conclusies van wetenschappers als Swaab zijn.

Swaab is niet de enige die zich schuldig maakt aan Cartesiaans denken. Sinds Descartes heeft het dualisme het denken over bewustzijn gedomineerd. Ondanks dat Cartesiaans substantiedualisme niet meer wordt onderschreven, wordt er nog steeds in termen van lichaam-geest gesproken, zelfs door kritische hedendaagse neurobiologen. Het is een hardnekkige neiging om het lichaam als iets anders te zien dan de geest. Wij zijn niet onze hersenen, zoals Swaab zegt. Dat is een mereologische drogreden. De hersenen zijn wel een belangrijk onderdeel van wie we zijn, maar onze hersenen - zonder ons lichaam - zijn wij niet.

LICHAAM EN BREIN

De eerste stap die Damasio zet in het boek *Self Comes to Mind* is het aantonen hoe vergaand de verwevenheid is tussen het lichaam en het brein. Damasio wijst ons erop dat het al langer bekend is dat het brein beelden maakt van het lichaam, maar daarbij wordt vaak over het hoofd gezien dat het brein ook beelden maakt van het brein dat beelden maakt van het lichaam. Zo beïnvloedt het lichaam de hersenen. Onze cognitie wordt van 'waarde' voorzien dankzij het lichaam. Tegelijk is het zo dat het brein ook de staat van het lichaam beïnvloedt. Er is sprake van communicatie over en weer; van tweerichtingsverkeer.

Ondanks hun zelfverklaarde monisme lijken neurowetenschappers en filosofen als Swaab nog steeds te neigen naar denken in termen van eenrichtingsverkeer. Of het lichaam vertroebelt de geest (de hersenen) óf de geest (de hersenen) stuurt het lichaam. Het denken in eenrichtingsverkeer voedt de misleidende intuïtie dat de geest iets is dat los is van het lichaam. Er wordt weliswaar niet meer gesproken over de geest als substantiële entiteit, maar door sommigen wordt er wel gespeculeerd over een brein dat speciale gebieden (met mysterieuze krachten) heeft die vrije wil en zelfbewustzijn mogelijk maken. In plaats van Swaab erop te wijzen waar het op staat (namelijk: wij zijn meer dan ons brein en alles wat ons lichaam bepaald wordt door ons bepaald), gaan filosofen⁵⁹ op zoek naar geheimzinnige gebieden in ons brein waar 'het-ik-dat-bepaalt' verstopt zou zitten. Dennett noemt dat mysterieuze gebied gekscherend: *the Cartesian Theatre*.⁶⁰

⁵⁹ Zie o.a. Kane, R. (1998). *The Significance of Free Will*. Oxford University Press.

⁶⁰ Dennett, D. (2003). p. 123

Damasio is ook streng voor zijn collega's. Neuropsychologie moet zich niet concentreren op het lokaliseren van functies in de hersenen, maar op het ontdekken van de relaties tussen componenten van de hersenen en hun cognitieve equivalent. Damasio zegt:

The goal of neuropsychological enterprise is thus to explain how certain cognitive operations and their components relate to neural systems and their components. Neuropsychology is not, or should not be, about finding the brain "localization" for a "symptom" or "syndrome".⁶¹

Volgens Damasio is de voornaamste taak van het brein het beheer van *biological value*; oftewel, het beheer van leven. In simpele hersenen bestaat er een directe link tussen subcorticale nuclei en het lichaam. Bij complexe breinen, met een ontwikkelde hersenschors, is er sprake van expliciete beelden van de staat van het lichaam. Een ingewikkeld brein verzamelt externe beelden en legt deze naast beelden van het lichaam om uiteindelijk tot een beslissing/handeling te komen. Let wel, dat gebeurt niet in een chronologische lineaire structuur, maar met constante interactie tussen het lichaam en allerlei hersengebieden en weer met de zintuigen. De reden dat er een intensieve verbinding bestaat tussen het brein en het lichaam kan Damasio niet vaak genoeg herhalen:

The business of managing life consists of managing a body. [...] Neurons are about life and about managing life in other cells of the body, and that *aboutness* requires two-way signaling.^{62 63}

HET ZELFPROCES

In Damasio's theorie over de menselijke geest staat het zelfproces centraal. Door naar de details van Damasio's *zelfproces* te kijken kunnen we zien hoe verweven het brein is met het lichaam en hoe dat zijn impact heeft op ons zelfbesef en daarmee op ons bewustzijn. Ik zal hier het zelfproces van Damasio samenvatten en dit vervolgens relateren aan de derde vrije wil voorwaarde van Swaab: een handeling moet *echt* uit jezelf komen.

Als Swaab spreekt over handelingen die *echt* uit jezelf moeten komen, bedoelt hij zeer waarschijnlijk handelingen die ter plekke *bewust* worden besloten, zonder inmenging van genetische disposities of (in de vroege jeugd) aangeleerde disposities. Handelingen waarbij het gevoel van eigenaarschap/agentschap terecht is en niet slechts een illusie.

Het fundamentele probleem van deze bijna Cartesiaanse benadering van het vrije wil vraagstuk laat zien dat hij over het hoofd ziet hoe ingenieus en bovenal non-lineair de aard van de subjectieve

⁶¹ Damasio, A. (1994) p. 53

⁶² Damasio, A. (2010). *Self comes to mind: constructing the conscious brain*. Pantheon Books. p. 90

⁶³ Let op: bij Damasio is *aboutness* niet gelijk aan bewustzijn. *Aboutness* vergelijkt hij met *the intentional stance*, zoals we die ook bij Dennett kennen. Op celniveau kan er al gesproken worden over intenties maar een brein als ook de cellen hebben geen *bewuste* intenties. Het waarneembare gedrag is alsof ze die wel hebben. "[...] neither a whole brain nor single cells deliberately intend anything with their behavior, but their stance is as if they do." Het is niet fout om te zeggen dat een cel iets wil net als de mens iets wil. De mens is er echter bewust van en heeft bovendien het luxeprobleem dat die veel meer dingen wil dan de cel. Fout is die vergelijking niet als je cartesiaans dualisme werkelijk links laat liggen.

ervaring is georganiseerd in de menselijke biologie. De subjectieve ervaring van 'zelf', met het daarbij horende gevoel (*feeling*) en rationeel besef (*sense*) van eigenaarschap van de eigen handeling/ beslissing, is een dynamisch fenomeen.

In het artikel *Where am I?* (1981) vraagt Dennett hardop: ben ik mijn brein? Kunnen we dat bevestigen na dertig jaar hersenonderzoek? Swaab suggereert van wel, zijn boek heet immers *Wij Zijn Ons Brein*. Swaab's vakgenoot Damasio schetst een ander beeld. Hierboven was al te lezen dat Damasio een sleutelrol ziet weggelegd voor het lichaam in het zelfproces. Een brein scheiden van een lichaam resulteert dus ook in een ernstige verstoring van het zelfproces.

Het zelf-fenomeen is vanuit verschillende perspectieven te bestuderen. Het zelf is een fenomeen dat bestudeerd kan worden vanuit een perspectief van introspectie of door een *second-person perspective*: door psychologen en neurowetenschappers. Het is een moeilijk te vatten concept: "*The self is indeed a movable feast.*"⁶⁴

Ten eerste, 'zelfheid' is een proces en niet een ding. Een proces dat altijd sluimerend aanwezig is, maar verschilt in de mate waarin je je ervan bewust bent: "*Now you sense it, now you don't, but you always feel it.*"⁶⁵

Damasio introduceert de term kaart ("map") als synoniem voor de 'zelf'-reprenterende patronen in het zenuwstelsel. Belangrijk om te weten is dat mentale beelden ("images") corresponderen met neurale patronen. Deze beelden kunnen van visuele aard zijn, maar deze beelden kunnen ook auditief, instinctief, tactiel, herinnerd of gefantaseerd zijn.

De hersenschors heeft een plattegrondachtige anatomie, perfect om objecten en acties te representeren. De hersenschors kan dat beter dan de lagere (oudere) hersengebieden, al kunnen die dat ook, weliswaar in een lagere "resolutie".⁶⁶ De kaarten van de hersenen bestaan uit tijdelijk actieve gebieden in combinatie met inactieve gebieden. Als bij een elektrisch reclamescherm zijn mentale beelden opgebouwd door actieve en inactieve 'lampjes'.

*"The human brain is a born cartographer, and the cartography began with the mapping of the body inside which the brain sits."*⁶⁷

Zelfprocessen, eveneens neurale patronen, zijn nodig om die beelden bewust te maken ("*to make them known*"). Gevoel en emotie hebben in deze zelfprocessen een sleutelrol. Mentale beelden kunnen bestaan in de hersenen zonder dat die bewust ervaren worden. Deze mentale beelden representeren fenomenen en evenementen buiten het brein. Deze evenementen hoeven niet buiten het lichaam te zijn. Sommige neurale patronen representeren zelfs andere patronen in het brein. Bepaalde combinaties van neurale activiteit resulteren in een geest met gevoelens ("*mind with feelings*"). Deze geest met gevoelens is het resultaat van een soort opschalen van cognitie en gevoel. Damasio vergelijkt het met lichaamsbeweging.

⁶⁴ Damasio, A. (2010) p. 171

⁶⁵ Damasio, A. (2010) p. 9

⁶⁶ Damasio, A. (2010) p. 66

⁶⁷ Damasio, A. (2010) p. 64

The contraction of a single microscopic muscle cell is a negligible phenomenon, whereas the simultaneous contraction of large numbers of muscle cells can produce a visible movement.⁶⁸

HET 3-STAPSGEWIJS OPGEBOUWDE ZELFPROCES VAN DAMASIO

Damasio omschrijft het zelfproces als iets dat in drie stappen wordt opgebouwd in de hersenen: 1) *the protoself*, 2) *the coreself* en 3) *the autobiographical self*. Elke volgende stap draagt onderdelen van het eerdere zelfproces met zich mee. Vanuit evolutionair perspectief is het eerste proces het oudst (dit proces komt dus ook voor bij lagere diersoorten) en het *autobiographical self*, ook wel “the social or spiritual me”, is de jongste zoon van modernatuur. De drie processen corresponderen met bepaalde functies van de hersenen.

Beelden die in je hersenen worden gevormd in combinatie met wakkerheid zijn uiteraard basisingrediënten van bewustzijn. De mate van wakkerheid die de hersenstam reguleert, is een antwoord op de vraag: hoe belangrijk zijn deze omstandigheden en objecten voor het eigen organisme? De waarde die (onbewust) wordt toegekend op deze manier bepaalt wat voor emoties we voelen, maar ook hoe alert we zijn; hoeveel aandacht we besteden aan die omstandigheden en objecten.

De bovenste hersenstam is essentieel voor het maken van mentale beelden van externe stimuli en van de staat waarin het lichaam verkeert. Deze beelden corresponderen met het protoself. De protoself (hersenen)structuren zijn verbonden met het lichaam, waardoor er constant interactie plaatsvindt. Het belangrijkste product van het protoself zijn oergevoelens (*primordial feelings*) zoals pijn en geluk. Damasio ziet deze oergevoelens beginnen met emoties en ervaren worden als gevoelens. Dit is voor Damasio een belangrijk onderscheid. Emoties zijn kaarten, subjectief ervaren als gevoelens, van het lichaam.

Het protoself geeft ons dankzij die kaarten het enkelvoudige gevoel dat wij één zijn en altijd geweest zijn. Het gevoel van 'ik ben één' (altijd geweest) is goed te vinden in het interne milieu van je lichaam. Damasio ziet daarvoor bewijzen in het feit dat je chemische interne milieu niet veel verandert door je leven heen. Zo veranderen de parameters van het chemische bad waar leven uit bestaat nauwelijks, of je drie jaar of tachtig jaar oud bent.⁶⁹ Zo ook de biologische essentie van emotieactieprogramma's in je lijf. De hersenstam is bovendien het hersengebied dat het minst ontwikkelt tijdens een leven.

De hersengebieden die kaarten van de externe wereld maken, zijn van een andere orde dan de hersenstam. Dit zijn de hoger orde hersendelen. De gebieden die kaarten van de buitenwereld maken, hebben de flexibiliteit om de meest uiteenlopende waarnemingen te kunnen verwerken, een soort tabula rasa. De hersenstam die kaarten van de interne wereld maakt, werkt eerder als een soort kleurplaat; er is dus bij elke waarneming van de toestand van het eigen organisme een

⁶⁸ Damasio, A. (2010) p. 20

⁶⁹ Damasio, A. (2010) p. 194

overeenkomst met de vorige waarneming. Damasio spreekt hier graag over “the infinite sameness of the narrow range of patterns related to the body’s interior and its tightly controlled regulation.”⁷⁰

Het begrijpen van dit verschil tussen lager orde en hoger orde hersengebieden betrokken bij (zelf)bewustzijn is essentieel in het begrijpen van het zelfproces.

Wat heb je nodig voor een interne protagonist? Wat zijn de voorwaarden voor het bestaan van subjectiviteit. Daar is meer voor nodig dan de ervaring van het eigen organisme, zoals het protoself dat realiseert. Er moet sprake zijn van 'zelf'-zichtbaarheid op een hoger kennisniveau (dus niet alleen een voelend-beseffen) en er moet een verbinding zijn met de omstandigheden waar het zelf in verzeild is. “It must *protagonize*.”⁷¹ Dat doet het zogenaamde *core self*, of kernzelf.

Het kernzelf komt gedeeltelijk voort uit het protoself (het is een soort opschaling van signalen), maar het is een proces dat ook andere hersengebieden aandoet.

Als het protoself een verandering ondergaat (stel het hormonale milieu verandert bij een interactie met een object), dan verbindt het kernzelf deze interne veranderingen aan een 'extern' object⁷². Het externe object heeft nu een emotionele lading gekregen en is naar voren gebracht in het aandachtsveld. Dit gebeurt in een sequentie van gebeurtenissen. Een object (actueel, herinnerd of gefantaseerd) komt in aanraking met het lichaam door gezien, gehoord of betast te worden vanuit een bepaald perspectief. Deze aanvaring zorgt voor veranderingen in het lichaam. De aanwezigheid van het object wordt gevoeld. Het object wordt saillant, in andere woorden: het krijgt een biologische waarde toegekend. Dat is evolutionair verklaarbaar: “after all, it is the object that matters most in order for the process to be adaptive.”⁷³

Het voortduren van een dergelijk non-verbaal narratief wijst de geest op een centraal element in het verhaal: de protagonist. Damasio noemt dit de *material me*. Eveneens worden de acties die het organisme produceert gekoppeld aan de protagonist, in één lijn met het gevoel dat het object oproept. De object-organismere relatie en handeling wordt als eigendom ervaren. Er is “a sense of ownership.”⁷⁴ De geest wordt aangevuld met een serie beelden die bewustzijn creëren: een beeld van het gemodificeerde protoself; een beeld van een objectgerelateerde emotionele reactie (een gevoel) en een beeld van het oorzakelijk object. “*The self comes to mind in the form of images, relentlessly telling a story of such engagements.*”⁷⁵

Dit is een belangrijk proces en een voorwaarde voor het ontstaan van het autobiografisch zelf. Damasio begrijpt de hang naar taal en sociale context (zoals Michael Gazzaniga⁷⁶ en anderen dat doen) in het verklaren van bewustzijn, maar pleit dat een autobiografisch zelf, noem het persoonlijkheid, niet mogelijk is zonder het kernzelf proces.

⁷⁰ Damasio, A. (2010) p. 199

⁷¹ Damasio, A. (2010) p. 202

⁷² Let op: een extern object kan ook een herinnerd of gefantaseerd object zijn.

⁷³ Damasio, A. (2010) p. 204

⁷⁴ Damasio, A. (2010) p. 203

⁷⁵ Damasio, A. (2010) p. 203

⁷⁶ Michael Gazzaniga, *The Mind's Past* (Berkeley: University of California Press, 1998).

Damasio onderbouwt het kernzelf proces door een aantal hersengebieden aan te wijzen die betrokken zouden kunnen zijn. Kandidaat voor het coördineren van gevoelens uit de hersenstam en beelden van externe objecten zijn de superior colliculi. Deze kunnen gegevens van diverse zintuigen verwerken, waardoor een synthese van de verschillende beeldenstromen mogelijk wordt. Een andere kandidaat is de thalamus, in het specifiek de associatieve nuclei van de thalamus, die perfect gesitueerd is om functionele verbindingen te maken tussen gescheiden groepen van corticale activiteit.

Het autobiografisch zelf wordt, door Damasio, neergezet als een tweeledig fenomeen. Het autobiografisch zelf kan plaatsvinden in volle bewustzijn waarbij je persoonlijke geschiedenis en fantasieën tot uitdrukking komen. "Autobiographical selves are autobiographies made conscious."⁷⁷ Tegelijk wordt het offline, buiten het bewustzijn, gevormd in het verloop van groeiende levenservaring. In dit tweede proces worden herinneringen gemaakt en/of herzien, meestal buiten het bewustzijn om.

Het autobiografische zelf wordt in de hersenen als volgt opgebouwd. Ten eerste moeten biografische gegevens zo opgeslagen worden dat ze associatief gegroepeerd zijn, maar tegelijk als losse objecten opgeroepen kunnen worden. Elk object moet het protoself kunnen modifieren om zo een *pulse of core self* mogelijk te maken, met de daarbij horende gevoelens van weten. Het object moet saillant worden gemaakt. Ten tweede moet het brein apparaten hebben die, in de complexiteit van geheugenopslag, de vereiste herinneringen met het protoself doen interacteren, teneinde die interactie oorzakelijk in verband te brengen met het juiste object.

Bij complexe autobiografische zelfprocessen (bijvoorbeeld als het subject nadenkt over morele vraagstukken) zullen meerdere objecten behandeld worden. Er zullen dan ook meerdere uitingen van het kernzelf (*pulses of core self*) moeten plaatsvinden in een bepaalde sequentie. Daar zijn coördinerende gebieden bij nodig.

Bij wijze van werkhypothese stelt Damasio voor dat het autobiografische geheugen uit twee samengevoegde mechanismen bestaat. Het eerste mechanisme is schatplichtig aan het core self-mechanisme. Een set van autobiografische herinneringen wordt bewust gemaakt als object in een *core self pulse*. Het tweede mechanisme realiseert een hersenwijde coördinatie van (1) geheugeninhoud die naar boven komt in de vorm van beeld, (2) dit beeld wordt gepositioneerd in interactieve modus met het protoself-systeem, en (3) de resultaten van deze interactie worden gedurende een bepaalde tijd vastgehouden in een staat van samenhang. Er zijn bepaalde systemen aan te wijzen die een dergelijke coördinerende rol spelen.

Damasio wil voorop stellen dat deze coördinerende apparaten in de hersenen geen Cartesiaanse Theaters zijn. Het zijn processen die geen mysterieus agentschap kennen, maar te werk gaan volgens spontane natuurlijke principes. Voorbeelden van principes zijn: de volgorde van binnenkomst telt en de toegekende biologische waardes tellen.

Waardering van waargenomen, herinnerde of gefantaseerde objecten komt tot stand dankzij genetische disposities en aangeleerde disposities. Waardes worden bij initiële perceptie toegekend

⁷⁷ Damasio, A. (2010) p. 210

en opgeslagen. Eventueel worden die waardes vervormd elke keer dat er sprake is van het terughalen van een herinnering aan het object. De coördinerende apparaten leveren beelden aan het protoself-systeem en houden de resultaten van de interacties (*pulses of core self*) vast in een samengevoegd patroon.

Damasio somt een aantal gebieden op waaronder associatieve nucleï van de thalamus, het claustrum en de Convergence-Divergence Zones (CDR's). De meest opvallende CDRegions zijn de *posteromedial cortices* (PMC's). Deze gebieden zijn verspreid door de hersenen. Vondsten uit de anesthesie, slaaponderzoek en brain-imaging studies (ook door Damasio uitgevoerd), laten zien dat deze gebieden vaak gebieden verbinden bij bewustzijn en tegelijk zelfprocessen faciliteren. Het zijn meestal geen informatievasthoudende plekken, maar eerder kruispunten van associaties die dingen naar boven kunnen halen en sleutelposities hebben bij zelfreflectie. Ik ga hier niet alle empirische vondsten omschrijven die Damasio betreft in zijn theorie, maar wat hij haalt uit fMRI onderzoek is al veelzeggend.

Er zijn diverse brain-imaging studies naar de PMC's te citeren om aan te tonen hoe de PMC's een rol in zelf-proces coördinatie lijken te spelen. Opvallend zijn bijvoorbeeld de fMRI-studies van Marcus Raichle.⁷⁸ Als mensen in een wakkere, maar niet actieve, staat zijn, dan is er een constante activiteit te vinden in PMC's en andere CDRegions. Er is dan sprake van een zogenaamde *default mode*. Zodra proefpersonen hun aandacht moeten richten op taken die externe stimuli betreffen, neemt die constante activiteit af. Zodra proefpersonen taken moeten uitvoeren betreffende hun autobiografie, morele beslissingen en/of toekomst verwachtingen, dan zijn die PMC- en CDRegions weer in *default mode* of zelfs actiever.

Kortom, Damasio ziet bewijzen dat de PMC's, en aanverwante CDR's, een zelfprocescoördinerende functie bekleden. Deze gebieden laten het zelf naar voren treden of naar de achtergrond afhankelijk van het feit of het een object betreft dat autobiografisch dan wel instrumenteel relevant is.

When we are unsolicited by the outside world, our self moves closer to center stage and may even move further forward when the object under scrutiny is our own person, alone or in its social setting.⁷⁹

De verschillende 'soorten' (noem het: niveau's) van het zelfproces zijn bovenop elkaar geëvolueerd, en dus niet in plaats van elkaar. Zo hebben bepaalde lager orde-zoogdieren alleen een protoself en een core self. Evolutie heeft geavanceerdere zelfprocessen bovenop de minder geavanceerde zelfprocessen gerealiseerd. Verschillende zelfprocessoorten, van geest tot bewuste geest tot bewuste geest die cultuur kan produceren, kwamen na elkaar naar voren, parallel aan de evolutie van de hersenen.⁸⁰ De verschillende zelfprocessen hangen dus enerzijds samen met gebieden in de hersenen en anderzijds met de gelaagdheid van bewustzijn.

⁷⁸ D. A. Gusnard and M. E. Raichle, "Searching for a Baseline: Functional Imaging and the Resting Human Brain," *Nature Reviews Neuroscience* 2 (2001), pp. 685-94.

⁷⁹ Damasio, A. (2010). p. 229

⁸⁰ Damasio gebruikt hier het woord: emerged (dat heb ik vertaald als kwamen naar voren). Damasio meent dat niet bewuste geest ook een emergentie is (uit neuronen die samen werken, neem ik aan). Hij positioneert zich

Wat leert Damasio's zelfproces ons over vrije wil? Damasio ziet bewustzijn als een symfonieorkest. Het is meer dan de som van de individuele musici. In eerste instantie is er geen dirigent. Die dirigent komt pas als het stuk begonnen is, maar dan wordt er ook echt gedirigeerd. Zodra een object (extern waargenomen, herinnerd of gefantaseerd) saillant wordt, dan starten hogere processen van deliberatie en beslissingen nemen.

The conductor undeniably exists in our minds, and nothing is gained by dismissing it as an illusion. [...] The coordinators are not mythical, sapient homunculi in charge of interpreting anything. And yet the coordinators do help with the assembly of an extraordinary media universe and with placement of a protagonist in its midst.⁸¹

Damasio toont zich bewust van het feit dat, ondanks dat hij zich op empirische bevindingen baseert, er nog veel aannames nodig zijn om de zelfprocestheorie hard te maken. Onze kennis van de werking van ons zenuwstelsel is nog maar bescheiden. Toch wenst Damasio zoveel mogelijk te verwijzen naar de empirie. Hij toetst zijn ideeën altijd aan objectieve kennis van de hersenen.

Het zelfproces is gelaagd en intens verbonden met alle elementen uit je biologie. Je zelfgevoel en -besef is niet te isoleren in tijd en plaats. Het is eerder een vloeibaar gegeven, dat nu weer meer betrokken is bij kwesties binnen en dan weer buiten het lichaam. Ondanks deze vloeibare staat is er weldegelijk ook enkelvoudigheid en continuïteit, en dat is dan weer te danken aan die vergaande verbinding met het lichaam.

Twee zaken vallen op als we Damasio's zelfprocestheorie lezen:

- 1) het 'zelf' is niet exclusief verbonden aan de hersenen, en
- 2) het 'zelf' is een proces. Niet een rigide entiteit.

Damasio laat specifiek beargumenteerd zien dat een handeling die uit *jezelf* komt niet alleen uit je hersenen komt. De hersenen spelen weliswaar een sleutelrol, maar het totaal van je organisme is nodig (gemedieerd door met name delen uit de hersenstam). Vooropgesteld zie ik hiermee het idee 'wij zijn ons brein' van Swaab ontkracht worden. Swaab maakt hiermee een mereologische drogreden.

Daarnaast zien we dat het maken van een beslissing voortkomt uit een proces. Niet uit een specifieke plek op een specifiek tijdstip. Dat sluit aan bij Dennett's positie. Dennett wijst ons erop dat beslissingen ook vooral niet gezien moeten worden als zaken die op een bepaald moment (in een bepaald hersengebied) genomen worden. Dit is in contrast tot wat door veel filosofen in gedachte-experimenten wordt gedaan. Zo wordt bijvoorbeeld in Van Inwagen's analyse van de mogelijkheid tot alternativiteit een specifiek moment 't' verondersteld waarop een beslissing gemaakt is en eventueel niet genomen had kunnen worden. Dennett verwerpt het mysterieuze tijdstip 't' waarop de beslissing tot handelen genomen wordt.

zo in het diverse vormen van dualisme debat door bewustzijn maar zelfs geest een emergentie te noemen. Interessant om hier Kauffman te citeren:

http://www.edge.org/3rd_culture/kauffman06/kauffman06_index.html

⁸¹ Damasio, A. (2010) p. 24

Losse tijdstippen bestaan niet, net zoals een punt in een 2D-ruimte niet bestaat in het echte leven. Dit zijn uiteraard slechts metafysische dan wel mathematische concepten, bruikbaar in hypothetische gedachte-experimenten. In de realiteit is een seconde een afspraak, niet een natuurprincipe.

Een beslissing die een mens maakt, heeft alles te maken met emotie (en daarmee met het lichamelijke, interne milieu), maar de mens heeft ook een autobiografisch zelf⁸², dat een duit in het zakje doet. Het is een samenspel van zelfprocessen dat dankzij het kernzelf ervaren wordt als eigendom: een gevoel van agentschap.

Zouden bewonderenswaardige kwaliteiten van de mens (altruïsme, abstract denken, ingaan tegen je impulsen etc.) die op vrije wil duiden, kunnen bestaan zonder deze zelfprocessen? Een beschadigd protoself, zoals bij patiënten met bepaalde hersenstambeschadigingen, kan fataal zijn voor die menselijkheid. Comateuze patiënten kunnen immers niet blijk geven van een vrije wil. Maar andersom geldt ook, schade aan gebieden die betrokken zijn bij het kernzelf, dan wel het autobiografische zelf, ontnemt de mens op een andere manier het *zelf*bewustzijn.

Maar is daarmee ook aangetoond dat beslissingen die mensen maken echt uit henzelf komen?

We hebben niet aangetoond dat deze 'zelf'-gerealiseerde handelingen en beslissingen vrij zijn. Daarover meer in de volgende paragrafen. Hierboven heeft Damasio vooral geloofwaardig kunnen maken dat handelingen/beslissingen die door de mens, als brein en lichaam, gemaakt worden ook echt producten van een zelfproces genoemd kunnen worden. Er is in mijn optiek een belangrijk verschil aan te duiden tussen een handeling/beslissing die uit jezelf komt en een handeling/beslissing die door jezelf gemaakt wordt. Dat laatste veronderstelt controle. Wat zeggen Dennett en Damasio hierover?

TEGEN JEZELF COMMUNICEREN, JEZELF AANSTUREN

Wat valt er te zeggen over zelfcontrole? Hoe kunnen we invloed op onszelf hebben? Zijn er ook neurofysiologische indicaties dat een brein zichzelf kan vormen en sturen? Is dat wat we werkelijk bedoelen als we zeggen: een handeling die echt uit jezelf komt?

Dennett ziet een sleutelrol weggelegd voor onze neuropsychologische mogelijkheden van geavanceerde interactie met anderen. Onze expertise als sociale dieren heeft ons niet alleen leren communiceren met anderen, maar ook met ons onszelf. Dennett zegt:

We wouldn't exist—as Selves “inhabiting complicated machinery,” as Wegner vividly puts it—if it weren't for the evolution of social interactions requiring each human animal to create within itself a subsystem designed for interacting with others. Once created it could also interact with itself at different times.⁸³

Uiteraard is taal een bijzonder hoogtepunt in de evolutie van het interacteren. Tegelijk heeft de opkomst van taal de mens ook adaptiever gemaakt. De mens is erdoor in staat gesteld zichzelf gemakkelijk te veranderen en aan te passen.

⁸² Het uitgebreide autobiografische zelf lijkt in dit geval ons te onderscheiden van lager orde dieren.

⁸³ Dennett, D. (2003) pp. 249-250

When language came into existence, it brought into existence the kind of mind that can transform itself on a moment's notice into a somewhat different virtual machine, taking on new projects, following new rules, adopting new policies. We are transformers.⁸⁴

Ten opzichte van andere dieren hebben onze hersenen een extra laag. Niet een anatomische laag, maar een functionele, virtuele laag, die wel in onze hersenen verwerkt is, maar dan op microniveau, en die ons in staat stelt met onszelf te communiceren. "we can ask each other to do things, and we can ask ourselves to do things."⁸⁵

We can engage in the practice of asking, and giving, reasons. It is this kind of asking, which we can also direct to ourselves, that creates the special category of *voluntary* actions that sets us apart."⁸⁶

In andere woorden, sommige van onze handelingen zijn zelfgevormd, op een moreel relevante manier. Een manier die ons verantwoordelijk maakt. Het zijn de beslissingen die je maakt na gereflecteerd te hebben op jezelf en je leven. Daarmee onderscheiden we ons van de veelheid aan evolutionaire creaties. Je kan een hond vragen dingen te doen, maar die hond kan zichzelf niet afvragen waarom je al die verzoeken hebt. Gorillas worden door de psycholoog Nicholas Humphrey beschouwd als *natuurlijke psychologen*, omdat ze het gemoed en de motieven van andere leden in de groep goed kunnen inschatten. Dat kunnen wij ook, zegt Dennett, maar daar bovenop zijn wij ook *auto-psychologen*: we kunnen ook *onszelf inschatten*.

Damasio kan ons uitleggen hoe dat werkt. Hij verwijst naar onderzoek van Giacomo Rizzolatti et al. (2001, 2004)⁸⁷. Zij hebben laten zien dat de motorische gebieden van apenhersenen, verantwoordelijk voor handbeweging, actief worden als ze de hand van de onderzoeker zien bewegen. Rizzolatti noemt dit het systeem van *mirror neurons*. MEG, fMRI en laesie-studies bij mensen wijzen ook op het bestaan van dergelijke *mirror neurons*. Damasio gaat ervanuit dat als hersenen de lichaamstoestand van iemand anders kunnen simuleren, ze ook hun eigen lichaamstoestand moeten kunnen simuleren.⁸⁸

Een voorbeeld van een dergelijk proces zou het volgende kunnen zijn: je springt weg terwijl je naar een film zit te kijken. De held in de film moest op dat moment wegspringen en omdat je zo meeleeft, spring je ook weg. Het hoeft niet altijd zo te zijn dat je medeleven zich uiterlijk toont, maar het is, volgens Damasio, wel waarschijnlijk dat motorische gebieden betrokken bij die emotie actief zijn, hoewel onderdrukt.

As we witness an action in another, our body-sensing brain adopts the body state we would assume were we ourselves moving, and it does so, in all probability, not by passive sensory

⁸⁴ Dennett, D. (2003) p. 250

⁸⁵ Dennett, D. (2003) p. 251

⁸⁶ Dennett, D. (2003) p. 251

⁸⁷ Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, 27, 169-192; Gallese, V. (2001). The shared manifold hypothesis. *From mirror neurons to empathy. Journal of consciousness studies*, 8(5-7), pp. 5-7

⁸⁸ Damasio, A. (2010) p. 103

patterns but by a preactivation of motor structures—ready for action but not allowed to act yet—and in some cases by actual motor activation.⁸⁹

De *ander* interpreteer je door visuele *ques*, taal en logische inferentie, maar ook door het waarnemen van zijn beweging. Daar maak je dan motorische representaties van in je hersenen. Bij een goede acteerprestatie wordt het als treffend ervaren als die acteur iemand anders zijn bewegingen (van type blik tot de manier van lopen en articuleren) kan nadoen.

Damasio concludeert dat het representeren van het lichaam evolutionair gezien een oud systeem is, maar wel mooie vaardigheden mogelijk heeft gemaakt voor ons mensen. Zo is het onder andere de bron van empathie en waarschijnlijk ook belangrijk in leerprocessen. We kunnen anderen begrijpen middels het simuleren van andermans lichaam in onze hersenen. We kunnen op dezelfde manier ook onze eigen lichaamstoestanden simuleren op basis van het geheugen. Communiceren met jezelf bouwt op dezelfde systemen als het communiceren met anderen. Damasio:

Body mapping of the most refined order undergirds *both* the self process in conscious minds *and* the representations of the world external to the organism. The inner world has opened the way for our ability to *know* not only that very inner world but also the world around us.⁹⁰

Hoe dicht het communiceren met anderen en met jezelf bij elkaar ligt, is zichtbaar in het dagelijks leven. Soms kan een ander bijvoorbeeld beter inschatten wat je van plan bent dan dat je dat zelf kan. “[We have a] capacity to be reminded”.⁹¹

Zo ook zijn de voorbeelden van schizofrenie sprekend. Een bepaalde stem in jezelf moet dan beschouwd worden als niet je eigen stem. Schizofrenie is een ongezonde werking van de hersenen, maar het laat wel zien dat er hersensystemen zijn die zelfcommunicatie mogelijk maken. Bewustzijn hebben zonder ernaar te kunnen handelen door een afwezige link naar die gebieden is een machteloze situatie. Zo’n bewustzijn kan niet verantwoordelijk gehouden worden. Zo ook kan een persoon de verbinding verliezen met zichzelf in tijd (wat waren toentertijd ook al weer mijn intenties om zus en zo te doen?). De ervaring van een zelf is een nuttige manier om in contact te blijven met intenties die je in het verleden had.

A *person* has to be able to keep in contact with past and anticipated intentions, and one of the main roles of the brain’s user-illusion of itself, which I call the self as a center of narrative gravity, is to provide *me* with a means of interfacing with myself at other times.⁹²

Het hersenonderzoek en de introspecties die Damasio en Dennett delen, laten zien dat het verdedigbaar is om te zeggen dat een handeling uit jezelf kan komen. Maar het is ook geloofwaardig te maken dat je jezelf kan aansturen. Kunnen we dit een handeling noemen die DOOR jezelf gemaakt is? Er valt wel iets voor te zeggen.

⁸⁹ Damasio, A. (2010) p. 104

⁹⁰ Damasio, A. (2010) pp. 106-107

⁹¹ Dennett, D. (2003) p. 252

⁹² Dennett, D. (2003) p. 253

3.2. De handeling moet met een reden zijn verricht

Het tweede criterium waar wij als mensen aan moeten voldoen volgens Swaab luidt als volgt: 'een handeling kan pas vrij zijn als die handeling met een reden is verricht'. Zijn wij dat rationele dier waar Aristoteles van uitging of zijn wij toch vooral dier en wordt het rationele deel overgewaardeerd? Zijn onze beslissingen rationeel⁹³ of impulsief⁹⁴?

Tijdens het onderzoek dat ik in 2005 heb gedaan voor mijn bachelor psychologie ontdekte ik dat mensen soms merkwaardig weinig redenen betrekken in hun beslissingsprocessen. Zelfs als het om gewichtige beslissingen gaat. Beschikbare wetenschappelijke literatuur wees al in die richting (o.a. Galotti (1995).⁹⁵). In ons onderzoek zagen we bevestigd dat eerstejaarsstudenten psychologie aan de Universiteit van Utrecht gemiddeld één argument betrokken in het beslissen welke bachelor-specialisatie ze gingen doen de jaren daarna.⁹⁶ De studenten hanteerden een zogenaamde *Take The Best* (TTB) -strategie. Mensen lijken dus vaak vooral op gevoel beslissingen te maken, aangezien het beslissen op basis van één argument geen afweging genoemd kan worden. Ondanks dat we vonden dat mensen wel een reden hadden tot handelen, was er niet echt sprake van 'redeneren'. Grappig genoeg rapporteren mensen zich zekerder te voelen over hun beslissing bij een dergelijke TTB-keuzestrategie dan als ze toch meerdere argumenten hebben afgewogen.⁹⁷

Mensen lijken dus altijd wel een reden voor hun gedrag te hebben, maar die ene reden grenst aan het handelen op gevoel. Daarom zou ik het criterium 'een handeling kan pas vrij zijn als deze met een reden is verricht' graag aangescherpt zien om zo een onderscheid te maken tussen beslissen op gevoel en beslissen na deliberatie. Een handeling zou ik pas vrij willen noemen als er meerdere redenen zijn afgewogen om tot die handeling te komen. Ik stel daarbovenop voor dat er een onderscheid gemaakt wordt tussen handelen naar de redenen van de natuurlijke wil (*natuurlijke reden*) of naar de redenen van de vrije wil (*vrije reden*).

⁹³ Mensen zijn in staat verschillende keuzestrategieën te hanteren en de keuzestrategie die gebruikt wordt hangt heel erg af van de situatie en wat voor gewicht in belangrijkheid hieraan gegeven wordt. Één van deze strategieën wordt de *rationele benadering* genoemd. Deze benadering neemt aan dat alle relevante en beschikbare informatie wordt gecombineerd in een optimale manier. De econoom en psycholoog Herbert Simon beschreef deze benadering als *substantieel rationeel*, omdat het overkoepelende doel van deze benadering het succes van de beslissing te maximaliseren is. In Simon, H. A. (1976) From substantive to procedural rationality. In S.I. Latsis (ed) *Method and appraisal in economics*. London, Cambridge University Press. pp. 129-148

⁹⁴ In contrast met Herbert Simon's *substantiële rationaliteit* is er de *Fast and Frugal Approach*. Een term te danken aan Gigerenzer, G. and Goldstein, D. (1996) in *Reasoning the fast and frugal way: models of bounded rationality*. Psychol. Rev. 103, pp. 650-6

⁹⁵ Galotti, K. M. (1995). A longitudinal study of real-life decision making: Choosing a college. *Applied Cognitive Psychology*, 9(6), pp. 459-484

⁹⁶ Burger, R., Szutkowski, K. & Witjes, R. (2005). *Beslissingsstrategieën in een Real Life Studiepad Keuze van Eerstejaars Bachelor Psychologie Studenten*. Universiteit van Utrecht, Utrecht.

⁹⁷ Galotti, K. M. (1999). Making a "major" real-life decision: College students choosing an academic major. *Journal of Educational Psychology*, 91, pp. 379-387.

Handelen naar een *natuurlijke reden* betekent dat je slechts gemotiveerd wordt door de korte termijn persoonlijke behoeftes. Handelen naar een *vrije reden* betekent dat je gemotiveerd wordt door lange termijn persoonlijke en/of niet-persoonlijke behoeftes.

We gaan in deze paragraaf kijken naar hoe Dennett en Damasio naar dit probleem kijken, oftewel, we gaan kijken of de neurofilosofie ons kan tonen dat de mens voldoet aan deze voorwaarden voor vrije wil. Aan het eind van deze paragraaf zullen we zien dat handelen op gevoel vrijheid niet in de weg staat. Eerst wordt onderbouwd dat de mens beschikt over beslissingen die gemaakt worden op basis van *vrije redenen*.

Dennett meent dat vrije redenen bestaan. Het feit dat wij mensen tot een vergevorderde vorm van coöperatie in staat zijn, maakt dat we kunnen handelen naar vrije redenen. Vrije redenen zijn morele redenen, aldus Dennett.

DE EVOLUTIE VAN COÖPERATIE

Het handelen naar niet-persoonlijke behoeftes, oftewel het handelen naar de belangen van de groep, is niet exclusief menselijk. Dennett verdedigt dat coöperatie niet exclusief menselijk is, bewust zijn van de eigen moraliteit wel.

Evolutie kan heel goed een diersoort voortbrengen die neigt naar coöperatie, vanwege de voordelen, maar dan hoeft die soort niet perse ook bewust zijn van de eigen moraliteit. Een levenssoort kan sociaal zijn zonder bewust te zijn dat het sociaal is. De mens lijkt daarentegen wel bewust te zijn van zijn eigen moraliteit. Er lijkt dus een evolutionaire ontwikkeling denkbaar met betrekking tot coöperatie.

Middels een gedachte-experiment laat Dennett zien hoe een evolutionair proces zou kunnen lopen om tot een samenleving te komen die coöperatie hoog in het vaandel heeft staan, een samenleving die gelijkheid en eerlijke verdeling expliciet viert als de juiste moraal. Dennett haalt daarvoor de evolutionaire speltheorie van Brian Skyrms naar voren (*Evolution of the Social Contract*, 1996). Skyrms introduceert de term ESS: *evolutionarily stable strategy*.

Eerlijke verdeling is een ESS, want in de meeste gevallen leidt vechten om de cake tot een slechter resultaat voor beide partijen dan wanneer er sprake is van eerlijke verdeling. Er zijn echter ook andere ESS denkbaar die ongelijkwaardigheid als constante hebben. Zo zijn er de *polymorphic traps*, waarbij uitbuiters (die 2/3 van de cake eisen) naast nederigen (die 1/3 van de cake eisen) leven. Een uitbouter zal negatieve resultaten boeken als hij de cake met een andere uitbouter moet verdelen (ze krijgen beide niets omdat ze het bevechten). De nederigen zullen altijd 1/3 deel van de cake krijgen, wie ze ook tegenover zich hebben. De uitbuiters zullen 1 op de 2 keer tegenover een nederige staan en krijgen dan 2/3 van de cake en 1 op de 2 keer tegenover een andere uitbouter en krijgen dan dus niets. Deze werkelijkheid is stabiel omdat de nederigen alsook de uitbuiters uiteindelijk gemiddeld 1/3 van de cake krijgen per contactmoment. Dat kan gezien worden als de traditie van deze groep; een moraal van ongelijkwaardigheid.

Er hoeft echter niet veel te gebeuren of deze groep gaat alsnog gelijkwaardigheid hoog in het vaandel dragen. Dit kan optreden door mutaties (zij het genetisch: verticale overdracht, zij het memetisch⁹⁸: horizontale overdracht). De nederigen kunnen besluiten de uitbuiters te gaan mijden, omdat ze daardoor een groter deel van de cake kunnen vragen aan gelijkgestemde nederigen, namelijk de helft. Ze zullen beter doorvoed raken, maar tegelijk zullen de uitbuiters nog slechts andere uitbuiters treffen en daardoor steeds minder inkomsten zien, met uitsterven tot gevolg. Deze stap naar gelijkwaardigheid wordt genomen vanwege de economische voordelen (leven is dan efficiënter) en evolutionaire voordelen (er ontstaat een stabiele verhouding tussen de leden). Is een dergelijke gelijkwaardigheid, die uit berekening ontstaat, wel echt moralistisch?

At the end of the process, do these agents know something their ancestors didn't? Sure they do: they know that fairness is just; they really *are* morally superior to their ancestors.⁹⁹

Voor typische vrije wil resultaten als moraliteit en opoffering heb je geen Cartesiaanse substantieelozee geest nodig. Deze menselijke producten kunnen heel goed ontstaan dankzij evolutionaire processen. Dergelijke producten komen niet plotseling tot stand, maar ontwikkelen zich langs een continue schaal waarin drie fasen te onderscheiden zijn: eerst vinden veranderingen in de wereld plaats door 'natuurlijke selectie' (een door uitputbare bronnen gedreven afvalrace). De volgende stap is 'onderbewuste selectie' (selecties gemaakt door organismen zelf op instinct zonder redenen). Uiteindelijk worden er veranderingen in de wereld aangebracht door 'methodologische selectie' (bewuste selecties gemaakt door organismen op basis van redeneren). Deze driestap is als eerste door Charles Darwin geformuleerd.¹⁰⁰ Het laat zien hoe bewust redeneren en moraliteit kan ontstaan uit blinde intentieloze natuurlijke selectie.

Our evolutionary journey has brought us to the traditional arena of philosophical and political investigation and debate, in which many ideas compete for our endorsement. [...] One of our most pressing tasks, as psychic engineers, is to see if we can secure the fundamental concept of a responsible moral agent, an agent who, unlike the cooperative prairie dog or loyal wolf or friendly dolphin, chooses freely for considered reasons and may be held morally accountable for the acts chosen.¹⁰¹

De mens en de samenleving sleutelen dagelijks aan de bestaande redenen om te handelen. Gewoontes en gangbare ideeën worden telkens herzien in onze moderne samenleving. Er zijn haast geen domeinen meer die niet bevraagd mogen worden. Heilige huisjes worden regelmatig omvergeworpen. Dat kan willekeurig lijken, maar het geeft ook vooral blijk van een vrijheid van redeneren. Er zijn echter ook beren op de weg: psychologen hebben keer op keer bewezen, in laboratorium- en veldonderzoek, dat we gemakkelijk te manipuleren zijn: we laten ons door emotie leiden, we laten ons redenen in de mond leggen en ga zo maar door.

⁹⁸ Overdracht middels kennisdeling. Ook wel fenotypische overdracht te noemen.

⁹⁹ Dennett, D. (2003) p. 265

¹⁰⁰ Darwin, C. (1859). *On the origins of species by means of natural selection*. London: Murray.

¹⁰¹ Dennett, D. (2003) p. 268

As we learn more and more about human weaknesses and the way the technologies of persuasion can exploit them, it can seem as if our vaunted autonomy is an unsupportable myth.¹⁰²

Uitbuiters bestaan er nog steeds die dankbaar gebruik maken van onze zwaktes. Maar dit is een situatie die zich blijvend ontwikkelt. De kennis van bepaalde technieken van verleiding, die de uitbuiters gebruiken, worden ook door de coöperatieven geleerd om weerbaarder te worden. In het boek *generatie Einstein* (2007)¹⁰³ omschrijven Boschma en Goren hedendaagse jongeren als minder vatbaar voor (traditionele) reclame. Ze zijn minder gevoelig voor trends en merken vergeleken met de generatie daarvoor. Generatie Einstein wil best kopen, maar zijn niet meer zo gemakkelijk te verleiden door op-het-gevoel-spelende-reclamecampagnes. Ze willen goede redenen hebben om een product aan te schaffen. Producenten passen hun beleid aan op deze nieuwe houding. Het in jongeren gespecialiseerde communicatiebureau Young Works¹⁰⁴ haakt in op de nieuwe houding en komt met een lijst van de best werkende benaderingsstrategieën. En zo wordt een *hide and seek*-spel gespeeld tussen de manipulators en gemanipuleerden.

Once we understand that this is an arms race, we can fend off the absolutism that sees only two possibilities: Either we are perfectly rational or we are not rational at all.¹⁰⁵

Het is juist omdat we rationeel zijn, pleit Dennett, dat we elkaar voor de gek kunnen houden. Daarom vinden we constant manieren om elkaar te misleiden, verleiden en manipuleren, maar ook weer strategieën te bedenken om ons hier tegen te weren. Het is tikkertje spelen zonder tijdslimiet of time out. Je zou het een spel tussen uitbuiters en nederigen kunnen noemen.

Coöperatie en moraliteit zijn denkbaar als resultaten van evolutie. Daarmee is geloofwaardig gemaakt dat handelen naar een vrije reden bestaat zonder dat we daarvoor een substantieel geest hoeven te bedenken.

Door het evolutionaire ontstaan van vrije redenen aan te tonen hebben we nog niet het probleem aangepakt van dat mensen vaak beslissen op basis van één reden (al dan niet een vrije reden); volgens de zogenaamde Take The Best strategie. Is deliberatie/ redeneren een illusie? Handelen we slechts op gevoel?

HANDELEN OP GEVOEL

Handelen naar een gevoel is niet hetzelfde als redenloos handelen, hebben we geleerd. Maar is daarmee delibereren een illusie? Damasio vindt van niet.

¹⁰² Dennett, D. (2003) p. 270

¹⁰³ Boschma, J., & Groen, I. G. (2007). *Generatie Einstein: slimmer, sneller en socialer: communiceren met jongeren van de 21ste eeuw*. Pearson Education.

¹⁰⁴ YoungWorks (2004). Young Works top 40: plug & play: de 40 belangrijkste hits in jongerencommunicatie, Young Works BV, Amsterdam.

¹⁰⁵ Dennett, D. (2003) p. 271

Damasio onderstreept het belang van emoties bij het begrijpen van geestprocessen en bewustzijn.¹⁰⁶ Emoties spelen een centrale rol in het verwerken van informatie en terughalen van informatie. Daarmee is ook te zeggen dat emoties een centrale rol in redeneren spelen.

Emoties zijn moeilijk te bestuderen vanwege de verscheidenheid aan soorten emotie en de daaraan te relateren processen. Een belangrijke versimpeling is het maken van onderscheid tussen emoties en gevoelens. In neurofysiologisch onderzoek worden deze twee als volgt onderscheiden.

Emoties zijn in eerste instantie acties met een cognitieve component. De acties zijn lichamelijk, van verandering in de gezichtsuitdrukking en houding tot verandering in de organen. De cognitieve component bestaat uit onze gedachten over die emoties. In het kort: emoties zijn actieprogramma's.¹⁰⁷ Gevoelens (*feelings of emotion*) zijn samengestelde percepties van hoe emotie in ons lichaam tot uitdrukking komt. Beelden van acties. "[...] the world of feelings is one of perceptions executed in brain maps."¹⁰⁸ Een belangrijk ingrediënt van gevoel is interoceptie (het lichaam verbeeld in de geest). Simpel gezegd: dieren kunnen levendige emoties hebben zonder noodzakelijk ook gevoelens te hebben.

De meeste emoties en hun lichamelijke en cognitieve effecten liggen vast, maar welke stimuli tot een emotie leiden is (gedeeltelijk) afhankelijk van cultuur en persoonlijke achtergrond. Er is weldegelijk sprake van een bepaalde mate van vormbaarheid.

Redeneren is niet exclusief een bewust proces. De kreet 'een handeling moet met een reden zijn verricht', vereist niet dat de reden (of het redeneren waar de reden uit voortkomt) in het bewustzijn moet plaatsvinden.

In het boek *Descartes' Error* (Damasio, 1994) onderbouwt Damasio het gegeven dat emotie en gevoel een belangrijke rol in redeneren spelen. Dat wil zeggen, het rationele deel van redeneren. Hij draagt diverse casestudies aan waarbij bepaalde schade aan de prefrontale cortex tot onbeslist en daarom irrationeel gedrag leidt. Bij de omschreven casestudies is vaak het algemene intellect, geheugen, taal en motoriek in orde. Op het eerste gezicht is er niets vreemds aan deze personen te merken, maar later blijkt dat deze personen moeilijk prioriteiten kunnen stellen tussen verschillende taken. Ze verliezen het zicht op het grotere doel en kunnen daardoor verzanden in details. Tegelijk is er een gereduceerd gevoel voor de ander en het zelf. Een afvlakking van emoties. Damasio heeft in dat boek het belang van emotie voor beslissingsprocessen, voor rationaliteit, aangekaart.

Handelen op basis van je gevoel heeft in onze westerse samenleving vaak een negatieve connotatie. 'Houd je hoofd koel!', wordt er dan gezegd. Nuchterheid en weloverwogenheid worden gevierd. Daar zit zeker iets in, geeft Damasio toe: impulsief handelen kan absoluut negatieve consequenties voortbrengen. Tegelijk heeft deze traditie het belang van gevoel voor de rationaliteit ondergesneeuwd.

Recente studies laten zien dat de intuïtie in bepaalde mate tot verstandigere keuzes leidt dan dat beredeneren dat doet. Damasio omschrijft een psychologisch onderzoek van de Nederlandse

¹⁰⁶ Damasio, A. (2010) p. 108

¹⁰⁷ Damasio, A. (2010) p. 115

¹⁰⁸ Damasio, A. (2010) p. 110

psycholoog Ap Dijksterhuis.¹⁰⁹ In dit onderzoek is getoond hoe in een beslissingsproces over kleine én grote aankopen onbewuste beslissingen beter uitpakken dan intensief gedelibereerde beslissingen. De resultaten van dit onderzoek suggereren dat onderbewuste processen nuttiger zijn dan bewuste deliberatie. Ook Swaab is ervan overtuigd dat ons handelen voor het leeuwendeel door het onderbewuste wordt bepaald. Onderbewuste beslissingen kunnen geregeld superieur zijn aan gedelibereerde beslissingen.¹¹⁰

Het overschatten van het onderbewuste fluit Damasio echter terug. De proefpersonen in Dijksterhuis' onderzoek hadden zonder bewustzijn het onderbewuste zijn werk niet kunnen laten doen. Bewustzijn speelt een belangrijke rol in het trainen van het onderbewuste en is ook een platform voor onderbewuste vaardigheden. De proefpersonen zijn bewust bij het ontvangen van de beslissingstaak, bij het moment van beslissen en bij het incasseren van het resultaat van hun beslissing. Het is heel handig dat onze hersenen rekenwerk (deliberatie) delegeren naar het onderbewuste, maar dat ontslaat ons niet van onze verantwoordelijkheid onderbewuste beslissingen te monitoren.

De ruimte in het bewustzijn is schijnbaar kleiner als het gaat om afwegen van redenen, kleiner dan de onderbewuste ruimte. Derhalve kunnen bewust gemaakt beslissingen soms slechter uitpakken dan beslissingen die 'intuïtief' gemaakt worden. Damasio meent dat deze kwaliteiten van het onderbewuste getraind kunnen en moeten worden door bewuste planningsprocessen, simulaties en testcases. *Practice makes perfect.*

De vrije wil lijkt dus met name tot uitdrukking te kunnen komen in lange termijn beslissingen. Daarbij kan je je eigen onderbewustzijn oefenen voor latere keuze momenten. Je kan je onderbewustzijn trainen om rationele keuzes te maken. Damasio laat in *Descartes' Error* zien dat de optimaliteit van de menselijke ratio te danken is aan een indrukwekkend doorgeëvolueerd zenuwstelsel, waarvan zelfs de lager orde hersendelen, betrokken bij emotie (lees: de oudere hersendelen in de evolutionaire geschiedenis van het brein), een rol spelen in de menselijke rationaliteit. Een door Descartes aangehangen tegenstelling van rationaliteit en emotionaliteit wordt ontkracht. Door neurofysiologische inzichten hoeven rationaliteit en emotionaliteit niet langer als concurrenten gezien te worden in het veroorzaken van gedrag, maar als partners.

Dennett reflecteert ook op het fenomeen handelen op gevoel. Hij verwijst daarin naar de psycholoog George Ainslie. George Ainslie is met name bekend om zijn experimentele werk over het contrast tussen korte en lange termijn behoeftebevrediging. Een mens, net als andere dieren, merkt George Ainslie op, neigt naar de directe bevrediging van behoeften, terwijl uitstellen vaak verstandiger is: de rationele beslissing.

De vraag dringt zich op: waar bestaat wilskracht uit? De kracht om directe behoeftebevrediging en verleidingen te voorkomen?

¹⁰⁹ Ap Dijksterhuis, "On Making the Right Choice: The Deliberation-without-Attention Effect," *Science* 311 (2006), 1005.

¹¹⁰ Swaab, D.F. (2010). Hoofdstuk XVIII. 'De Vrije Wil: Een Plezierige Illusie' in *Wij Zijn Ons Brein*. p. 296.

Kennis van de beslissingen die je neemt onder bepaalde omstandigheden, kan je helpen in toekomstige beslismomenten. Ainslie: “self-control [is] a matter of self-prediction.”¹¹¹

Het beslissingsproces om te handelen naar een impuls of deze juist in te houden bestaat volgens Ainslie uit conflicterende voorkeuren die op neuraal niveau concurreren om macht. Dat is een proces van concurrerende maar ook interacterende argumenten en gevoelens. Voor een externe toeschouwer is het onmogelijk te voorspellen wat uiteindelijk de beslissing gaat zijn.

Van deze processen kun je kennis nemen om zo te ontdekken welke rationalisaties je soms verkeerd maakt. Je kunt dus vooraf al bewust rekening houden met de (onderbewuste) deliberatie¹¹² die je later met jezelf gaat voeren als je nog voldoende afstand tot de situatie hebt.

Dennett herneemt dit en omschrijft dit als volgt: het is belangrijk jezelf “vooraf te determineren” zodat je achteraf “niet anders kan.” Door je moreel weerbaar te maken, heb je de mogelijkheden verkleind als het op beslissen aankomt. Het is paradoxaal maar dat maakt je nou juist een vrijer mens. Een gedachte die ingaat tegen de aanname van Swaab dat je, om vrij te zijn, veel mogelijkheden moet hebben en ook anders had kunnen doen op het moment van beslissen. Het ‘intertemporele onderhandelen’ van Ainslie leert ons een nieuwe kijk op de vrije wil en laat ons zien dat handelen op gevoel niet noodzakelijkerwijs betekent dat er geen redeneren heeft plaatsgevonden.

Damasio’s uitgebreide neuropsychologische onderzoek naar rationaliteit en beslissingsprocessen laat dat ook zien. Deliberatie bestaat, maar niet perse in de vorm zoals we die in eerste instantie zouden verwachten. Deliberatie vindt vaak plaats in het onderbewuste maar dit neemt niet weg dat het bestaan geloofwaardig is gemaakt van handelingen die genomen worden na deliberatie.

3.3. De mogelijkheid tot alternativiteit

DE TERREUR VAN HET ONBEWUSTE

Hoe zit het nu met de mogelijkheid tot alternativiteit? We hebben het reeds in hoofdstuk 1 over de mogelijkheid tot alternativiteit in de context van de metafysica gehad. Maar hoe zit dat als we een neurofilosofisch perspectief kiezen? Swaab spreekt immers ook graag over neurofysiologische vorm van determinisme. Hij meent dat de mens bijna uitsluitend gedetermineerd wordt door zijn genen en jonggeleerde disposities. Wij worden door processen in onze hersenen gestuurd die buiten ons bewustzijn liggen, is het idee. Is er dan nog ruimte om te denken dat we ook anders kunnen doen als dat waar is? Dit vraagstuk beantwoorden we door Dennett’s behandeling van Benjamin Libet’s onderzoek naar de *Readiness Potential* te bespreken.

Benjamin Libet heeft in 1981 een onderzoek gedaan naar het *Readiness Potential* (RD).¹¹³ Libet heeft een proef bedacht die zou kunnen aantonen dat er een verschil is in tijd tussen wanneer iemand zich

¹¹¹ Ainslie, G. (2001). *Breakdown of will*. Cambridge University Press. p 87

¹¹² Later zullen we lezen dat het woord deliberatie centraal staat in Damasio’s ideeën over de vrije wil.

¹¹³ Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, 106: pp. 623-642.

bewust is van een beslissing en wanneer die beslissing waarneembaar is in de hersenen (RD). Hij vond een verschil tussen de objectief waarneembare beslissing in de hersenen (RD) en de subjectief waargenomen beslissing (*reported time of decision*). Er zat hier gemiddeld een verschil tussen van 300 tot 500 milliseconden. Je bent niet degene die de beslissing bewust maakt, werd geconcludeerd. Later is ontdekt dat je de reeds onbewust genomen beslissing (RD) wel kunt tegenhouden middels bewuste processen, maar dan nog, zeggen aanhangers van Libet, heb je maar beperkte macht, de beslissing is immers al genomen. Dat heeft sommigen doen concluderen dat de vrije wil niets anders is dan die 100 milliseconden waarin je de reeds besloten beslissing kunt stoppen. De neurowetenschapper Vilayanur Ramachandran noemt dit “free won’t” in contrast met *free will*.¹¹⁴

Dennett vindt het problematisch om door dit onderzoek de vrije wil af te schrijven. Wie zegt dat je niet al eerder had bedacht dat je gaat handelen als de wijzer van de klok op een bepaald cijfer valt? Waarom moet je vrije beslissing samenvallen met het moment dat je achter de knoppen zit?

Een tweede probleem, volgens Dennett, is dat Libet veronderstelt dat er in de hersenen een intersectie, een samenkomen, van twee trajecten te vinden is. De plek waar we bewust worden van onze beslissing:

- De bewustwording van onze beslissing om te bewegen, en
- De bewustwording van de oriëntatie van de wijzer van de klok

“Waar gebeurt dat dan?”, vraagt Dennett zich af.¹¹⁵ In het Cartesiaans Theater? In de primaire visuele cortex of in de frontale lobe? En wat zegt dat dan over ons agentschap?

The governing idea is presumably that you can only *act* where you *are*, so if you are not *in* the faculty of practical reasoning when a decision is made there, *you didn’t make it*.¹¹⁶

Dennett meent dat er een denkfout zit in de aanname van Libet dat je een beslissing alleen kan *vetoën*: “Why couldn’t *you* have been thinking (“unconsciously”) about whether to veto *Flick!* ever since *you* decided (“unconsciously”) to flick, half a second ago?”¹¹⁷. Waarom kunnen onderbewuste processen niet eveneens een veto starten voordat je de beslissing überhaupt bewust wordt?

Patricia Churchland (1981) heeft aangetoond dat het 350 milliseconden duurt voordat iemand zich een licht-stimulus bewust wordt.¹¹⁸ Processen in de hersenen kosten tijd en gebeuren voor een belangrijk deel buiten je bewustzijn, maar je bent je er wel bewust van als het relevant wordt.

Een *Self-contained You hypothesis* is misschien hetgeen veel mensen bewezen willen zien en dat wordt helaas door Libet’s onderzoek verworpen. Deze mensen willen graag geloven dat er een plek in de hersenen is waar alles tegelijk gebeurt en waar je bewust bent van alles (tegelijk). Maar het

¹¹⁴ Ramachandran, V. S. (1998). As quoted in B. Holmes, ‘*Irresistible illusions*’. *New Scientists*, 159, 2150.

¹¹⁵ Dennett, D. (2003) pp. 232-234

¹¹⁶ Dennett, D. (2003) p. 234

¹¹⁷ Dennett, D. (2003) p. 237

¹¹⁸ Churchland, P. S. (1981). The timing of sensations: Reply to Libet. *Philosophy of Science*, 48(3), 492-497.

blijkt dat rekentaken (e.d.) moeten worden verdeeld in de ruimte en in de tijd van het brein. Dennett noemt Libet's 300 milliseconden-gat "an artifact of mis-imagined theory, not a discovery."¹¹⁹

Belangrijk is het ook om kennis te nemen van de kunstmatigheid van Libet's onderzoek. Je kunt heel snel reageren op een fenomeen als je, je van te voren daarop hebt ingesteld. Bij het onderzoek van Churchland kunnen proefpersonen binnen 300 milliseconden met een bewuste intentionele reactie komen dankzij het feit dat ze zich hebben ingesteld op de taak. Dennett noemt dat *mini-Luther cases*.¹²⁰ Hij maakt een vergelijking met de bekende quote van Martin Luther: "Here I stand, I can do no other, so help me God.". De moreel relevante beslissing wordt vaak gemaakt voordat de fase van het uitvoeren in zicht is. De proefpersonen maken de vrije wil relevante beslissing in aanloop naar het onderzoek van Libet. Namelijk, op het moment dat ze beslissen wel of niet deel te nemen aan het onderzoek. Nadat ze hebben toegezegd en voor Libet's klokje gezet worden, kunnen de hersenen van de proefpersoon op de automaat. De beslissingen worden vanaf dan uitgevoerd door automatische dispositie-achtige systemen.

We do this all the time. Our lives are full of decisions to act when the time is ripe. [...] We are the authors and executors of these policies, even though they are compiled from parts we can only indirectly monitor and control.¹²¹

Als Libet iets heeft aangetoond, dan is het dat bewuste beslissingen tijd kosten. Libet's experiment leert ons dat onze vrije wil, als ook andere geestelijke vaardigheden, zaken zijn die tijd kosten en niet op één abstract tijdstip kunnen plaatsvinden. Als er iets niet bestaat, dan is het de door vele metafysici maar ook neurofysiologen geliefde tijd 't'.

You are not out of the loop; you *are* the loop. You are that large.¹²²

Het uitvoeren van je vrije wil gebeurt op het gebied van lange termijn beslissingen, zoals: 'ik beslis mee te werken aan Libet's onderzoek als proefpersoon volgende week en dan doe ik alles wat hij me opdraagt'.

Alternatieven afwegen doe je tijdens een episode waarbij je even jezelf moet loskoppelen van externe taken en de externe omstandigheden en niet bezig moet zijn met ronddraaiende wijzers van een klok voor je neus. Op zo'n moment moet jezelf overgeven aan automatismen; aan je onderbewustzijn. Of je aandacht wordt naar buiten gericht, naar omstandigheden waarnaar je moet handelen, of je bent bewust aan het delibereren, waarbij je je aandacht naar binnen richt, naar je eigen gedachteprocessen. Het is het één of het ander. In het laatste geval raakt je bewustzijn betrokken bij het monitoren, verwerken en manipuleren van informatiestromen. Deliberatie is een proces dat offline gebeurt, je trekt jezelf terug uit de werkelijkheid.

Onderzoeksresultaten zoals die van Benjamin Libet kunnen bij overhaaste interpretatie de waarde doen onderschatten van zelfgeoriënteerde bewuste controle. Onderbewuste processen die gedrag sturen bestaan en zijn een aanzienlijke steun in het proper functioneren van het bewustzijn.

¹¹⁹ Dennett, D. (2003) p. 241

¹²⁰ Dennett, D. (2003) p. 238

¹²¹ Dennett, D. (2003) pp. 238-239

¹²² Dennett, D. (2003) p. 242

Onderbewuste processen zijn echter, voor een belangrijk deel en op diverse manieren, onderhevig aan het bewustzijn.

Er valt te verdedigen dat de jeugd van de mens zo uitzonderlijk lang duurt omdat het een lange tijd duurt voordat deze veelheid aan nuttige onderbewuste processen gevormd zijn en er een bewustzijn geïnstalleerd is dat intentioneel en doortastend controle kan uitoefenen op die onderbewuste processen. Damasio zegt:

Consciousness is not devalued by the presence of nonconscious processes. Instead, the reach of consciousness is amplified. And, assuming the presence of a normally functioning brain, the degree of one's responsibility for an action is not necessarily diminished by the presence of healthy and robust nonconscious execution of some actions.¹²³

Tijdens de evolutie zijn onderbewuste geestelementen in de hersenen langzaam de dienaren geworden van de bewuste processen. Doordat we niet bewust hoeven te zijn van de stappen die je moet zetten tijdens het lopen, heeft ons bewustzijn tijd om na te denken over de volgende handeling ('wat zeg ik straks nadat ik aangebeld heb?').

Bewust controle uitoefenen is een traject dat tijd kost en zelden of nauwelijks betrokken is bij handelingen die in een reflex worden gedaan. Delibereren en plannen kost tijd en vindt plaats ruim voordat de daarbij horende handelingen gedaan zullen worden. De handeling zelf is dan vaak een onbewust proces. "Conscious deliberation is about *reflection over knowledge*."¹²⁴

Damasio zegt dat beelden de valuta van deliberatie zijn. De vroegsensorische gebieden zijn neuro-anatomisch zo gedetailleerd dat ze perfect zijn voor het produceren van beelden: actueel, herinnerd of geconstrueerd. Deliberatie maakt gebruik van dezelfde arena als zintuiglijke waarneming in het manipuleren van beelden. Deliberatie is daarbij een soort simulatie van wat zou kunnen gebeuren.

Conscious decisions begin with reflection, simulation, and testing in the conscious mind; that process can be completed and rehearsed in the nonconscious mind, from which freshly selected actions can be executed.¹²⁵

De aandacht wordt van buiten naar binnen verlegd door waarde toekennende gevoelens. Het bestaan van bewuste deliberatie is het soort fenomeen dat de onmisbaarheid van bewustzijn onderbouwt. Bewustzijn is dus allesbehalve een functieloze kwebbeldoos zoals menigeen uit het onderzoek van Libet concludeerde.

Dat lijkt te kloppen. In 2011 heeft Baumeister et al.¹²⁶ een literatuurstudie gedaan, waarin de controle van bewustzijn over het gedrag werd onderbouwd. Opvallend is dat bij dergelijke experimenten het bewustzijn met name het gedrag op de lange termijn beïnvloedt en niet het gedrag dat à la minuut wordt gevraagd door de onderzoeker. Directe controle over motorische

¹²³ Damasio, A. (2010) p. 270

¹²⁴ Damasio, A. (2010) p. 271

¹²⁵ Damasio, A. (2010) p. 272

¹²⁶ Vohs, K. D., & Baumeister, R. F. (Eds.). (2011). *Handbook of self-regulation: Research, theory, and applications*. Guilford Press.

bewegingen is niet de belangrijkste zorg van het bewustzijn, concluderen Baumeister et al, een duidelijke kritiek op Libet's aanname dat controle een illusie is.

Bewuste deliberatie is dus nooit volledig vrij van de natuurlijke wil (de genetisch aanleg e.d.), maar andersom geldt het ook: een onbewuste handeling heeft in veel gevallen een oorsprong in een bewuste beslissing en dat maakt ons aansprakelijk voor die onbewuste handeling.

Wat heeft dan de overhand in het sturen van gedrag? Door het bewustzijn gecontroleerde onderbewuste processen of de extern en genetisch gemanipuleerde onderbewuste processen? Het antwoord is waarschijnlijk dat dat per situatie, per dagdeel, per persoon verschilt. Niet iedereen liet zich in Stanley Milgram's experiment brainwashen tot (potentiële) moordenaar. Niet alle varianten van dit experiment waren even succesvol. Damasio:

I suspect that most of us, weak but well-meaning sinners, operate on both registers, more on one or more on the other, depending on the situation and the hour of the day.¹²⁷

Het voelt echter nagenoeg altijd alsof we een bewuste beslissing hebben genomen. Dat komt omdat elke waarneming van je eigen gedrag een interactiviteit veroorzaakt tussen perceptie, lichaam en lichaammonitorende hersendelen. Er is daarom altijd een gevoel van weten (*feeling of knowing*). Een *core self impuls* begeleidt de handeling. Daniel Wegner noemt dit de *somatic marker of personal authorship*.¹²⁸ In de hersenen is er een systeem dat een 'zelf-gevoel' koppelt aan een handeling. We zijn niet slechts getuigen van de handeling, maar voelen daar ook iets bij: een gevoel van eigendom. Zonder dat gevoel zouden we geen plannen kunnen maken voor gedrag dat we in de toekomst willen vertonen en we zouden ook geen verantwoordelijkheid kunnen nemen.

Laat het pleidooi van Damasio en Dennett voor intertemporele deliberatie zien dat we een mogelijkheid tot alternativiteit hebben? Ja en nee. Nee, in de zin dat je op het moment 't', het moment van handelen, geen alternatieve mogelijkheid hebt. Tegelijk ook ja. Als we het idee kunnen loslaten dat een handeling besloten wordt op moment 't', dan is er een sterke onderbouwing in de argumenten van Damasio en Dennett te vinden voor het bestaan van de mogelijkheid tot alternativiteit. We kunnen 'geanticiperde uitkomsten' voorkomen of toelaten. Dat doen we door ons bewust voor te bereiden op het geanticiperde beslissingsmoment. Denk aan een roker die wil stoppen en zich psychologisch voorbereidt op de verleiding die gaat plaatsvinden in de kroeg later die week. Of denk aan de studente die niet voor altijd in de kroeg wil blijven werken en zich goed voorbereidt op haar tentamens (ook al is haar directe behoefte de boeken uit het raam te gooien).

In dit hoofdstuk hebben we gezien hoe de drie voorwaardelijke aspecten van vrije wil, zoals ik die van Swaab heb overgenomen, middels de neurofilosofische benadering bij de mens worden waargenomen. Een aantal zaken blijft echter open: in welke mate zijn we bereid pragmatisch naar de werkelijkheid te kijken zoals de neurofilosofie dat in werkelijkheid doet? Kunnen we het metafysische moment 't' loslaten? Kunnen we Descartes' nalatenschap, als het gaat om het lichaam-geestdualisme en de tegenovergesteldheid van ratio en gevoel, definitief afzweren? Kunnen we accepteren dat de mens niet meer is dan een door de evolutie vormgegeven machine? En, kunnen we accepteren dat

¹²⁷ Damasio, A. (2010) p. 280

¹²⁸ Wegner, D. (2002) *The Illusion of Conscious Will*, Cambridge MA: The MIT Press

het brein een prachtig en belangrijk orgaan is met betrekking tot wie we zijn, maar niet meer dan dat? Kunnen we accepteren dat vrije wil niet hetzelfde betekent als onbegrensde vrijheid?

DE PSYCHOLOGISCHE ONTWIKKELING VAN MORALITEIT

Swaab ziet onze genen als leidend in ons gedrag. De rest is te wijten aan onze periode als foetus en de eerste paar jaar van ons leven. Paradoxaal genoeg lijken we vooral vormbaar in de periode van ons leven dat we totaal geen controle over onszelf hebben. Het kleine beetje vrijheid, in de zin van vormbaarheid van de hersenen, dat de genetica ons rest, wordt verbruikt in onze vroege jaren door opvoeding en omstandigheden. Als we volwassen zijn kunnen we volgens Swaab nog maar verwaarloosbaar weinig invloed uitoefenen op de ontwikkeling van onze hersenen en dus van ons leven. De neurowetenschappelijke feitelijkheid dat belangrijke hersengebieden vooral in onze vroege levensjaren worden gevormd (m.b.t. taal, seksualiteit, karakterkenmerken, spiritualiteit) doet Swaab de vrije wil verwerpen.¹²⁹

Dennett is zich bewust van de feiten over vroege vormbaarheid, maar omschrijft deze vroege levensfase anders: dit is een fase dat we van amoreel, onvrij zuigelingschap naar volwassen moreel agentschap kunnen gaan. Het is waar, de hulp van je ouders (of voogden) is daarbij nodig.

De eerste stap in het worden van een persoon is afhankelijk van het feit of mensen om je heen (je ouders of voogden) je communicatief kunnen maken. Als dit niet het geval is dan is er sprake van pech en krijg je een lagere status in de samenleving. Na de eerste stap kunnen andere talenten ontplooi worden. Sommige aspecten aan je individuele kwaliteiten zijn genetisch bepaald, andere zaken zijn niet genetisch, maar wel aangeboren (als bijvoorbeeld je moeder veel rookte tijdens de zwangerschap). Soms kun je ook gewoon pure pech of geluk hebben; iedereen draagt een (*pseudo-*) *random number generator* in zich. De omstandigheden (leven in armoede en/of oorlog, verwend en/of verwaarloosd zijn etc.) waarin je opgroeit, kun je niet kiezen. Dat zijn allemaal dingen die buiten je controle liggen. Er zijn veel (*pseudo*)wetenschappelijke claims over hoe een kind opgevoed moet worden. Deze zijn niet altijd even bruikbaar, maar steeds vaker wel. Er wordt vooruitgang geboekt in de wetenschappen van psychologische ontwikkeling en pedagogiek. Kritisch reflecteren op vormen van opvoeden en opleiden is dan ook zeker noodzakelijk voor de vorming van de vrije wil. Daarbij formuleert Dennett een vuistregel:

Scientific knowledge is the Royal road—the only road—to evitability.¹³⁰

Het punt is dat onze opvoeding een grote impact heeft op wie we zijn, maar we worden er niet perse onvrij van. Evengoed kan onze opvoeding ons juist tot hele vrije individuen maken. Er is immers ook wetenschappelijk bewijs dat een goede opvoeding bepaalde ongewenste genetische aanleg kan bestrijden.¹³¹ De genen en ook de opvoeding zijn bepalend in welke mate wij ons ontwikkelen tot morele agenten en beschikken over 'vrije redenen'. Ik ben het dan ook absoluut eens met Swaab als

¹²⁹ Swaab, D.F. (2010). Hoofdstuk XVIII. 'De Vrije Wil: Een Plezierige Illusie' in *Wij Zijn Ons Brein*. p. 292

¹³⁰ Dennett, D. (2003) p. 275

¹³¹ Collins, W. A., Maccoby, E. E., Steinberg, L., Hetherington, E. M., & Bornstein, M. H. (2000). Contemporary research on parenting: the case for nature and nurture. *American Psychologist*, 55(2), 218.

hij ons oproept om verantwoordelijk om te gaan met de vroege jaren van het kind. Net als de evolutie is ook je context een belangrijke factor in het vormen van wie je bent. Swaab laat zien hoe een individu gevormd wordt door genen en opvoeding, tot wie hij is.

Het laten zien hoe iemand gevormd wordt doet echter niets af aan wie die persoon is. Wederom is er het Cartesiaanse nalatenschap te bespeuren in de argumentatie van Swaab. Omdat onze persoonlijkheid valt te verklaren in termen van *nature and nurture*, en niet een magische substantiële component omvat, verliest die persoonlijkheid aan waarde. De hedendaagse neurowetenschap zorgt voor een demystificatie van geestfenomenen. Maar demystificatie *an sich* vind ik geen argument tegen het bestaan van de vrije wil. Swaab vergroot nog de afstand tussen de empirische werkelijkheid en de vrije wil door te stellen dat we in ieder geval verre van absoluut vrij zijn. Daar heeft Swaab gelijk in, we zijn niet absoluut vrij, maar dat is een vrijheid die ook niet wenselijk is omdat absolute vrijheid hetzelfde is als willekeur. Vrije wil is meer dan vrijheid. Het is ook gerichtheid, het is ook wil. Die gerichtheid moet ergens vandaan komen. *Nature and nurture* zijn uitstekende kandidaten om die gerichtheid te realiseren.

4. Conclusie en discussie

4.1. De bruikbaarheid van de neurofilosofie in het vrije wil-debat

Dennett heeft een belangrijk deel van zijn carrière besteed aan het in twijfel trekken van metafysische bewijzen tegen de vrije wil. In hoofdstuk twee en drie is te lezen hoe Dennett de argumentatie van metafysische Libertariërs bekritiseert. Hij gelooft wel in determinisme, maar ziet geen reden waarom dit het concept van vrije wil in de weg staat. Een belangrijk deel van zijn argumentatie komt erop neer dat theoretische kritiek niet in de weg staat van de praktische/pragmatische mogelijkheid van vrije wil. Hij laat zien hoe metafysische concepten als het tijdstip 't' en metafysisch determinisme door hun abstracte aard misleidend werken in het denken over de vrije wil.

Met deze keuze verlangt hij van ons het Cartesiaans dualisme definitief af te zweren. Er is niets meer dan wat naturalistisch waar te nemen is. De empirie heeft hij als bondgenoot in het verwerpen van Descartes. Het Cartesiaans dualisme is al lange tijd in diskrediet onder filosofen en ook onder neurowetenschappers, maar de consequenties van dit afzweren worden nog maar mondjesmaat getrokken. Ik ben bereid mee te gaan met Dennett om definitief afscheid te nemen van het Cartesiaans dualisme en de volgende stap te zetten, teneinde een beter beeld te krijgen van het vrije wil-probleem en een eventuele oplossing.

Het pragmatisme van Dennett is wel noodzakelijk om dit neurofilosofische project te laten slagen. De vrije wil uitsluitend verklaren vanuit één verklaringshouding zorgt voor een rem op de vooruitgang van ons begrip. De zwakte daarvan zagen we in Peter van Inwagen's (meta)fysische benadering en bij de designbenadering zoals die onder andere door Dick Swaab of Paul en Patricia Churchland wordt bedreven.

Dennett's heterofenomenologische methode is daarom een aanrader. Om vrije wil beter te leren begrijpen, moeten we de verschillende verklaringshoudingen waarop dit concept betekenis heeft uitproberen en de resultaten tot elkaar verhouden. Damasio zien we die methode toepassen. In hoofdstuk drie zagen we dat zijn neurofilosofie opmerkelijke inzichten bood in de aard van de menselijke vrijheid.

De neurofilosofie vind ik daarom een vruchtbare interdisciplinaire benadering als het gaat om het vrije wil-probleem, maar wel onder de voorwaarde dat er rekening gehouden wordt met verschillende verklaringshoudingen en hun relaties.

4.2. Verschillen en overeenkomsten tussen Dennett en Damasio

Op veel vlakken zitten Dennett en Damasio op één lijn. Beiden zijn compatibilisten. Ze benadrukken ook beiden de gevaren van het Cartesiaanse lichaam-geestdualisme. Dennett verwerpt metafysische

vrije wil-kritiek gebaseerd op concepten als determinisme en causaliteit. Deze kritiek valt samen te vatten als een kritiek op de erfenis van Cartesiaans dualisme.

Als monisme waar is, dan moet een pragmatisch concept van de vrije wil gehanteerd worden. Dennett meent ook bij Damasio het gebruik van de heterofenomenologische methode te herkennen.¹³² Damasio gaat op eenzelfde manier om met het switchen tussen verklaringshoudingen om bepaalde geestfenomenen completer te begrijpen.

Her en der zijn verschillen in uitgangspunten waarneembaar. Wat zegt dat over hun positie in de vrije wil-discussie?

Aansluitend op Swaab heb ik de drie meest belangrijke (traditionele) voorwaarden gespiegeld aan de capaciteiten van de mens zoals Damasio en anderen die omschrijven. Het spiegelen van neurofilosofische kennis aan de voorwaarden van vrije wil is een onderneming die ook Dennett doet, al verwijst hij eerder naar empirische kennis uit de AI en mathematische wetenschap. Damasio betreft juist heel expliciet neurofysiologische kennis erbij, maar hij doet slechts mondjesmaat uitspraken over de vrije wil, al ligt het wel geregeld op het puntje van zijn tong. Je ziet bij hem wel neurowetenschappelijke kennis betrokken worden bij belangrijke *deel*onderwerpen uit het vrije wil-debat. Zo spreekt hij over het handelen naar ratio en gevoel, het zelfproces en indirect over de mogelijkheid tot alternativiteit. Uit Damasio's werk kunnen dus argumenten gehaald worden die gebruikt kunnen worden om te onderbouwen dat de mens voldoet aan de drie bovenstaande voorwaarden voor vrije wil.

De neurofilosofie laat zien dat er nieuwe manieren zijn waarop naar het vrije wil-vraagstuk gekeken kan worden. En daarbij komen ook nieuwe inzichten over de capaciteiten van de mens, als een lichaam dat geavanceerde hersenen bezit. Let op: de neurofilosofie ziet de mens niet als lichaam en (substantiële) geest. De integratie tussen lichaam en geest is wat zelf(besef) mogelijk maakt, aldus Damasio. Daarin spelen emoties, die hun oorsprong vinden in het lichaam, een belangrijke rol. "*Body and brain are engaged in a continuous interactive dance*"¹³³. Een verandering in de één kan grote gevolgen hebben voor de ander.

A small alteration on the brain side of the system can have major consequences for the body state (think of the release of any hormone); likewise, a small change on the body side (think of a broken tooth filling) can have major effect on the mind.¹³⁴

Om de Cartesiaanse valkuil definitief achter ons te laten, zou het wellicht beter zijn om voortaan te spreken over de mens als een wezen met lichaam *of* geest (i.p.v. lichaam *en* geest). Het zijn niet twee naast elkaar bestaande elementen, maar ze zijn hetzelfde. Het enige verschil is dat je een mens in andere termen beschrijft, afhankelijk van de verklaringshouding die je aanneemt. In hoofdstuk twee hebben we gezien dat het spreken over de mens als lichaam gedaan wordt in de zogenaamde *design stance* en het spreken over de mens als geest wordt gedaan in *the intentional stance*. Zo is er ook de

¹³² Dennett, D. (2005). A Third Person Approach To Consciousness. *Sweet dreams: Philosophical obstacles to a science of consciousness*. The MIT Press. pp. 42-44

¹³³ Damasio, A. (2010) p. 96

¹³⁴ Damasio, A. (2010) p. 96

fysische, evolutionaire, introspectieve en de morele houding van verklaren en begrijpen. Al deze perspectieven hebben bestaansrecht omdat ze verschillende soorten inzichten kunnen geven over dezelfde fenomenen. In dit geval op de vraag of de mens een vrije wil heeft. Daarbij is de fysische houding van verklaren beperkt bruikbaar in het bespreken van de voorwaarden van vrije wil, daar lijken Dennett en Damasio het over eens. Dennett vindt dat geestfenomenen veel te complex zijn voor de fysische benadering en Damasio spreekt nagenoeg nooit in fysische termen. Beiden vinden ze het niet een onjuiste benadering, maar wel een onpraktische.

Zijn er ook verschillen tussen de posities van Damasio en Dennett met betrekking tot vrije wil-specifieke onderwerpen?

Dennett vergelijkt de menselijke geest graag met een computer. Dat doet hij geregeld en op verschillende manieren. Dat is een benadering waar Damasio kanttekeningen bij plaatst. Er zijn vele overeenkomsten tussen organismen en machines, maar er zijn ook fundamentele verschillen. Een machine, zoals een Boeing 777, heeft niet de neiging naar homeostase. Homeostase zorgt voor een drang tot overleven bij elke cel van het organisme. Het totale organisme neigt daardoor naar behoud van alle micro- en macrocomponenten. De computers van een Boeing 777 neigen slechts naar behoud van het geheel, niet van alle micro- en macrocomponenten. Elk onderdeel kan vervangen of verwijderd worden zonder weerstand van het geheel.¹³⁵ De vergelijkingen tussen machines (computers) en mensen moeten dus altijd met de nodige terughoudendheid en realiteitszin gedaan worden, is Damasio's positie.

Ook de bekende naturalist Paul Churchland is kritisch naar het gebruik van software in het modelmatig testen van geestfenomenen. Dennett bekritiseert de positie van Churchland dat virtuele machinesoftware inferieur zou zijn aan hardware in het denken over, en onderzoeken van neurale netwerken. Dennett pleit daarmee dat Churchland, een naturalist en determinist, zich laat verleiden om toch de menselijke hardware te mystificeren. Dennett ziet geen onderbouwing van dit mystificeren in de biologie.¹³⁶ Dennett lijkt dus wat vrijgevinger dan Damasio als het gaat om het vergelijken van mensen en apparaten.

Damasio's aanname is dat organische materie, de cellen waaruit we bestaan, iets bijzonders is vanwege de homeostase. Elke cel heeft een drang tot overleven. Deze 'will to survive' ziet hij als een natuurprincipe. Een principe dat alle ontwikkelingen in de biologie mogelijk maakt. Tegelijk geeft hij toe dat hij niet durft te zeggen waar dit natuurprincipe vandaan komt en vindt dit ook niet belangrijk om te verdedigen. Het is er gewoon, een uitgangspunt om verder te kunnen. Je zou daarbij kunnen zeggen dat Damasio de biologie een *status aparte* geeft ten opzichte van deterministische softwaremodellen (zoals het door Dennett geprezen model *The Game of Life*). Damasio voorkomt zo, bewust of onbewust, de demystificering van de mens. Is dit geloofwaardig?

Damasio's argument is zeer beperkt onderbouwd. Het argument van behoud van micro- en macrodelen is niet indrukwekkend. Er is toch een computer te maken die geprogrammeerd is met een drang tot overleven en behoud van alle micro- en macrodelen? Bovendien is een mens, net als

¹³⁵ Damasio, A. (2010). p. 45

¹³⁶ Dennett, D.C. (2003) p. 108

een Boeing 777, bereid onderdelen van zijn lichaam te laten vervangen ten goede van het geheel. Sterker nog, een mens wordt door de tijd constant afgebroken en herbouwd op celniveau. Stel dat een Boeing voortaan gemaakt wordt uit organische materialen, is er dan nog steeds een fundamenteel verschil tussen een mens en een dergelijke machine?

Churchland lijkt eveneens geen goede verdediging van de status aparte van organische materie te presenteren, merkt Dennett op.

Ik ga hier wat dat betreft toch eerder mee met Dennett. Het is waar dat het vergelijken van computersimulaties en apparaten met de mens een beperking heeft, maar dit komt niet doordat biologische materie een status aparte verdient. De beperking zit hem in het feit dat de complexiteit van de menselijke machinerie niet geëvenaard kan worden door welke niet-biologische machine dan ook. Machines en/of softwareprogramma's die de complexiteit en ingenieursheid van de mens benaderen, bestaan op dit moment niet. Maar dat wil niet zeggen dat machines en softwareprogramma's niet bruikbaar zijn in het verklaren van bepaalde deelfuncties van bijvoorbeeld hersendelen, organen of simpele organismen.

Damasio geeft toe dat je je niet moet laten intimideren door de complexiteit van de menselijke biologie en brein. Damasio staat dus wel open voor simpele modellen als startpunt in onze kennis over het vrije wil-vraagstuk. Een fundamentele onenigheid over deze methodiek zie ik dan ook niet tussen Dennett en Damasio. Damasio:

Being awake, having a mind, and having a self are different brain processes, concocted by the operation of different brain components. They merge seamlessly on any given day, in a remarkable functional continuum inside our brains, permitting and revealing different manifestations of behavior. But they are not "compartments" as such. They are not rooms divided by rigid walls because biological processes are not at all like artifacts engineered by humans. Still, in their messy, fuzzy, biological way, they are separable, and if we do not try to discover how they differ and where the subtle transitions occur, we have no prayer of understanding how the whole thing works.¹³⁷

KRITIEK OP DAMASIO

Steven Rose, in een recensie van Damasio's *Self Comes To Mind* (2010), merkt op dat niet breinen een bewustzijn hebben, maar dat personen een bewustzijn hebben. Het is een mereologische drogreden om het tegenovergestelde te beweren.¹³⁸

Is dit terechte kritiek? Nee. Steven Rose heeft op twee punten Damasio niet goed begrepen:

Ten eerste, Damasio beschrijft in detail hoe nauw verbonden het functioneren van het brein is met de staat van het lichaam. Anders gezegd: Damasio ondergraaft juist de klassiek aangenomen superioriteit van het brein over het lichaam. Het is een wederkerige relatie. Die mereologische

¹³⁷ Damasio, A. (2010) pp. 165-166

¹³⁸ Rose, S., *Self Comes to Mind: Constructing the Conscious Brain* by Antonio Damasio – review, *The Guardian*, Saturday 12 February 2011

drogreden maakt Dick Swaab wel door te zeggen 'Wij zijn ons Brein'. Damasio ziet de mens echter als een geheel. Bij Damasio is niet het brein de mens, maar de mens is de mens.

Ten tweede, Damasio bedient zich van de heterofenomenologische methode en hanteert dus verschillende perspectieven om dieper tot de kern van het probleem (of de oplossing) te komen. Dat wil niet zeggen dat hij het morele standpunt van het bestaan van personen ontkent. Dat wil alleen maar zeggen dat Damasio verschillende perspectieven kiest met betrekking tot het vraagstuk van (zelf)bewustzijn, en daarmee neurofysiologische, subjectieve alsook intentionele en socioculturele fenomenen betreft bij het onderzoek. De persoon en de biologische mens zijn niet substantieel verschillend (pas op voor Descartes!). De persoon en de mens zijn slechts verschillend in het licht van verschillende verklarende perspectieven.

KRITIEK OP DENNETT

Ik bekritiseer Dennett's positie dat de neurofilosofie een nieuw concept van de vrije wil zou voortbrengen. Volgens hem moeten we oude concepten (zoals de Libertijnse variant van vrije wil) laten varen, omdat die vol met tegenstrijdigheden zit, en op zoek gaan naar "a variety of free will worth wanting".¹³⁹ Dat vind ik verwarrend, want Dennett (alsook Damasio) komt veelal met argumenten die juist de traditionele vormen en voorwaarden van de vrije wil verdedigen. De verklaringshoudingentheorie biedt juist kansen om eeuwenoude problemen in het vrije wil-debat te ontzenuwen. Zijn concept van de vrije wil is niet zozeer vernieuwend, maar zijn methodiek wel.

Een belangrijker probleem in Dennett's werk vind ik zijn latrelatie met de metafysica. Hij zegt geregeld dat de metafysica meer verwarring dan vooruitgang oplevert in het vrije wil-debat. Hij is bovendien kritisch op het gebruik van *the physical stance*. De *physical stance* is niet gelijk aan het gebruik van de metafysica, maar er zit weldegelijk een overlap. Dennett pleit vooral voor zijn heterofenomenologie (waarin de fysische benadering opzij gezet wordt) als methode om vooruitgang te boeken.

Dennett zoekt echter geregeld de metafysica op om zijn pragmatische concepten te kunnen positioneren ten opzichte van ongelukkige metafysische ideeën. Hij spreekt zelfs over "ontologie" in de verschillende verklaringsniveaus van de heterofenomenologie.¹⁴⁰ Hij ziet bepaalde zaken wel bestaan in bepaalde begripsniveaus en in andere niet. Om dit inzichtelijk te maken, gebruikt hij taal zoals die past bij een metafysicus.

Dennett zou wel de metafysica in ere willen herstellen, maar alleen als pragmatisme een rol krijgt daarin. Dennett is wat dat betreft een aanhanger van het werk van Stephen White (1991).¹⁴¹ Dennett vat White's positieve samen als volgt:

Don't try to use metaphysics to ground ethics, he argues; put it the other way around: Use ethics to fix what we should mean by our "metaphysical" criterion.¹⁴²

¹³⁹ Dennett, D. (2003) p. 123

¹⁴⁰ Dennett, D. (2003) p. 40

¹⁴¹ White, S. L. (1991). *The unity of the self*. Cambridge, MA: MIT Press.

Dennett zou in mijn optiek vooral zijn strijd met de wetenschappelijke gebruiken van de metafysica aan moeten gaan. Hij maakt het voor mij echter niet geloofwaardig dat de metafysica geen plek kan hebben in een pragmatisch concept van de vrije wil. Hij zegt ook weliswaar nergens dat de metafysica onbruikbaar is, maar hij geeft de metafysica geen plek in zijn heterofenomenologische methode. Ik denk dat dit een fout is. De metafysica is broodnodig in het formuleren van bepaalde concepten die relevant zijn in ons begrip van het vrije wil-fenomeen. Zelfs als dit fenomeen, in Dennett's woorden, een geëvolueerd fenomeen is en niet een natuurprincipe.

Ik heb ook mijn bedenkingen over hoe gemakkelijk Dennett de fysische benadering opzij zet. Middels een simpele rekensom in *The Game of Life* van John Conway laat Dennett zien hoe onwenselijk complex het is om gedrag van een organisme te verklaren gebruikmakend van de fysische benadering. Er zijn ontelbaar veel mogelijkheden waarop het grid van *Life* gevuld kan zijn met een *pixel*. Dat is waar, maar ik vind het wel denkbaar dat je in *Life* bestudeert wat een verandering in de natuurwet doet met een organisme. Maar een natuurwet kan soms relatief eenvoudig zijn en getest worden op microscopische schaal. Zeker in *Life*. De fysische benadering gaat toch zeker niet alleen over de wereld bestuderen door naar gedrag van de kleinste deeltjes te kijken? Het zou ook moeten gaan over wat de natuurwetten zijn en hoe die invloed hebben op die werkelijkheid.

Naar mijn mening oversimplificeert Dennett de potentie van het fysische perspectief. Dennett's terechte observatie dat het heel complex is gedrag van organismen te bestuderen op een fysisch niveau is wat mij betreft niet voldoende overtuigend om daarmee het fysische niveau te weren van het vrije wil-debat. Dennett haalt nog het argument van Hume aan¹⁴³ door te stellen dat natuurwetten a priori niet te kennen zijn. Het kan zomaar zijn dat de natuurwet luidt dat de aarde draait rondom de zon, behalve op 17 september 2016. Naar Hume zegt hij dat kennis van alle constanten in de voorgaande staten van de werkelijkheid ons niet garandeert dat in de toekomst die constanten er nog zijn. Zo probeert Dennett de *physical stance* nog maar eens in diskrediet te brengen door ons te doen twijfelen aan de betrouwbaarheid van natuurwetten. Dat vind ik een raar argument voor een pragmatist als Dennett. Hij wil wel genoeg nemen met de discutabele voorspellende kracht van wetmatigheden in het design of intentioneel niveau, maar op fysisch niveau mag dat niet. Ik vind dat Dennett onvoldoende de irrelevantie van de fysische benadering heeft aangetoond.

Let op: dit neemt wat mij betreft niet weg dat hij geloofwaardig heeft gemaakt dat de andere verklaringshoudingen relevant zijn en dat het schakelen tussen verklaringshoudingen relevant is in onze zoektocht naar een oplossing voor het vrije wil-probleem.

Anderen twijfelen aan de houdbaarheid van de heterofenomenologie. De filosoof Charles E.M. Dunlop¹⁴⁴ doet een goede poging om Dennett's voorkeur voor heterofenomenologie te ondermijnen in zijn recensie van Dennett's *Sweet Dreams* uit 2005. Dunlop bekritiseert met name het feit dat

¹⁴² Dennett, D. (2003) p. 297

¹⁴³ Dennett, D. (2003) p. 33

¹⁴⁴ Dunlop, C. E. (2006). Review of "Sweet Dreams: Philosophical Obstacles to a Science of Consciousness". *Essays in Philosophy*, 7(2), 2.

Dennett *first-person experiences* (subjectieve ervaringen) negeert, als niet ontkent. Dunlop doelt dan voornamelijk op het feit dat Dennett weigert subjectieve ervaringen toe te laten in een serieuze studie naar de aard van het bewustzijn. Dunlop beaamt de moeilijkheden in het bestuderen van subjectieve ervaringen, maar vindt het verkeerd van Dennett om het bestaan van subjectieve ervaringen (qualia e.d.) te ontkennen.

Ik denk dat Dunlop overdrijft als hij zegt dat Dennett subjectieve ervaringen ontkent in de heterofenomenologie. In hoofdstuk 2 heb ik laten zien dat Dennett introspectie weldegelijk een plek geeft in de heterofenomenologische methode. Het is echter waar dat Dennett dat slecht toelaat in de vorm van een controleperspectief. Dunlop lijkt graag te willen dat subjectieve ervaringen centraal staan in het bestuderen van geestfenomenen. Toch vind ik Dunlop's kritiek niet terecht vanwege twee redenen.

1) Ik ben het eens met Dennett dat om te bepalen wat bewustzijn is het niet nodig is te weten wat de specifieke eigenschappen van bewustzijn zijn (wat in traditionele introspectie wel gewenst werd). Slechts de algemene eigenschappen zijn relevant. In andere woorden: ik hoef niet de kleur te weten om te omschrijven wat een auto is. Kleur is een specifieke eigenschap die per auto kan afwijken, zo ook mijn subjectieve ervaring van kleur. Die kan anders zijn dan die van jou en toch hebben wij beiden een bewustzijn die in algemene termen hetzelfde is. De inhoud kan genegeerd worden in de zoektocht naar een antwoord op: "Wat is het bewustzijn?"

2) Vrije wil bestaat bij de gratie van de objectieve wereld. Betekenissen die wij met elkaar kunnen delen (overtuigingen, omschrijvingen, inschattingen etc.) staan los van subjectieve ervaringen. Als ik afzie van het eten van vlees, dan is dat omdat negatieve informatie over vleeseten mij via objectieve bronnen gecommuniceerd is. Het zijn juist deze elementen die bepalen in welke mate ik onafhankelijk van de natuur kan handelen. Vrije wil is daarom een onderwerp dat middels de heterofenomenologische methode bestudeerd kan worden. Dennett zegt dat het geheugen misschien wel het meest essentiële onderdeel van het bewustzijn is. Zo geldt dat ook voor vrije wil. De mate waarin we vrij zijn is afhankelijk van de mate waarin wij culturele informatie kunnen verwerken en paraat hebben in beslissingsmomenten.

4.3. De belofte van de neurofilosofie

Naast Dick Swaab zijn ook Victor Lamme en Marian Joëls vrije wil-sceptisch en spreken dat uit. Wat zijn hun argumenten? In eerste instantie lijken ze aanhangers van genetisch determinisme. Hun positie is dubbelzinnig, bij nadere bestudering. Bij zomergasten 2011 zegt Dick Swaab enerzijds dat therapie niet helpt, maar beveelt wel TBS en begeleiding aan in plaats van gevangenisstraf.¹⁴⁵ Het getuigt niet van een echte overtuiging van het harde genetische determinisme.

Dick Swaab beroept zich, net als Lamme, voor een belangrijk deel op anekdotische case studies waarbij mensen onvrijwillig de raarste dan wel gruwelijkste dingen doen. Hier worden geen gezonde

¹⁴⁵ Corstius, J. B. (Interviewer) & Swaab, D. (Interviewee). (2011). Retrieved from a VPRO television broadcast the 31th of July 2011.

mensen besproken. Dergelijk anekdotisch bewijs moet niet genegeerd worden, maar wel gezien worden als wat het is.

De neurofilosofie maakt het mogelijk aan te tonen dat de vrije wil compatibel is met een gedetermineerd brein. We hebben in hoofdstuk drie gezien hoe de drie klassieke voorwaarden van vrije wil gerealiseerd zijn in de mens door naar Damasio's neurofilosofische resultaten te kijken. De neurofilosofische methode opent ook de weg naar interessantere vragen over de vrije wil. Vragen die gesteld moeten worden om onszelf daadwerkelijk vrijer te maken.

De vrije wil is een vloeibaar gegeven. We zijn als mens niet onbegrensd vrij, hebben we eveneens geleerd, maar de mate waarin we vrij zijn is rekbaar. We moeten ontdekken hoe we de vrijheid kunnen maximaliseren. Daarvoor moeten we breed onderzoek doen: gebruik maken van diverse verklaringshoudingen. Daar schreef de Amerikaanse filosoof en psycholoog John Dewey in 1922 al over:

The neglect of sciences that deal specifically with facts of the natural and social environment leads to a side-tracking of moral forces into an unreal privacy of an unreal self.¹⁴⁶

De vragen die we zouden kunnen gaan stellen, kunnen gaan over onze intuïtie dat dieren geen vrije wil hebben. Hoe kan dat? Een hint naar een antwoord is het gegeven dat Dennett en Damasio het eens lijken te zijn dat een neurofysiologisch geavanceerd bewustzijn een voorwaarde is voor vrije wil.

Door neurofysiologisch onderzoek te koppelen aan onze concepten over bewustzijn leren we dat andere diersoorten dan de mens ook vormen van bewustzijn hebben. Het menselijk bewustzijn is echter geavanceerder, zo ook de mate van onze vrijheid ten opzichte van onze 'dierlijke' impulsen (directe behoeftebevrediging e.d.). Er zijn gradaties in hoe vrij een bewust wezen kan zijn in het uiten van zijn wil. Het is geen categoriaal verschil, maar een gradueel verschil. Is dit houdbaar? Hoe verloopt die gradatie? Waar trekken we de grens als het gaat om diersoorten en de mate waarin je hen de vrijheid ontnemt?

Je zou ook vragen kunnen stellen over de invloed van opvoeding, educatie en cultuur op onze vrijheid. Damasio, net als Dennett, ziet de menselijke neurofysiologische vermogens tot communiceren, informatie verzamelen en erop voortbouwen (ontwikkelingen van wetenschap, kunst, techniek, politiek etc.) als geëvolueerde capaciteiten die ons vrij(er) maken. Wat levert de memenleer ons op in ons streven naar vrijheid?

Of je zou vragen kunnen stellen over wanneer een criminele handeling wel of niet uit vrije wil uitgevoerd is. Welke lichamelijke dan wel hersenprocessen waren daarbij betrokken? Waar trekken we de grenzen qua aansprakelijkheid?

Om antwoorden te vinden op dergelijke praktische vragen moeten we accepteren dat er een vrije wil bestaat, maar dat onder bepaalde omstandigheden deze beperkt is. We moeten accepteren dat de vrije wil een vloeibaar concept is, geen categorie, geen principe. De vrije wil is net zo echt als de zuurstof die we inademen.

¹⁴⁶ Dewey, J. (1992), *Human Nature and conduct*, p. 10

4.4. Is de vrije wil van Damasio en Dennett een wenselijke vrije wil?

In een video-interview voor het online magazine wired.com (2010)¹⁴⁷ zegt Damasio tegen Jonah Lehrer dat bewustzijn de basis van cultuur is. Onze bewuste verwevenheid met cultuur maakt weer een bepaalde mate van vrijheid van natuur mogelijk. Daarom kunnen we een stuk chocolade of vette hamburger afslaan. Mijn vraag is: is dit de echt vrije wil? In andere woorden: is dit een vrije wil die we wensen te hebben?

Swaab vindt dat de vrije wil alleen kan bestaan als er sprake is van *volledige* vrijheid en de hedendaagse neurobiologie laat zien dat er absoluut geen sprake is van *volledige* vrijheid.¹⁴⁸ Dennett en Damasio zijn het hierover eens. Volledige vrijheid bestaat niet. In het boek *Elbow Room* (1984) houdt Dennett een vurig pleidooi dat we dat ook niet moeten willen hebben. Beperking geeft je ook een uitgangspunt om naar te handelen. *All we need is a little bit of elbow room*. Onze natuurlijke wil duwt ons en de samenleving in een bepaalde richting, geeft ons een doel in het leven. De intense verbinding tussen de hersenen en de rest van het lichaam speelt daarin een belangrijke rol. Onze capaciteit om te kunnen reflecteren op die voortgang en daarnaar te handelen om uiteindelijk er beter uit te komen, is dat beetje elleboogruimte dat we willen.

4.5. Nawoord

Plato (348-347 v.Chr.) was bewust van de kracht van de kunst, maar sceptisch over de integriteit van de kunstenaar en met name de theatermaker. Plato kon niet accepteren dat kunst aanspraak deed op de waarheid. De waarheid in de toen actuele tragedies bestond uit een ironische formulering van de menselijke conditie. Steeds werd de dubbelzinnigheid van het leven benadrukt in die stukken door uitspraken in de vorm van “the king is dead, long live the king”. De tweeledigheid van ironie kon geen waarheid zijn, omdat echte waarheid ondubbelzinnig is. Waarheid kwam volgens hem namelijk voort uit de ondeelbare onsterfelijke ziel.¹⁴⁹

Anekdotisch voor deze scriptie is het dat Plato's omarmen van een ondeelbare onsterfelijke ziel een puriteins idee is waar de neurofilosofie lijnrecht tegenover staat. De neurofilosofie leert ons dat lichaam en geest niet gescheiden zijn en dat de geest opgedeeld is, net zoals het brein gespecialiseerde gebieden kent.

Anekdotisch in het algemeen is het dat Plato's boeken lezen als theaterscripts, vol dialogen.

¹⁴⁷ Leher, J. (Interviewer) & Damasio, A. (Interviewee). (2010). Self Comes to Mind [Interview transcript]. Retrieved from Wired Science Blogs/Frontal Cortex: <http://www.wired.com/wiredscience/2010/11/self-comes-to-mind/>

¹⁴⁸ Swaab, D.F. (2010). Hoofdstuk XVIII. 'De Vrije Wil: Een Plezierige Illusie' in *Wij Zijn Ons Brein*. p. 292

¹⁴⁹ Plato, Book X, in *The Republic*, (Benjamin Jowett trans.) geschreven circa 38 v.Chr., Mineola NY: Dover Publications, Inc. pp. 252-77