

JASPER VAN DEN HERIK

---

# QUALIA GEDISKWALIFICEERD?

EEN ELIMINATIE VAN DENNETT'S ELIMINATIVISME

---





# **QUALIA GEDISKWALIFICEERD?**

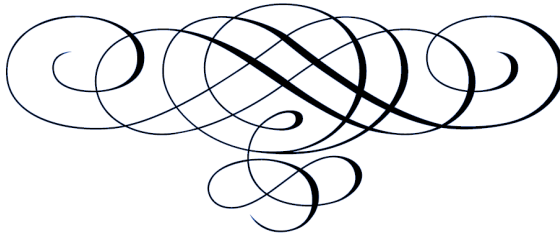
EEN ELIMINATIE VAN DENNETT'S ELIMINATIVISME



BACHELORTHESIS

# QUALIA GEDISKWALIFICEERD?

EEN ELIMINATIE VAN DENNETT'S ELIMINATIVISME



JASPER VAN DEN HERIK

2012

Ter afronding van de bacheloropleiding Filosofie aan de  
ERASMUS UNIVERSITEIT ROTTERDAM

**Auteur: Jasper van den Herik**

E: [info@jaspervandenherik.nl](mailto:info@jaspervandenherik.nl) // T: 0644804267



**Begeleider: Dr. Tim de Mey**



**Adviseur: Dr. Awee Prins**



# INHOUDSOPGAVE

Inleiding..... 1

## *PROLOOG*

**DENNETT'S THEORIE GESITUEERD.....7**

## *DEEL I*

**WAT ZIJN QUALIA?.....11**

1.1 Fenomenale ervaring en het fysische..... 11  
1.2 Het geïnverteerde spectrum argument..... 13  
1.3 Welke mentale toestanden hebben qualia?..... 13  
1.4 Qualia volgens Dennett..... 14  
1.5 Het bewijs voor qualia..... 15

## *DEEL II*

**HET OPPOMPEN VAN DE INTUÏTIE: 'MARY' EN DE  
PROBLEMEN MET NIET-FYSISCHE QUALIA.....17**

2.1 Is er iets met Mary?..... 18  
2.2 De implicaties van niet-fysische qualia..... 19  
2.3 Epifenomenale qualia?..... 21  
2.4 En de Mary-intuïtie dan?..... 22

## *EMPIRISCH INTERMEZZO*

**(ON)UITSPREEKBAARHEID ALS FACTOR? MOLY-  
NEUX EN DE MODERNE WETENSCHAP.....25**

## *DEEL III*

**VAN METAFYSICA NAAR EPISTEMOLOGIE: CHUR-  
CHLANDS MATERIALISTISCHE INVULLING VAN  
DE MARY-INTUÏTIE.....29**

3.1 Meerdere vormen van kennis in een fysicalistische wereld.... 30  
3.2 Churchlands vertrouwde hypothese..... 32

3.3 De mogelijkheid van intrinsieke eigenschappen in een functionalistische theorie.....	34
3.4 Weten-dat het <i>dit</i> is om rood te zien.....	35
3.5 Implicaties van de vertrouwdheidshypothese van Churchland.....	36

*DEEL IV*

**DENNETT'S METAFORENOORLOG: EEN FICTIONALISERING VAN HET BEWUSTZIJN.....39**

4.1 Heterofenomenologie: een verschuiving van het explanandum.....	40
4.2 Van Cartesiaans Theater naar meervoudige conceptversies...41	
4.3 Bewustzijn als virtuele machine.....	43
4.4 Het bewustzijn als linguïstische machine: de narratieve fictionalisering .....	44

*FILOSOFISCH INTERMEZZO*

**EEN BEROEP OP WITTGENSTEIN: TAAL SPELEN VERSUS ONTOLOGIE.....47**

*DEEL V*

**DENNETT'S ELIMINATIVISME.....51**

5.1 Reactieve disposities en bierdrinkers.....	51
5.2 De blauwe banaan.....	53
5.3 EEG-Mary.....	54
5.4 Twee verhalen, maar waar zijn de argumenten?.....	55
5.5 Spelregels voor 'Mary'.....	55
5.6 RoboMary.....	57
5.7 Dennetts herinterpretatie van de logische structuur van 'Mary'.....	59
5.8 Terug naar de menselijke vermogens: RoboDennett.....	61
5.9 Een heterofenomenologische analyse van RoboDennett.....	63
5.10 De behavioristische trekken van de heterofenomenologie....	64
5.11 De onderdeterminatie van 'RoboMary'.....	65



*DEEL VI*

**QUALIA: VAN REACTIEVE DISPOSITIES NAAR DE  
VERTROUWDHEIDSHYPOTHESE.....67**

- 6.1 Defictionaliseren: de grondslag van de fenomenologie  
in het brein.....67
- 6.2 Hoe theoretische kennis onze perceptie niet verandert.....69
- 6.3 Een wittgensteiniaanse analyse van Churchlands  
neurofysiologisch geïnspireerde introspectie.....70
- 6.4 De onreduceerbaarheid van het eerste-persoonsperspectief:  
de vertrouwdheidshypothese geherformuleerd.....72

*CONCLUSIE*

**DE ONMOGELIJKHEID VAN EEN DERDE-PER-  
SOONSTHEORIE VAN BEWUSTZIJN.....75**

**REFERENTIELIJST.....79**

**BRONVERMELDING FIGUREN.....88**



# INLEIDING

*Your body is made up of some trillions of cells, each one utterly ignorant of all the things you know. If we are to explain the conscious Subject, one way or another the transition from clueless cells to knowing organizations of cells must be made without any magic ingredients.*

DANIEL C. DENNETT (2001, p. 229)

*But that is the beauty of it! In a proper theory of consciousness, the Emperor is not just deposed, but exposed, shown to be a cunning conspiracy of lesser operatives whose activities jointly account for the miraculous powers of the Emperor. Banished along with the Emperor are what might be called the Imperial Properties: the two most mysterious varieties being the Qualia Enjoyed by the Emperor and the Imperial Edicts of Conscious Will.*

DANIEL C. DENNETT (2003b, p. 15)

*The most formidable challenge to any sort of materialism and functionalism comes from the friend of phenomenal qualia.*

DAVID K. LEWIS (1983, p. 130)

*Het bewustzijn verklaard* luidt de overambitieuze titel van Dennetts (1991/1999) boek waarin hij zijn theorie over bewustzijn uiteen zet. Vanuit een wetenschappelijk – dat wil zeggen materialistisch en evolutionair – kader probeert Dennett een oplossing te vinden voor het probleem dat filosofen sinds Descartes heeft beziggehouden: wat is de relatie tussen lichaam en geest. Of, hoe is het mogelijk dat in

een materialistische wereld, uit een verzameling cellen die elk op zichzelf niet bewust zijn, ons bewustzijn ontstaat.

Het boek van Dennett heeft tot veel reacties geleid. Criticasters menen dat het boek beter *Het bewustzijn vermeden* (McGinn, 1995) of *Het bewustzijn genegeerd* (Block, 1993) had kunnen heten. Dennett (2003b) meent namelijk dat het bewustzijn gezien kan worden als een goochelshow: er lijkt voor ons een innerlijke wereld te bestaan, maar dit is slechts een truc. Wanneer het bewustzijn op een wetenschappelijke wijze beschouwd wordt kan deze truc worden doorzien. De innerlijke wereld die voor ons allen zo evident is, wordt in de theorie van Dennett gefictionaliseerd. Hoewel het lijkt alsof wij vanuit een eerste-persoonsperspectief de wereld beschouwen verdwijnt deze wereld onder het wetenschappelijke vergrootglas.

Een dergelijke these, die ingaat tegen bijna eenieders intuïties en vrijwel alle filosofische theorieën, verdient een zeer degelijke onderbouwing. In *Consciousness Explained* probeert Dennett echter vooral de bestaande metaforen waarmee we nadenken over bewustzijn te vervangen door een nieuwe set aan metaforen (Kirk, 1993). Het is hierdoor vaak onduidelijk wat Dennetts theorie zelf inhoudt (McGinn, 1995). Robinson (1993) noemt Dennetts methode daarom ook de 'Jericho-methode'. Het lijkt erop dat Dennett denkt dat wanneer hij maar vaak genoeg rond een probleem heen loopt het vanzelf uiteenvalt. Het is daarom moeilijk om Dennetts positie te beschrijven. Interpretaties van Dennett lopen uiteen van reductionistisch (Hutto, 1995), eliminativistisch (Ramsey, 2011; Mason, 2005), speculatief empirisch (Kirk, 1993), metaforisch (McGinn, 1995) tot simpelweg incoherent (Block, 1993).

## Het probleem van fenomenale ervaring

In zijn reacties op 'Mary', het kennisargument van Frank Jackson (1982, 1986) dat wellicht het bekendste gedachte-experiment van de twintigste eeuw (Graham en Horgan,

2000) is, wordt hij echter gedwongen om zijn ideeën te expliciteren. In deze scriptie zal ik daarom dit gedachte-experiment als rode draad nemen. Met betrekking tot onze fenomenale ervaringen is de *die-hard* fysicist<sup>1</sup> Dennett zeker een eliminativist, waarbij hij meent dat het voor zijn theorie van bewustzijn cruciaal is dat ze verworpen worden.

Maar wat zijn deze fenomenale ervaringen? Ze zijn het best te beschrijven aan de hand van een voorbeeld: wat gebeurt er wanneer je een rode roos ziet? In het wetenschappelijke verhaal wordt kleurloos licht van een bepaalde golflengte door de roos gereflecteerd en opgevangen in de receptoren van je retina, waarna je hersenen deze input verwerken tot het rood van de roos. Maar is hiermee het hele verhaal verteld? Er lijkt hiernaast immers ook iets te zijn wat het *voor jou* is om het rood te zien, een eigenschap van de ervaring die zich moeilijk laat reduceren tot de fysische, wetenschappelijk beschrijfbare processen die aanleiding geven tot diezelfde ervaring.

Deze fenomenale eigenschappen van onze perceptuele ervaringen spelen al lange tijd een belangrijke rol in debatten rond het lichaam-geest probleem en de mogelijkheid van een wetenschappelijke theorie van bewustzijn. Onder de noemer qualia voeren filosofen levendige debatten over het metafysische statuut en de epistemische implicaties van deze kwalitatieve eigenschappen. In reactie op de wetenschap, die ons heeft laten zien dat 'in de wereld'<sup>2</sup> slechts kleurloze elektromagnetische straling, geurloze chemische verbindingen en klankloze trillingen van luchtdeeltjes voorkomen, hebben filosofen geprobeerd om de kleuren, geuren en klanken die we ervaren een plek te geven in de waarnemer. De qualia, als ze niet te reduceren zijn tot de fysische

---

1 In deze scriptie gebruik ik 'materialisme' en 'fysicalisme' door elkaar.

2 In de wereld betekent hier in de objecten zelf, in plaats van in de waarnemer, die in een materialistische visie zelf natuurlijk ook in de wereld is (zie ook Dennett, 1991)

processen, lijken dan metafysisch een apart soort entiteiten, en vormen zo een belangrijk argument tegen het materialisme.

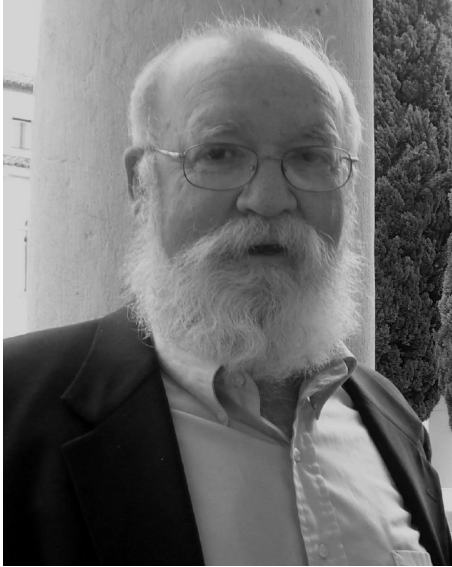
Nu zien we ook waarom voor een materialistische filosoof als Dennett hier dus een belangrijke uitdaging ligt. Om onze fenomenale ervaringen onder te brengen in een materialistische theorie moet een adequaat alternatief gevonden worden voor onze dualistische intuïties. Het is dus begrijpelijk dat Dennett in het genoemde boek een apart hoofdstuk inruimt om in te gaan op de uitdaging die de qualia vormen.

### **Een eliminatie van Dennetts eliminativisme: doel en opbouw van deze scriptie**

In deze scriptie zal ik laten zien dat zelfs wanneer we Dennetts theorie volledig accepteren hij geen duidelijke, laat staan beslissende, argumenten kan geven voor zijn eliminatie van onze fenomenale ervaringen. De argumenten die hij geeft zijn ondergedetermineerd, en laten ruimte om andere materialistische theorieën, in het bijzonder die van Churchland, aan te nemen. Ik zal in deze scriptie dan ook betogen dat Dennetts fictionalisering van het bewustzijn overbodig is. Hiernaast zal ik laten zien dat een degelijke analyse van het probleem met betrekking tot qualia laat zien het onmogelijk is om een adequate wetenschappelijke, ofwel derde-persoonstheorie te geven.

Om bovenstaande conclusie te kunnen ontwikkelen en verdedigen is deze scriptie opgebouwd uit verschillende delen. In het eerste deel zal ik bekijken wat qualia zijn. In het tweede deel bestudeer ik het gedachte-experiment van Frank Jackson en de implicaties van zijn aanname van niet-fysische qualia. Vervolgens geef ik een beknopt overzicht van de mogelijke materialistische reacties op Jackson's gedachte-experiment, met speciale aandacht voor de *vertrouwdheidshypothese* van Churchland.

In het vierde deel geef ik een introductie van de belangrijkste elementen uit Dennetts theorie van bewustzijn die van belang zijn om in het volgende deel te concluderen dat zijn verdediging van de eliminatie van fenomenale ervaringen ondergedetermineerd is. In deel zes zal ik vervolgens laten zien dat de vertrouwdheidshypothese van Churchland een adequaat alternatief vormt dat niet wordt uitgesloten door Dennetts argumentatie. Bovendien zal ik hier laten zien dat een derde-persoons of wetenschappelijke theorie van bewustzijn onmogelijk is door de idiosyncratische opbouw van de menselijke hersenen.



Figuur 1: Daniel C. Dennett



## PROLOOG

# DENNETTS THEORIE GESITUEERD

*I finished a presentable draft of my dissertation in the minimum time (six terms or two years) and submitted it with scant expectation that it would be accepted on first go. On the eve of submitting it, I came across an early draft of it, and compared the final product with its ancestor. To my astonishment, I could see Ryle's influence on every page.*

DANIEL C. DENNETT (2008)

*But when [Frank Jackson] goes on to proclaim a sharp distinction between "behaviorism, functionalism, eliminativism, and instrumentalism as competing answers," he sees differences that don't hold up, in my opinion. My answer, in any event, mixes elements from all of these, and denies that there is any good reason to cleave to a less eclectic answer.*

DANIEL C. DENNETT (1993)

Zoals in de inleiding al aangegeven is het onduidelijk onder welke noemer we de theorie van Dennett moeten brengen. Hijzelf meent dat het bestaande jargon van de *philosophy of mind* ernstig ontoereikend is (Dennett, 1993). Voor criticasters van Dennett is het daarom vaak onduidelijk welke posities hij aanvalt, of juist onderschrijft. In deze proloog bespreek ik kort de twee belangrijkste invloeden op Dennetts denken, het behaviorisme en het functionalisme.

De eerste kandidaat is het behaviorisme, een stroming in de filosofie die groot was in het midden van de twintigste eeuw. Behavioristen probeerden het gedrag van mensen te verklaren zonder daarbij gebruik te maken van mentale concepten. In plaats daarvan probeerden ze het gedrag van de mensen volledig in stimulus-respons relaties te duiden. Oorzaken van gedrag, het studieobject van de behavioristen, zijn altijd extern aan het subject. Wanneer een mentaal concept gebruikt wordt in een verklaring moet dit geëlimineerd worden en in termen van gedrag worden vertaald of geparafraseerd (Graham, 2010).

Van groot belang – zo blijkt uit bovenstaand citaat uit Dennetts autobiografie – is de invloed van Gilbert Ryle, de analytisch (of logisch) behaviorist, bij wie Dennett studeerde in Oxford. Ook Wittgenstein, een andere belangrijke inspiratiebron voor Dennett, wordt soms tot de behavioristen gerekend, hoewel deze classificatie zeer wordt betwist (Graham, 2010). In deze scriptie is een filosofisch intermezzo opgenomen over de invloed van Wittgenstein op Dennett. Hier zal ik laten zien dat – hoewel Wittgenstein zelf wellicht geen behaviorist was – Dennetts ontologische interpretatie van het *private language argument* eenvoudig tot behavioristische conclusies kan leiden.

In de filosofie van Dennett vinden we verder behavioristische trekken in zijn methode van heterofenomenologie, waarbij we om het bewustzijn te begrijpen slechts moeten afgaan op derde-persoonstoegankelijke data, ofwel gedrag en uitspraken van mensen. Meer over deze methode volgt in deel IV van deze scriptie.

*Dennett 1991 defends a position with strong affinities to behaviorism, though it is better classified as functionalist.*

Graham (2010)

De tweede belangrijke kandidaat om de Dennetts theorie van bewustzijn te classificeren, is het functionalisme, een

stroming die in het verlengde ligt van het behaviorisme. De simpele stimulus-respons relaties worden vervangen door functionele relaties tot gedrag en andere mentale toestanden. Het functionalisme breed gedefinieerd beschouwt mentale toestanden dan ook als exhaustief gekarakteriseerd door de functionele rol die ze spelen, niet door hun concrete, materiële realisatie (Levin, 2010; zie ook Churchland, 1989). In *Het bewustzijn verklaard* stelt Dennett zelf expliciet dat zijn theorie een vorm van functionalisme is: “Misschien zal uiteindelijk toch een variant van het functionalisme triomferen (in feite verdedigt dit boek een bepaalde versie van het functionalisme).” (Dennett, 1991/1999, p. 44).

Een goed voorbeeld om het functionalisme te begrijpen is de Chinese natie van Ned Block (1990), bedoeld als tegenargument tegen het functionalisme: Wanneer we alle Chinezen een mobiele telefoon geven en in dezelfde functionele verhouding tot elkaar laten staan als de neuronen in een brein (ze verwerken dezelfde informatie, hebben dezelfde verbindingen etc.) heeft China als geheel dezelfde mentale eigenschappen als een mens heeft. Dennetts reactie hierop is *biting the bullet*: China heeft in dat geval inderdaad mentale eigenschappen.

In deze scriptie stel ik, vanwege de beschikbare ruimte, het functionalisme niet ter discussie, maar ga hierin mee met Dennett (en ook Churchland, zie deel III deze scriptie).



# WAT ZIJN QUALIA?

*The sensation of colour cannot be accounted for by the physicist's objective picture of light-waves. Could the physiologist account for it, if he had fuller knowledge than he has of the processes in the retina and the nervous processes set up by them in the optical nerve bundles and in the brain? I do not think so.*

ERWIN SCHRÖDINGER (1952, p. 154)

*Nothing you could tell of the physical sort captures the smell of a rose.*

FRANK JACKSON (1982, p. 39)

## 1.1 Fenomenale ervaring en het fysische

Een uitspraak van het type “deze roos is rood” dicht een eigenschap toe aan een object in de wereld, in dit geval de roos. Maar schijn lijkt te bedriegen. Uit de moderne wetenschap weten we dat we de rode roos kunnen zien omdat hij licht van een bepaalde golflengte, die binnen de spectrale gevoeligheid van onze kegeltjes ligt, reflecteert. Dit licht op zichzelf is niet gekleurd, maar brengt samen met de specifieke werking van onze ogen en onze hersenen, in ons de sensatie van rood voort.

De Britse empirist John Locke (1690) maakte om deze relatie tussen de wereld en de waarnemer te funderen het onderscheid tussen primaire eigenschappen van objecten, zoals extensie en vorm, en secundaire eigenschappen, zoals kleur, geur en smaak. Deze secundaire eigenschappen zijn geen eigenlijke eigenschappen van het object zelf, maar zijn vermogens om sensaties in een waarnemer te veroorzaken die berusten op de primaire eigenschappen van het ob-

ject. In een dergelijke visie is een roos zelf dus niet rood, maar heeft slechts het vermogen om in een mens – in Locke's termen – het simpele idee van rood te veroorzaken. De kleuren zitten dus niet 'in de wereld', maar in de waarnemer.

De fenomenale aspecten van ons mentale leven, zoals de rode kleur van de roos, de geur van versgezette koffie, of de klank van een klarinet, hebben dus een subjectief karakter: ze lijken niet toegankelijk te zijn voor iemand anders dan onszelf. Er is iets *wat het is om* rood te zien dat niet besloten ligt in het rode object dat waargenomen wordt maar dat slechts gevormd wordt in de waarneming van een rood object. Het is om deze reden dat Thomas Nagel (1974) betoogt dat het onmogelijk is om te weten *wat het is om* een vleermuis te zijn. Dit 'wat-het-is-om' lijkt niet besloten te liggen in de fysische informatie. Hoeveel we ook weten van een vleermuis, we zullen nooit weten wat het is om er één te zijn. Maar wat zijn deze illustere ervaringen dan?

De qualiafielen, of qualia freaks, zoals de voorvechters van qualia zichzelf graag noemen, hebben hier een antwoord op gevonden. De eerste die het woord qualia in zijn huidige betekenis noemde was Clarence Irving Lewis (1929):

*There are recognizable qualitative characters of the given, which may be repeated in different experiences, and are thus a sort of universals; I call these "qualia." But although such qualia are universals, in the sense of being recognized from one to another experience, they must be distinguished from the properties of objects. Confusion of these two is characteristic of many historical conceptions, as well as of current essence-theories. The quale is directly intuited, given, and is not the subject of any possible error because it is purely subjective.*

Objecten in de wereld hebben dus het vermogen om qualia te produceren in de waarnemer, waarbij qualia de introspectief toegankelijke, fenomenale en puur subjectieve aspecten van onze mentale levens zijn (Tye, 2009). De directe introspectieve toegankelijkheid duidt op een indirect realisme: de qualia zitten tussen het subject en de wereld in. Hoewel we ons kunnen vergissen over de eigenschappen van de wereld kunnen we ons niet vergissen over onze waarneming daarvan.

## 1.2 Het geïnverteerde spectrum argument

Een klassiek gedachte-experiment om de onafhankelijkheid van onze fenomenale ervaring ten opzichte van de fysieke objecten te expliciteren is het geïnverteerde spectrum argument, dat ook al beschreven is door John Locke<sup>3</sup> (1690). Hoe kunnen we weten dat het rood dat persoon A ervaart hetzelfde rood is als dat persoon B ervaart? Het lijkt denkbaar dat de fenomenale ervaring van de ene persoon het spiegelbeeld is van de andere. Deze qualia bemiddelen zo tussen het object en het bewustzijn van het subject: ze zijn de oorzaak van de cognitieve respons van het subject als gevolg van de waarneming van het object, terwijl ze tegelijkertijd onafhankelijk zijn van de effecten, en dus kunnen variëren terwijl de effecten hetzelfde blijven (Block, 1990).

## 1.3 Welke mentale toestanden hebben qualia?

Naast onze perceptuele ervaringen als zien, ruiken en proeven zijn er ook andere mentale toestanden die een specifiek subjectief gevoel hebben. Denk bijvoorbeeld aan lichamelijke sensaties als pijn en honger. Maar ook voor emoties en gemoedstoestanden is er een duidelijk subjec-

---

3 Locke was niet de bedenker van het geïnverteerde spectrum gedachte-experiment. Tim de Mey (persoonlijke communicatie, 22 juli 2012) meent dat Pierre Gassendi dit gedachte-experiment voor het eerst bedacht.

tief gevoel van wat het is om een dergelijke toestand te ervaren (Tye, 2009). Over andere mentale toestanden bestaat meer discussie. Moeten we de ervaring van het begrijpen van een zin bijvoorbeeld ook tot de toestanden rekenen waarbij een fenomenale ervaring optreedt (Strawson, 1994, zoals beschreven in Tye, 2009)? Hoewel dit interessante problemen zijn, vallen ze buiten de reikwijdte van deze scriptie.

In deze scriptie focus ik daarom exclusief op perceptuele ervaringen. Hierbij sluit ik mij aan bij de filosofische traditie. De meeste debatten rond qualia gaan over kleur, denk aan het zojuist genoemde geïnverteerde spectrum argument, maar ook het kennisargument van Jackson dat in het tweede deel van deze scriptie aan bod zal komen.

## 1.4 Qualia volgens Dennett

Hoewel Dennett (1988) het bestaan van qualia ontkent, geeft hij wel een duidelijke definitie. Hij is het eens met de brede definitie van qualia als het fenomenale karakter van onze ervaringen, ofwel de wijze waarop een ervaring is voor mij, waarbij dit fenomenale karakter strikt onafhankelijk is van alle disposities en reacties die voortkomen uit een quale. Hij gooit de qualia hierbij op één hoop met andere concepten die een brug proberen te slaan tussen het bewustzijn en de wereld, als *raw feels*, *sensa*, fenomenale kwaliteiten, intrinsieke eigenschappen van bewuste ervaringen en de kwalitatieve inhoud van mentale toestanden (Dennett, 1991).

Daarnaast geeft hij een viertal eigenschappen die qualia in zijn ogen ook moeten hebben (Dennett, 1988). Qualia zijn volgens hem

1. *onuitspreekbaar*, het is onmogelijk om *wat het is om* groen te zien met iemand te bespreken. Je kan wel andere kleurwoorden gebruiken, door bijvoorbeeld te zeggen dat groen tussen blauw en geel ligt, of andere groene objecten benoemen, maar het specifieke



fenomenale karakter van kleuren in het algemeen is onbeschrijfbaar. Het is onmogelijk om aan een blinde persoon uit te leggen wat kleuren zijn.

2. *intrinsieke eigenschappen*<sup>4</sup> van een ervaring. Dit veronderstelt een atomisme waarin de quale niet te reduceren is tot iets anders. Hieruit volgt bovendien de onmogelijkheid van de analyse van qualia.
3. *privaat*, er is geen enkele objectieve test, of die nu fysiologisch of louter gedragsmatig is, waardoor het mogelijk zou zijn om een interpersoonlijke vergelijking te maken tussen qualia.
4. *direct toegankelijk* voor het bewustzijn. Omdat qualia eigenschappen zijn van *mijn* ervaring is er een directe toegang mogelijk tot de qualia en is het niet mogelijk om je te vergissen met betrekking tot qualia.

## 1.5 Het bewijs voor qualia

Omdat qualia per definitie slechts vanuit het eerste-persoonsperspectief toegankelijk zijn, is de studie ervan gecompliceerd. Vaak wordt gepoogd het bestaan ervan via gedachte-experimenten – of intuïtiepompen in de woorden van Dennett (1988, 1991) – aan te tonen. Eén van deze gedachte-experimenten zijn we al tegengekomen: het geïnverteerde spectrum argument. Eind twintigste eeuw deed Jackson (1982, 1986) echter een voorstel voor een nieuw gedachte-experiment dat voor- en tegenstanders van qualia tot op de dag van vandaag bezighoudt. In het volgende deel bekijk ik dit gedachte-experiment en de problemen met de niet-fysische qualia die hieruit lijken te volgen.

---

4 Volgens Churchland en Churchland (1985) volgt uit het intrinsieke karakter niet dat een eigenschap niet analyseerbaar is. Zie ook Churchlands verdediging van intrinsieke eigenschappen in 3.3 van deze scriptie.



Figuur 2: Mary's kamer

# HET OPPOMPEN VAN DE INTUÏTIE: 'MARY' EN DE PROBLEMEN MET NIET-FYSISCHES QUALIA

*Inderdaad kunnen we ons allemaal levendig voorstellen dat ze voor het eerst een rode roos ziet en uitroept: 'Dus zo ziet rood eruit!'*

DANIEL C. DENNETT (1991/1999, p. 437)

Hoewel de intuïtie dat qualia een niet-fysisch element toevoegen aan onze fysische wereld sterk is bij velen is een brute stipulatie hiervan retorisch niet sterk. Het is waarschijnlijk daarom dat de intuïtiepomp van Frank Jackson (1982) zoveel aandacht heeft gekregen en de gemoederen tot op de dag van vandaag beroert. Het gedachte-experiment maakt niet zozeer een nieuwe kijk op qualia mogelijk, maar het expliciteert en voedt de intuïtie dat qualia niet te reduceren zijn tot een naturalistische beschrijving. In dit deel bekijk ik het gedachte-experiment van Jackson. Ook bekijk ik waarom zijn niet-fysische conclusies<sup>5</sup> tot meer problemen leiden dan ze oplossen.

---

5 Jackson is midden jaren negentig – zo'n 15 jaar na de eerste publicatie van zijn gedachte-experiment – radicaal van positie gewisseld (Stoljar & Nagasawa, 2004). Hij meent niet langer dat er niet-fysische qualia zijn. Voor een beschrijving van deze ommekeer en een korte introductie van Jackson's nieuwe positie zie Jackson (1998).

## 2.1 Is er iets met Mary?

In zijn gedachte-experiment (Jackson, 1982; Jackson, 1986) zit de briljante kleurenwetenschapster Mary vanaf haar geboorte in een kamer waar alles zwart/wit is, inclusief haar eigen lichaam<sup>6</sup>. Via zwart/wit beeldschermen en boeken heeft Mary toegang tot alle *complete* – dat wil zeggen ideale – theorieën betreffende de neurofysiologie, biologie, scheikunde en fysica van kleur en kleurwaarneming, maar ook alle informatie over bijvoorbeeld kunst en de affectieve reacties van mensen op kleuren. Ze weet dus welke golflengten het visuele systeem van de mens beïnvloeden, hoe licht de retina stimuleert, hoe deze signalen door de hersenen worden verwerkt, hoe mensen oordelen vormen over kleuren, welke kleuren een alarmfunctie hebben of als plezierig worden ervaren etc.

Wanneer Mary alles weet wat er te weten is met betrekking tot kleuren en kleurervaringen verlaat ze de zwart/witte kamer en wordt ze geconfronteerd met een rode roos. De vraag is nu of ze op dit moment iets nieuws leert. Intuïtief zullen de meeste mensen hier positief op antwoorden, ze weet nu immers *wat het is om* rood te zien, iets wat ze terwijl ze nog in haar zwart/witte kamer opgesloten zat, niet wist. Maar haar kennis omvatte wel alle fysische informatie, dus hetgeen ze leert kan niet fysisch zijn, aldus Jackson. Ergo, het fysicalisme betreffende het lichaam/geest probleem is niet houdbaar<sup>7</sup>.

---

6 Een alternatief voor de zwart/witte kamer is het uitschakelen van kleurwaarneming in Mary zelf. Van belang is dat Mary op geen enkele wijze kleurervaringen heeft, ook niet door bijvoorbeeld in haar ogen te wrijven of door gekleurd te dromen. Hoewel er onenigheid is over de mogelijkheid om het gedachte-experiment daadwerkelijk uit te voeren zal ik in deze scriptie ervanuit gaan dat het logisch mogelijk is om Mary af te sluiten van elke kleurervaring.

7 Deze intuïtie, dat Mary iets nieuws leert wanneer ze voor het eerst een kleur ziet zal ik in deze scriptie de Mary-intuïtie noemen. 'Mary' verwijst naar het gedachte-experiment zoals hier toegelicht.

Voor de materialist is het duidelijk dat de conclusie van dit gedachte-experiment moet worden verworpen; er bestaan voor haar immers enkel fysische objecten en eigenschappen. Maar voordat ik in het derde deel bekijk hoe Mary's bijzondere situatie materialistisch geherinterpreteerd kan worden beschouw ik in dit deel eerst de implicaties van Jackson's metafysische conclusie, het bestaan van iets niet-fysisch.

## 2.2 De implicaties van niet-fysische qualia

Als we de metafysische conclusie van Jackson accepteren, moeten we ook het bestaan van twee soorten informatie aannemen (Lewis, 1988). Immers, wanneer de fenomenale ervaring niet besloten ligt in de fysische eigenschappen van objecten en waarnemer, geeft deze ervaring ons andere informatie over de wereld, met name informatie die niet kan worden gereduceerd tot de fysische informatie. Dit noemt Lewis (1988) de hypothese van de fenomenale informatie. De enige manier waarop deze hypothese zinvol is, is wanneer er niet-fysische dingen of processen zijn, of fysische dingen met niet-fysische eigenschappen, op basis waarvan het mogelijk is om een verschil waar te nemen in de wereld terwijl er geen fysisch verschil is.

Een minimaal fysicalistische theorie moet echter uitgaan van superveniëntie, ofwel geen verschil zonder een fysisch verschil<sup>8</sup>. Wanneer het dus mogelijk is, zoals de qualia freak aanneemt, dat een quale een verschil mogelijk maakt tussen twee situaties die fysisch hetzelfde zijn – de quale

---

8 Superveniëntie betekent dus dat elk verschil in mentale toestand gespiegeld moet zijn in de fysieke toestand. In de woorden van Jackson (1998b, p. 12): “Any world which is a minimal physical duplicate of our world is a duplicate simpliciter.” Het fysicalisme kan dan gedefiniëerd worden als: “Among worlds where no natural properties alien to our world [zoals qualia, JC] are instantiated, no two differ without differing physically; any two such worlds that are exactly alike physically are duplicates.” (Lewis 1983, p. 364).

volgt immers niet uit de fysische eigenschappen – dan moet het fysicalisme verworpen worden.

Echter, wanneer we de qualia freaks in alle opzichten gelijk geven, en dus uitgaan van fenomenale informatie leidt dit tot opmerkelijke conclusies. Wanneer we aannemen dat wat het is om rood te zien niet alleen afhangt van ons visuele systeem in combinatie met licht van een bepaalde golflengte, dan had rood er ook anders uit kunnen zien. Denk ook terug aan het geïnverteerde spectrum argument. Maar, voor Mary ziet rood er niet anders uit dan het er voor haar uitziet, waardoor de niet-fysische quale - ofwel de fenomenale informatie - alle mogelijkheden behalve de feitelijke uitsluit. Alternatieven zijn voor Mary zowel van te voren als naderhand ondenkbaar. Dus hoewel de niet-fysische qualia veronderstellen dat de fenomenale informatie zou kunnen variëren in dezelfde fysische situatie, is dit een vooronderstelling die we nooit kunnen toetsen (Lewis, 1988).

Hiernaast levert de hypothese van fenomenale informatie een probleem op voor de wijze waarop wetenschap bedreven wordt. Immers, het zien van kleuren stuurt het gedrag van mensen, en wanneer we ervan uitgaan dat in de waarneming een toevoeging ontstaat op de fysische informatie moeten we concluderen dat niet-fysische informatie een effect heeft op fysische processen. Eén van de grondgedachten van de fysica is echter de causale geslotenheid van de fysische wereld. Dit betekent dat fysische processen altijd door andere fysische processen veroorzaakt zijn (Jaegwon Kim, 1993). Deze causale geslotenheid van de wereld impliceert tevens de fysische geslotenheid, bekend van bijvoorbeeld de wet van behoud van energie. Elke niet-fysische inbreuk op de fysische wereld sluit de causale geslotenheid van de wereld echter uit. De hypothese van fenomenale informatie sluit dan dus uit dat de huidige wetenschap volledig zou zijn.

## 2.3 Epifenomenale qualia?

Een verweer tegen deze laatste stelling is het veronderstelde epifenomenale karakter van qualia. Hoewel qualia volgens Jackson (1982) wel veroorzaakt worden door fysische processen hebben zij op hun beurt geen invloed op die fysische processen. Maar dit epifenomenale karakter brengt nieuwe problemen met zich mee voor de qualia freak. Wanneer qualia immers epifenomenaal zijn ten opzichte van de fysische wereld, dan kan het feit dat Jackson uitgaat van het bestaan van qualia niet veroorzaakt worden door de qualia zelf (Dennett, 1991). Epifenomenen kunnen namelijk geen enkel fysisch effect hebben, en dus niet causaal verantwoordelijk zijn voor Jacksons uitspraken dat qualia bestaan. Jackson (1982) meent echter dat qualia, hoewel ze geen effecten hebben op de fysische wereld, wel een effect kunnen hebben op andere, niet-fysische mentale processen. Wanneer echter de qualia werkelijk epifenomenaal zijn, kunnen deze niet-fysische mentale processen ook geen effecten hebben op de fysische wereld. Epifenomenale qualia leiden dus tot een extreem solipsisme, waarbij er voor elk subject een mentale wereld bestaat die geen enkele invloed kan hebben op de mentale wereld van een ander of op de (gedeelde) fysische wereld.

Deze solipsistische conclusie heeft echter nog een onbedoeld effect. Stel dat er twee werelden zijn, waarbij wereld A louter fysisch is en wereld B hier een exacte kopie van is met toevoeging van de epifenomenale qualia zoals Jackson die voorstelt<sup>9</sup>. Omdat de qualia causaal ineffectief zijn met

---

9 Over mogelijke werelden waarin exacte fysische replica's van mensen rondlopen die geen fenomenale ervaringen hebben bestaat een uitgebreid debat onder de noemer filosofische zombies. Dennett stelt dat ze niet voorstelbaar zijn (immers veronderstelt een exacte fysische replica voor de materialist een exacte replica per sé), terwijl Chalmers filosofische zombies wel denkbaar acht. Voor Dennetts mening zie Dennett (1995) waaruit het volgende citaat komt: "Sometimes philosophers clutch an insupportable hypothesis to their bosoms and run headlong over the cliff edge. Then,

betrekking tot de fysische wereld zal Jackson in zowel wereld A als B precies hetzelfde artikel schrijven over epifenomenale qualia. In wereld A zal Jacksons artikel niet waar zijn, omdat hier geen qualia voorkomen, terwijl in wereld B hetzelfde artikel wel waar zal zijn. Indien Jacksons these dat qualia, wanneer ze bestaan, epifenomenaal zijn dus correct is, dan kunnen we niet weten of we in wereld A of B leven. Met andere woorden, als qualia inderdaad epifenomenaal zijn, kunnen we nooit weten of ze bestaan.

Het gedachte-experiment van Jackson stuit dus op enkele bezwaren. Nemen we zijn conclusie over dat er niet-fysische qualia bestaan, dan moeten we het bestaan van fenomenale informatie ook aannemen, wat leidt tot de merkwaardige conclusies dat onze fenomenale wereld anders zou kunnen zijn – iets dat we nooit kunnen toetsen – en dat onze hele fysica fundamenteel incorrect is. De oplossing voor deze problemen is de aanname dat qualia epifenomenaal zijn. Dit leidt echter tot een extreem solipsisme en is bovendien fundamenteel onbewijsbaar. Het aannemen van niet-fysische qualia lijkt dus meer problemen op te leveren dan het oplost.

## 2.4 En de Mary-intuïtie dan?

Maar de Mary-intuïtie - dat Mary iets leert wanneer ze haar zwart/witte kamer verlaat - staat nog steeds overeind. Wanneer we op basis van bovenstaande kritiek de qualia niet willen aannemen, moeten we een andere oplossing vinden voor de (schijnbare?) ontdekking die Mary doet wanneer ze met de rode roos wordt geconfronteerd. In het vol-

---

like cartoon characters, they hang there in mid-air, until they notice what they have done and gravity takes over. Just such a boon is the philosophers' concept of a zombie, a strangely attractive notion that sums up, in one leaden lump, almost everything that I think is wrong with current thinking about consciousness." (Dennett, 1995, p. 322). Zie Cottrel (1996) voor het onderscheid tussen Dennett en Chalmers.



gende deel bekijk ik daarom de mogelijkheden voor een fysicistisch filosoof om een antwoord op 'Mary' te geven. Maar voordat ik daaraan begin eerst een empirisch intermezzo over de rol van onuitspreekbaarheid van qualia.



# (ON)UITSPREEKBAARHEID ALS FACTOR? MOLYNEUX EN DE MODERNE WETENSCHAP

*Suppose a man born blind, and now adult, and taught, by his touch to distinguish between a cube, and a sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and t'other, which is the cube, which the sphere. Suppose then the cube and the sphere placed on a table, and the blind man to be made to see: Quare, whether by his sight, before he touched them, he could now distinguish, and tell, which is the globe, which is the cube.*

JOHN LOCKE (1690, II.ix.8)

In persoonlijke communicatie laat Bill Lycan (Dennett, 2006) weten dat Mary in haar gevangenschap haar toekomstige fenomenale ervaring van kleur niet kan afleiden uit een verzameling van onpersoonlijke wetenschappelijke informatie. Dit verleidt Dennett ertoe te onderzoeken of de onuitspreekbaarheid van fenomenale ervaring een belangrijk aspect is. Volgens Dennett is deze onuitspreekbaarheid slechts het gevolg van de hoge mate van informatie, en is deze nog nooit door iemand principieel bewezen. Uit het feit dat we op dit moment niet in staat zijn om een beschrijving te geven, kan niet afgeleid worden dat het principieel onmogelijk is.

Dennett maakt hierbij de vergelijking met een wel uitspreekbare eigenschap, bijvoorbeeld die van driehoekig zijn, en stelt vervolgens dat: “*Someone who has never seen or touched a triangle can presumably be told in a few wel-*

*l-chosen words just what to expect, and when they experience their first triangle, they should have no difficulty singling it out as such on the basis of the brief description they had been given*” (Dennett, 2006, p. 21). Maar is dit wel zo? Zijn mensen die nog nooit vormen visueel hebben waargenomen meteen in staat om deze te herkennen op basis van een paar simpele woorden?

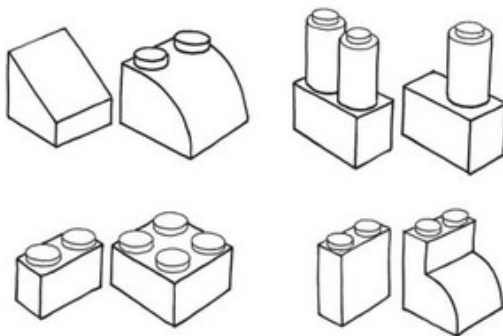
Het probleem waar Dennett hier terloops overheen stapt is een bekend filosofisch probleem dat uit de tijd van Locke stamt. In 1688 schreef William Molyneux een brief aan Locke (Locke, 1690; zie ook Degenaar en Lokhorst, 2011), waarin hij hem vroeg of een blinde man, die op basis van tast al bekend was met het verschil tussen een kubus en een bol, deze visueel uit elkaar kon houden mocht zijn visuele perceptie plotseling hersteld worden. Een interessant filosofisch probleem, dat lange tijd vooral een gedachte-experiment leek.

Met de moderne medische wetenschap is het echter mogelijk om de visuele perceptie van blindgeboren kinderen te herstellen. Een recente studie hiernaar leverde interessante uitkomsten op (Held et al., 2011). Het bleek namelijk dat proefpersonen objecten zowel met hun tastzin als met hun net herstellende visuele perceptie uit elkaar konden, terwijl ze tegelijkertijd de twee perceptuele modaliteiten niet aan elkaar konden koppelen. Een object dat ze hadden gevoeld konden ze dus vervolgens niet op basis van visuele perceptie herkennen, en vice versa. Zelfs voor primaire eigenschappen lijkt er dus bij eerste kennismaking een aha-moment te zijn: “Dus zo ziet een kubus eruit!”.

De kinderen die als proefpersonen fungeerden zijn natuurlijk geen alwetende wetenschappers en vormen dus geen goede vergelijking met de briljante Mary. Maar deze experimentele resultaten laten wel zien dat bij visuele perceptie van een eenvoudige geometrische vormen meer aan de hand is dan de simpele beschrijving die Dennett voorstelt. Het visueel herkennen van vormen is iets dat je moet

leren wanneer je visuele perceptie plotseling hersteld wordt<sup>10</sup>. De simpele beschrijving van Dennett volstaat niet om de ervaring van simpele geometrische vormen uitputtend te beschrijven.

De empirische resultaten met betrekking tot het probleem van Molyneux geven ons eerder een aanwijzing om primaire eigenschappen – naast de secundaire eigenschappen als kleur – ook tot die eigenschappen te rekenen waar Mary verbaasd over zou zijn wanneer ze ze voor het eerst zou zien. Wanneer Mary volledig blind zou zijn – en niet slechts kleurenblind – is het allesbehalve duidelijk of ze kubussen en bollen direct uit elkaar zou kunnen houden wanneer haar visuele perceptie hersteld zou zijn.



Figuur 3: Voorbeelden van de gebruikte stimuli (Held et al., 2011)

---

10 Vijf dagen na hun operatie scoorden de proefpersonen als normale personen op de test van Held et al. (2011)



Figuur 4: Paul M. Churchland

# VAN METAFYSICA NAAR EPIS- TEMOLOGIE: CHURCHLANDS MATERIALISTISCHE INVULLING VAN DE MARY- INTUÏTIE

*Our capacity for recognizing a range of [...] inarticulate features in our subjective experience is easily explained on materialist principles [...] and signifies nothing about their metaphysical status.*

PAUL M. CHURCHLAND (1989, p. 169)

Voordat ik in deel IV en V de positie van Dennett uiteenzet is het van belang om eerst te bekijken welke reacties een fysicalist kan hebben op het gedachte-experiment van Mary. Om de anti-fysicalistische conclusie te voorkomen, lijkt het immers alsof de Mary-intuïtie verworpen moet worden. Als Mary daadwerkelijk alle fysische kennis heeft, hoe kan ze dan nog iets bijleren? Dit is dan ook de visie van Dennett. Verrassend is wellicht dat dit een minderheidspositie in het debat rond qualia vormt.

Verreweg de meeste materialistische filosofen houden vast aan de Mary-intuïtie. Met andere woorden: zij veronderstellen een epistemisch contrast tussen Mary<sub>1</sub>, de alwetende Mary die nog geen kleurervaringen heeft gehad, en Mary<sub>2</sub>, die deze kleurervaringen wel heeft gehad<sup>11</sup>. De meta-

---

11 Indelingen tussen posities waarin het epistemisch contrast en daarmee de Mary-intuïtie behouden blijft en posities die beide verwerpen zijn bijvoorbeeld *thick* versus *thin* (Graham en Horgan, 2000), *gappy* versus *non-gappy* (Mandik, 2010) en Type-B versus Type-A materialisten (Chalmers, 2007). De verschillende auteurs

fysische conclusie van Jackson wordt hier dan vervangen door een epistemologische: de kennis die Mary in haar opsluiting verkrijgt, is van een andere vorm dan de kennis die ze verkrijgt wanneer ze een kleur ervaart.

### 3.1 Meerdere vormen van kennis in een fysicalistische wereld

Hoe kan het zijn dat alle fysische kennis met betrekking tot kleuren en kleurendervaring niet leidt tot een kennis van de fenomenale ervaring van kleuren in een louter fysische wereld? Churchland (1989) licht dit toe met een voorbeeld van een golfer. Een competente golfer heeft een motorrepresentatie van een goede golf-*swing* in zijn motorcortex. Op basis hiervan kan hij een golfbal wegslaan. Waarschijnlijk heeft hij ook een propositionele representatie van een golf-*swing*. Dit stelt hem in staat met anderen over het golfen te praten en om mensen die net beginnen te golfen te instrueren hoe ze zich deze motorische representatie aan kunnen leren. De twee systemen zijn echter onafhankelijk van elkaar. Een goede golfer hoeft niet te kunnen uitleggen hoe hij zijn *swing* uitvoert: zolang de bal maar in de hole gaat, heeft hij geen propositionele representatie nodig. En een goede sportarts kan een prima propositionele representatie hebben terwijl hij een golfbal nooit in de hole zal kunnen krijgen.

Deze analyse van kennis veronderstelt dat, zelfs indien Mary alles zou leren over golfen, zij na het verlaten van haar kamer niet direct zou kunnen golfen. Om een vaardigheid aan te leren is immers oefening nodig. Wanneer de kennis die Mary verkrijgt op basis van haar fenomenale ervaring van kleur dus van een andere vorm is dan de propositionele kennis die ze op kan doen in haar gevangenschap kan de Mary-intuïtie behouden blijven zonder het fysicalisme op te geven. Om dit in te zien is het goed om naar de lo-

---

delen verschillende posities soms net iets anders in. Dennett wordt in ieder geval gezien als *thin, non-gappy* Type-A materialist.



gische vorm van 'Mary' te kijken, waarbij  $Mary_1$  beschouwd wordt, dat wil zeggen voordat ze een kleurervaring heeft gehad:

- |  |  |
|--|--|
| (i) $\forall x ( Fx \rightarrow Kmx )$ | $Fx$ : x is fysisch;                                   |
| (ii) $\exists x ( \neg Kmx )$          | $Kyx$ : y weet dat x;                                  |
| (iii) $\exists x ( \neg Fx )$          | $m = \text{Mary}$ ; en x zijn<br>alle kenbare objecten |

Ofwel: (i) voor alle kenbare objecten geldt dat als ze fysisch zijn Mary ze kent. Er is echter (ii) iets dat Mary niet kent, de Mary-intuïtie. Met modus tollens kunnen we dan concluderen dat (iii) er iets moet zijn dat niet fysisch is. Het is nu duidelijk dat wanneer de twee predikaten  $Ksx$  van elkaar zouden verschillen de antifysicalistische conclusie niet langer volgt:

- |  |   |
|--|---|
| (iv) $\forall x ( Fx \rightarrow K(p)mx )$ | $K(z)sx$ : s weet x op wijze z;                                     |
| (v) $\exists x ( \neg K(y)mx )$            | $p = \text{propositioneel}$ ;<br>$y = \text{niet-propositioneel}$ ; |

Deze fysicalistische herinterpretatie van 'Mary' behoudt dus de Mary-intuïtie, maar dan in de vorm van (v). Er is een andere vorm van kennis die Mary nog niet heeft in haar opsluiting, maar die ze wel verkrijgt wanneer ze een kleurervaring heeft. Omdat het een andere vorm van kennis betreft die niet te reduceren valt tot de volledige propositionele kennis die ze in haar zwart/witte opsluiting had, leidt dit dus niet tot de metafysische conclusie dat er niet-fysische qualia bestaan.

Er zijn verschillende mogelijkheden om deze andere vorm van kennis te definiëren. Lewis (1980) en Nemirow (1990) stellen bijvoorbeeld dat Mary een nieuwe vaardigheid leert wanneer ze kleurervaringen krijgt<sup>12</sup>. Na haar vrij-

---

<sup>12</sup> Lewis (1983) geeft echter wel toe dat Mary informatie in een *bruikbare vorm* mist. Zijn vaardigheidspositie komt hierdoor behoorlijk dicht bij de vertrouwdehypothese van Churchland, die ik

lating krijgt ze slechts het vermogen om een kleur voor te stellen, te onderscheiden en te herinneren, zonder dat ze nieuwe propositionele kennis opdoet. Om een verdediging tegen Dennetts eliminativisme te geven – zoals ik deel VI van deze scriptie zal doen – is echter vooral de theorie van Churchland (1989) interessant, omdat Churchlands positie het dichtst bij die van Dennett ligt<sup>13</sup> en toch ruimte maakt voor de Mary-intuïtie. Ik zal deze daarom hier kort introduceren.

### 3.2 Churchlands vertrouwdheidshypothese

Volgens Churchland (1989) moeten we de waarneming van kleuren zien als een vertrouwdheid met een ervaring (*knowledge by acquaintance*). Dit is een derde vorm van kennis naast propositionele kennis en vaardigheden. Zolang Mary in haar kamer is opgesloten, mist ze niet slechts bepaalde vaardigheden zoals Lewis (1988) stelt, maar is ze niet bekend met een complexe representatie van een toestand van haar visuele systeem. Dit doet niets af aan het feit dat ze in haar taalcentrum een gedetailleerde beschrij-

---

daarom verder zal gebruiken. Het is bovendien onduidelijk wanneer de ervaring slechts de vaardigheden is *wat* je je herinnert of voorstelt. Churchland (1989) geeft hier een expliciet antwoord op. Lewis (1983, p. 131-132) legt wel helder uit waarom het bestaan van meerdere vormen van kennis het afleiden van de ervaring uit propositionele kennis onmogelijk maakt: “*Now imagine a pattern-recognizing device that works as follows. When exposed to a pattern it makes a sort of template, which it then applied to patterns presented to it in future. [...] There is no reason to think that any such device must have [another] faculty: a faculty of making templates for patterns it has never been exposed to, using its stored information about these patterns. If it has a full description about a pattern, but no template for it, it lacks an ability but it doesn't lack information. (Rather, it lacks information in usable form.)*”.

13 Beiden worden gezien als eliminatief materialisten, zijn fel gekant tegen onze psychologie van de koude grond, ofwel volkspychologie en menen dat een reductie van onze volkspychologische concepten tot wetenschappelijke concepten bij zal dragen aan ons begrip van het bewustzijn (Churchland, 1985; Dennett, 1991).

ving heeft van de hersenprocessen die leiden tot deze complexe representatie, maar geeft wel aan waarom het – zelfs in een physicalistische wereld – onmogelijk is om uit deze propositionele kennis de ervaring van een kleur af te leiden.

Wanneer het oog geprikkeld wordt door gekleurd licht treedt er een specifieke activatie van het neurologische systeem op die verantwoordelijk is voor de kleurwaarneming:

*In creatures with trichromatic vision (i.e., with three types of retinal cone), color information is coded as a pattern of spiking frequencies across the axonal fibers of the parvocellular subsystem of the optic nerve. That massive cable of axons lead to a second population of cells in a central body called the lateral geniculate nucleus (LGN), whose axonal projections turn to several areas of the visual cortex at the rear of the brain's cerebral hemispheres, to V1, V2, and ultimately to V4, which area appears especially devoted to the processing and representation of color information.*

Churchland (1989, p. 165-166)

In hersengebied V4 is dus een representatie van de kleuren wanneer we ze waarnemen. Deze voortalige representatie kan volgens Churchland worden geïntrospecteerd op een wijze die overeenkomstig is met onze perceptuele vermogens<sup>14</sup>. Doordat het visuele systeem gescheiden is van ons linguïstisch vermogen hebben we geen directe linguïstische toegang tot de hersenprocessen die de kleurwaarne-

---

<sup>14</sup> Churchland (1985, p. 16) zegt bijvoorbeeld “our introspective discrimination and recognition”. Verdere analogen die hij trekt tussen perceptuele processen en introspectieve processen in dit artikel ondersteunen de aanname dat introspectie een hogere-orde perceptuele representatie is van lage orde representaties, in het geval van kleuren in gebied V4. Zie ook Shoemaker (1984).

ming veroorzaken, maar slechts tot de representatie van de kleur. Hierdoor kan kennis verkregen via dit visuele systeem niet terug worden gevoerd tot een beschrijving op het niveau van propositionele kennis. Maar de complexe representatie in V4 maakt wel al onze vaardigheden met betrekking tot kleuren mogelijk: het discrimineren, herkennen en voorstellen. De kleursensaties zoals wij die in onze volkpsychologie formuleren en gebruiken, zijn volgens Churchland dus niet te beschrijven in een derde-persoonsperspectief en zullen onze wetenschappelijke theorieën altijd ontgaan. De enige manier waarop we dus kennis kunnen nemen van een kleursensatie is door die kleur tegen te komen in de wereld en ernaar te kijken<sup>15</sup>.

Wel stelt Churchland (1985, 1989) dat de wetenschap in de toekomst wellicht wel een beschrijving kan geven van wat er precies gebeurt in de hersenen wanneer we kleuren zien. Als dit mogelijk zal blijken te zijn – een empirisch feit volgens Churchland (1985) – dan kan het epistemisch contrast tussen  $Mary_1$  en  $Mary_2$  volledig verdwijnen omdat de kennis door vertrouwdheid dan gereduceerd kan worden tot propositionele kennis. Uiteindelijk is de *Mary*-intuïtie dus niet houdbaar. In het laatste hoofdstuk kom ik uitgebreid terug op deze reductie die Churchland voorstelt, waar ik zal betogen dat deze reductie onmogelijk is, en het door Churchland veronderstelde onderscheid tussen beide vormen van kennis noodzakelijk moet blijven bestaan.

### **3.3 De mogelijkheid van intrinsieke eigenschappen in een functionalistische theorie**

Op het eerste gezicht lijkt het functionalisme incompatibel met intrinsieke eigenschappen. Als een mentale toe-

---

15 Of door een directe stimulatie van de neuronen die projecteren naar V4. Een goed voorbeeld hiervan is een drukkingsfosfeen, ofwel de kleuren die je ziet als je hard in je ogen wrijft. De kleuren die hierbij ontstaan zijn echter altijd gerepresenteerd in V4. Voor een opmerkelijk voorbeeld van directe interne stimulatie in een synestheet, zie Ramachandran en Hubbard (2001).

stand immers slechts gedefinieerd wordt door zijn functionele rol lijkt er geen mogelijkheid meer om de intrinsieke eigenschappen van deze toestand een plek te geven in het functionalisme. Churchland en Churchland (1981) menen echter dat deze redenering berust op een misvatting omtrent het functionalisme. Het geïnverteerde spectrum argument is mogelijk in een functionalistisch perspectief. We kunnen ons voorstellen dat de representaties in hersengebied V4 worden omgewisseld, zodat wat eerst als groen gezien werd nu als rood gezien wordt en vice versa. De *functionele rol* van de twee representaties is nu omgedraaid. Volgens de Churchlands hoeft de functionalist slechts aan te nemen dat er *iets* is dat de functionele rol van de representatie van rood op zich kan nemen, zonder dat hierbij moet worden aangenomen dat deze representatie van rood uitputtend beschreven wordt *door* deze functionele rol.

### 3.4 Weten-dat het *dít* is om rood te zien

Maar één probleem lijkt nog steeds overeind te staan. Stel dat Mary twee ervaringseigenschappen kent, eerst op een derde-persoons wijze, als twee neurologische eigenschappen. Wanneer ze hierna met deze eigenschappen op een directe eerste-persoonswijze wordt geconfronteerd zou ze het volgende kunnen zeggen:

*So this is what this experiential property is like, and that is what that one is like. I now know that this one is not that one, and that one is not this one. I now know 'which way around' the experiential properties in fact happen to be. They are this way around, rather than the other way around.*

Bigelow & Pargetter (2004, p. 189)

Dit lijkt op het eerste gezicht een propositionele ontdekking, en vormt ook de aanleiding voor de intuïtie dat Mary

iets leert wanneer ze voor het eerst met een kleur wordt geconfronteerd. Bigelow en Pargetter (2004) menen echter, dat hoewel het voor Mary wellicht kan lijken alsof de fenomenale eigenschappen omgedraaid hadden kunnen zijn, dit geen werkelijke mogelijkheid is. De toestand waarmee ze eerst alleen in het derde-persoonsperspectief vertrouwd was is identiek met één van de ervaringstoestanden (namelijk degene die het geval blijkt te zijn) en niet met de andere. Hoewel voor Mary dus, voordat ze op een eerste-persoonswijze vertrouwd was met het fenomenale karakter van beide eigenschappen, en beide opties dus epistemisch en logisch<sup>16</sup> mogelijk waren, was slechts één van beide nomologisch mogelijk. De andere optie, die dus slechts epistemisch (of logisch) mogelijk is, heeft geen corresponderende mogelijke wereld om hem waar te maken. In het geval dat haar neurofysiologie is zoals die feitelijk is, zal slechts één kleurervaring mogelijk zijn in een materialistische, functionalistische<sup>17</sup> wereld.

### **3.5 Implicaties van de vertrouwdheidshypothese van Churchland**

Volgens Churchland zijn er dus verschillende manieren waarop we toegang hebben tot dezelfde informatie. Wanneer Mary voor het eerst in haar leven vanuit het eerste persoonsperspectief een kleur ervaart, krijgt ze dus niet meer informatie over de wereld (dat de kleuren ook anders

---

16 Deze epistemische mogelijkheid komt overeen met de logische mogelijkheid. Omdat het logisch mogelijk was voor Mary dat de fenomenale ervaringen anders aan de hersentoestanden waren gekoppeld dan ze feitelijk zijn was dit voor haar een epistemische mogelijkheid die slechts door het doen van waarnemingen bepaald kon worden.

17 Dit betekent dus niet dat voor functionalistisch equivalente organismen aan Mary groen er hetzelfde uit moet zien, zolang de kwalitatieve eigenschappen maar eenzelfde functionele rol vervullen (zie ook Churchland, 1985). Alleen voor Mary zelf is er slechts één nomologische mogelijkheid.

hadden kunnen zijn, was immers slechts een epistemische of logische mogelijkheid), maar alleen een nieuwe toegang tot diezelfde, fysische informatie. De Mary-intuïtie kan dus behouden worden.

In deel V bekijk ik Dennetts materialistische alternatief voor de anti-fysicalistische qualia: een verwerping van de Mary-intuïtie. Maar om dit alternatief voor onze fenomenale ervaringen goed te kunnen beoordelen geef ik eerst een korte schets van Dennetts visie op bewustzijn in het volgende deel.





# DENNETT'S METAFORENOORLOG: EEN FICTIONALISERING VAN HET BEWUSTZIJN

*Dat is alleen maar een metaforenoorlog, hoor ik u zeggen, maar metaforen zijn niet alleen maar metaforen. Metaforen zijn het gereedschap van denken. En aangezien niemand over het bewustzijn na kan denken zonder metaforen, is het belangrijk om over het beste gereedschap te beschikken. En zie eens wat we met ons gereedschap hebben opgebouwd. Had u zich dat kunnen voorstellen zonder dat gereedschap?*

DANIEL C. DENNETT (1991/1999, p. 499)

In *Het Bewustzijn Verklaard* geeft Dennett (1991) ons een aantal nieuwe manieren om over bewustzijn na te denken. Het blijkt echter lastig te zijn om uit deze metaforen een duidelijke theorie over bewustzijn af te leiden, omdat het boek bestaat uit veel verschillende, zelfs tegenstelde ideeën en onduidelijke voorstellen (Block, 1993; McGinn, 1995; Kirk, 1993). In de eerste drie paragrafen van dit deel geef ik een korte weergave van Dennetts belangrijkste metaforen om hier in de vierde paragraaf een conclusie aan te verbinden die als basis kan dienen om Dennetts visie op het probleem van de qualia en het fenomenale bewustzijn te beschouwen in het volgende deel. In dit deel geef ik geen kritiek op Dennetts metaforen en visie op het bewustzijn, maar bekijk ik alleen van waaruit hij vertrekt wanneer hij het probleem van qualia benadert.

## 4.1 Heterofenomenologie: een verschuiving van het explanandum

In de fenomenologische methode wordt slechts uitgegaan van het eerste-persoonsperspectief, waarbij de wereld buiten het subject als het ware tussen haken gezet wordt, i.e. de *epoché* van Husserl. Maar het subject zelf kan zich volgens Dennett (1991) vergissen over zijn eigen mentale toestand: “*U bent niet gezaghebbend over wat er in u gebeurt, maar wel over wat er in u lijkt te gebeuren.*”<sup>18</sup> (Dennett, 1991/1999, p. 114). Elke uitspraak over een innerlijke wereld is een uitkomst van subjectieve ad-hoc theorieën gebaseerd op de volkspychologie (*folk psychology*) die een culturele constructie is. De innerlijke wereld *zelf* kan niet bestudeerd worden vanuit het derde persoonsperspectief. Een wetenschappelijke methode die onderzoek doet naar het bewustzijn van mensen moet daarom in tegenstelling tot de fenomenologie niet de wereld, maar de interne toestanden die aanleiding geven tot observeerbaar en meetbaar (verbaal) gedrag van mensen tussen haken zetten. Dit is de heterofenomenologie, waarbij de bruikbare elementen van Husserls theorie zijn gebruikt<sup>19</sup> (Dennett, 2007).

De intentionele objecten van het bewustzijn worden in deze methode fictionaliseerd: “*Wat zijn dat voor voorwerpen en waar zijn ze van gemaakt? [...] 'Niets!' Waarvan is de heer Bommel gemaakt? Nergens van. Heer Bommel is een fictief voorwerp, net als de voorwerpen die door heterofenomenologen beschreven, benoemd en genoemd worden.*” (Dennett, 1991/1999, p. 113). De manifeste verschijningsvorm van het bewustzijn aan het subject zelf wordt door Dennett niet als een ontologische indicator beschouwd, dit

---

18 In deel VI kom ik terug op deze cryptische uitspraak van Dennett.

19 Dennetts interpretatie van Husserl is (natuurlijk) niet onbetwist.

Helaas valt het buiten het bereik van deze scriptie om hier dieper op in te gaan. Voor een verwerping van zowel Dennetts methode en als zijn conclusies vanuit een husserliaans perspectief zie bijvoorbeeld Vallor (2008). Voor een verdediging van zijn Husserl-interpretatie zie Dennett (1996).

is immers slechts de manifeste ontologie van het subject zelf. De wetenschappelijke verschijningsvorm kan hiervan verschillen en is gebaseerd op de wetenschappelijke ontologie (Dennett, 2007).

De heterofenomenoloog verschilt van de *lone wolf* autofenomenoloog (Dennett, 2007) in zijn aannames met betrekking tot de mentale toestand van het subject dat op de onderzoekstafel ligt. Hij is alleen verplicht om het bestaan van overtuigingen aan te nemen, maar blijft agnostisch met betrekking tot de grond voor deze overtuigingen (Dennett, 2003). De fenomenologie lijkt er alleen maar te zijn, maar dit betekent niet dat hij er ook daadwerkelijk is. De explananda van de heterofenomenologie zijn dus niet de intentionele, bewuste ervaringen van de klassieke fenomenoloog, maar het gedrag en de uitspraken van proefpersonen zoals die vanuit de derde persoon kunnen worden opgetekend.

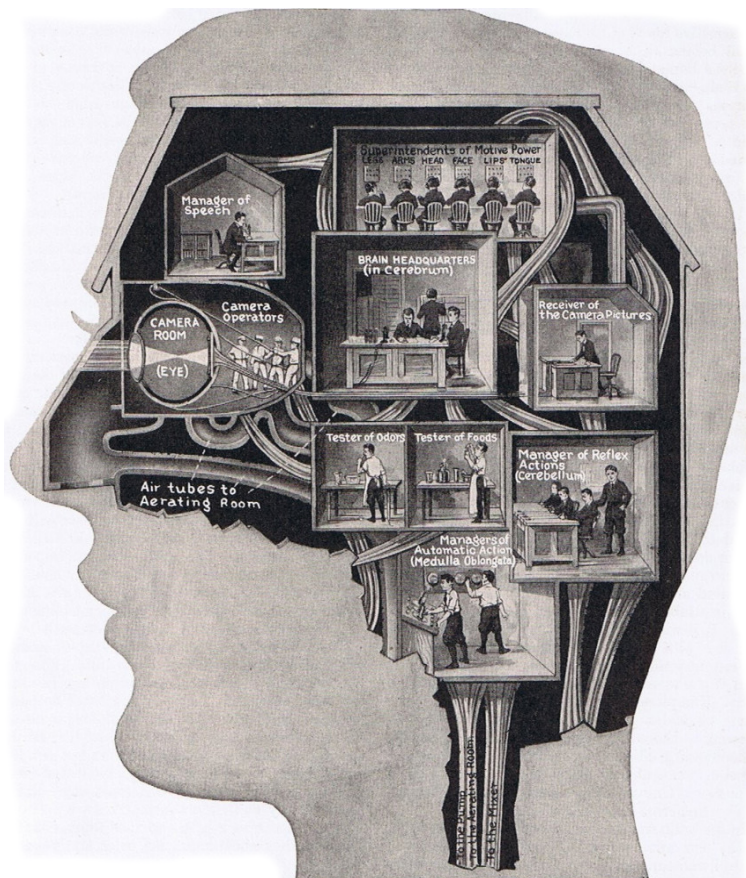
## 4.2 Van Cartesiaans Theater naar meervoudige conceptversies<sup>20</sup>

Een voorbeeld van de vergissingen die mensen maken als ze hun blik naar binnen richten is volgens Dennett de aanname van een Cartesiaans Theater. Het Cartesiaans Materialisme dat hieraan ten grondslag richt is de theorie van Descartes zonder het dualisme en komt voort uit Descartes' conceptie van de pijnappelklier als schakelpunt tussen het materiële lichaam en de immateriële ziel, waarbij slechts die processen die de pijnappelklier bereiken tot het bewustzijn van de immateriële ziel kunnen geraken (Dennett, 1991). Het Cartesiaans Theater is dan de plek waar

---

20 Hoewel ik in zijn algemeenheid de vertaling van Maas en Smeets (Dennett, 1991/1999) volg vind ik de weergave van *Multiple Drafts* als meervoudige versies niet afdoende. De *drafts* hebben een expliciete verwijzing naar een schrijfproces in zich, waarbij gesproken wordt over een nog niet afgeronde versie. Er bestaan dus niet meerdere volledige versies naast elkaar, maar meerdere concepten waaruit steeds één versie geconstrueerd wordt.

het bewustzijn zetelt, de centrale bestuurskamer zagezegd. Er zijn geen hedendaagse filosofen die letterlijk een dergelijk Cartesiaans Theater aannemen in hun theorieën stelt Dennett, maar het speelt op de achtergrond een belangrijke rol over hoe zowel leken als filosofen denken over bewustzijn.



Figuur 5: een Cartesiaans geïnspireerde weergave van wat er gebeurt in de hersenen, met het Cartesiaans Theater als Brain Headquarters.

In plaats daarvan stelt Dennett (1991/1999, p. 148) dat “het perspectief van de waarnemer in zijn hersenen zowel in ruimte als in tijd is uitgesmeerd”, en bestaat uit een pandemonium van conceptversies van narratieven die gelijktijdig verschillende mogelijke delen van verhalen vertellen. Door op verschillende momenten en plaatsen te peilen, ontstaat uit deze meervoudige conceptversies één verhaal dat gerapporteerd wordt aan de buitenwereld. Dit is echter geen definitieve versie, maar kan continu veranderen afhankelijk van welke conceptversies gepeild worden. Hoe dit proces vervolgens bewust wordt, kunnen we volgens Dennett begrijpen aan de hand van de Joyceaanse machine.

### 4.3 Bewustzijn als virtuele machine

*De enige toegang die we hebben tot wat er in ons brein gebeurt, heeft een sequentiële gedaante die sprekend lijkt op de Von Neumann-architectuur.*

Dennett (1991/1999, p. 241)

Het menselijke bewustzijn is een virtuele machine die gevormd wordt door cultureel diep ingesleten 'mentale gewoontes'. Deze gewoontes worden 'geïnstalleerd' in de jeugd en kunnen worden begrepen als een memencomplex<sup>21</sup>. De virtuele machine vormt een *stream of consciousness* en wordt daarom door Dennett (1991) ook wel een joyceaanse machine genoemd. Doordat de joyceaanse machine een serieel proces is dat bovenop de parallelle meervoudige conceptversie-architectuur van de hersenen draait, brengt het rust in het pandemonium van gelijktijdige processen

21 Memen vormen op dezelfde wijze waarop genen de drager van biologische evolutie zijn de drager voor culturele evolutie. Memen zijn hierbij zichzelf vermenigvuldigende eenheden van informatie die tussen mensen kunnen worden uitgewisseld. Zo kunnen evolutionaire methodes worden ingezet om de ontwikkeling van cultuur te beschouwen. Het idee van memen komt van Dawkins (1976).

die niet congruent zijn, en maakt het mensen mogelijk om hun aandacht te richten en problemen op te lossen. De joyceaanse virtuele machine is volgens Dennett bovendien een voldoende en noodzakelijke voorwaarde voor bewustzijn.

De metafoor van de virtuele machine heeft een belangrijke uitwerking. Net zoals het vanuit de verschijning van een tekstverwerker op het beeldscherm niet mogelijk is om af te leiden hoe de onderliggende hardware is opgebouwd (Microsoft Word draait immers net zo goed op een Mac als op een PC) is voor het subject ook de innerlijke werking van zijn eigen hersenen ontoegankelijk.

#### **4.4 Het bewustzijn als linguïstische machine: de narratieve fictionalisering**

Dennetts nieuwe metaforen lijken te wijzen op een sterk taalkundig georiënteerde visie op bewustzijn. Als explanandum hebben we slechts de uitspraken en overig gedrag van mensen, waarbij we (in ieder geval tijdens de data-verzamingsfase) agnostisch blijven over de achterliggende processen. Het agnosticisme van de heterofenomenologie maakt in Dennetts denken echter al snel plaats voor een fictionalisering van het bewustzijn waarbij onze bevindingen ten aanzien van wat er in ons bewustzijn gebeurt niet langer verwijzen naar mentale objecten, maar slechts verhalen zijn, verzinssels zonder referentiële basis<sup>22</sup> in de realiteit van neurotransmitters, neuronen en hersenprocessen.

Dit fictionalisme komt voort uit het feit dat volgens Dennett het bewustzijn gezien moet worden als een virtuele joyceaanse machine die op seriële wijze de parallelle meervoudige narratieve conceptversies peilt en op basis hiervan

---

22 Dennett ontkent natuurlijk niet dat het gehele bewustzijn uiteindelijk terug te voeren is op de hersenen. Hij meent alleen dat de objecten waaraan gerefereerd lijkt te worden in het bewustzijn geen basis hebben in het brein. Dit is dus in directe tegenspraak met Churchland die stelt dat in het geval van kleur de referent de representatie is hersengebied V4. Hierover meer in deel VI.

een semi-definitief verhaal samenstelt dat verbaal gerapporteerd kan worden. Omdat het bewustzijn grotendeels een culturele uitvinding is die niet in de genetische maar pas in de memetische evolutionaire processen gevormd is, wordt de functie van taal extra benadrukt en het schisma tussen de 'echte' hersenprocessen en 'fictieve' mentale objecten verder vergroot. Onze taal maakt het ons dus niet mogelijk om aan elkaar te vertellen wat in ons voortalige brein omgaat, omdat we altijd vertellen over (schijnbare) mentale objecten die gelijktijdig met de taal evolueerden en dus nog niet bestonden in ons voortalige brein. Na het verlaten van het Cartesiaanse Theater moeten we volgens Dennett concluderen dat er geen eerste-persoonsperspectief is los van de verhalen die we hierover vertellen.

De vraag is in hoeverre er ruimte is voor het fenomenale bewustzijn in deze gefictionaliseerde linguïstische machine. Voordat ik deze vraag bespreek in deel V ga ik eerst in op Dennetts gebruik van Wittgenstein, een belangrijke inspiratiebron voor zijn eliminativistische visie op qualia.



Figuur 6: Ludwig Wittgenstein



# EEN BEROEP OP WITTGENSTEIN: TAAL SPELEN VERSUS ONTOLOGIE

*Now someone tells me that he knows what pain is only from his own case! – Suppose everyone had a box with something in it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. – Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. – But suppose the word "beetle" had a use in these people's language? – If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a something: for the box might even be empty. – No, one can 'divide through' by the thing in the box; it cancels out, whatever it is.*

LUDWIG WITTGENSTEIN (1953, §293)

Met bovenstaand citaat doet Dennett in het hoofdstuk *Qualia Gediskwalificeerd* een beroep op de late Wittgensteins demystificatie van bewustzijn. Beide denkers proberen een alternatief te vinden voor de reïficatie van mentale objecten. Dennett stelt hierbij dat zijn filosofische onderneming gezien kan worden “als een soort herhaling van Wittgensteins aanval op de 'objecten' van de bewuste ervaring” (Dennett, 1991/1999, p. 527). Hij vertelt dat hij in zijn tijd in Oxford, net als veel studenten in die tijd, Wittgenstein als grote voorbeeld had. Omdat veel van zijn medestuden-

ten volgens Dennett niets begrepen van Wittgensteins werk besloot Dennett geen wittgensteiniaan te worden, maar dat wat hij geleerd had uit de *Philosophische Untersuchungen* toe te passen. Dit zou in het bijzonder moeten gelden voor zijn eliminativisme met betrekking tot qualia.

Maar de vraag is of Dennetts filosofie wel een radicalere versie is van Wittgenstein. Na bovenstaand citaat van Wittgenstein gegeven te hebben, stelt Dennett: “*Maar wat betekent het? Betekent het dat qualia echt zijn maar geen uitwerking hebben? Of dat er helemaal geen qualia zijn?*” (Dennett, 1991/1999, p. 428). Wittgenstein lijkt echter alleen geïnteresseerd in het taalspel, en niet in de ontologische implicaties of de fenomenen zelf. Anders dan het revisionistisch eliminativisme van Dennett, wil Wittgenstein de zaken laten zoals ze zijn (Hutto, 1995). Om te komen tot een oplossing voor het filosofische probleem aangaande de objecten van het bewustzijn wil Wittgenstein geen onderzoek doen naar deze objecten zelf, maar naar de uitspraken die we hierover doen. Dennett daarentegen is vooral geïnteresseerd in de eigenschappen van de vermeende objecten van ervaringen en probeert op basis hiervan het bestaan van deze ervaringen te ontkennen. Beide denkers gebruiken dus een andere methode (Mason, 2005).

Wittgenstein en Dennett komen wel overeen in de mening dat termen die we gebruiken om private fenomenen als ervaringen en pijn te verwoorden, nergens naar verwijzen. Met andere woorden, er is geen mentaal object waaraan gerefereerd wordt. Maar Wittgensteins positie lijkt een stuk genuanceerder dan die van Dennett. Hoewel Wittgenstein de geest niet langer als een object wil beschouwen wil hij ons spreken over bewustzijn wel degelijk serieus nemen (Hutto, 1995). Om te zien hoe Wittgenstein deze nuance aanbrengt moeten we eerst zijn visie op betekenis door gebruik beschouwen.

Wittgenstein stelt dat woorden hun betekenis altijd krijgen in het gebruik: simpele ostentieve definities kunnen

immers altijd anders worden geïnterpreteerd. Maar om een woord correct te kunnen gebruiken moet er een publieke norm zijn die duidelijk maakt wanneer het woord correct gebruikt wordt. De betekenis van taal wordt volgens Wittgenstein dus geconstitueerd in het gebruik ervan. Maar met betrekking tot private fenomenen kan een dergelijke betekenis niet worden geconstrueerd. Er is immers geen publieke norm waartegen het gebruik kan worden afgemeten (Mason, 2005). Het volgende citaat maakt dit duidelijk:

*'I know how the colour green looks to me' – surely that makes sense! – Certainly: what use of the proposition are you thinking of? [...] Imagine someone saying: 'But I know how tall I am!' and putting his hand on the top of his head to prove it.*

Wittgenstein (1953, §278-9)

Door je hand op je hoofd te leggen, weet je natuurlijk nog niet hoe groot je bent ten opzichte van anderen. Diezelfde onmogelijkheid is er wanneer we onze innerlijke, private ervaringen als criterium nemen voor onze woorden. Wittgensteins oplossing op dit dilemma kan wellicht het beste beschreven worden aan de hand van pijn. “*A child has hurt himself and cries: and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain behaviour.*” (Wittgenstein, 1953, §244). Wanneer iemand zegt “Ik heb pijn” is er in de visie van Wittgenstein dus geen object pijn waarvan het subject zich gewaar wordt en waarvan hij melding maakt met zijn uitspraak. In plaats daarvan is de uitdrukking een aangeleerde expressie van een sensatie.

Het wordt nu dan ook duidelijk hoe we het kevervoorbeeld van Wittgenstein kunnen begrijpen. De laatste zin van dezelfde paragraaf licht toe wat Wittgenstein bedoelt: “*That is to say: if we construe the grammar of the expression of sensation on the model of 'object and designation' the*

*object drops out of consideration as irrelevant.*” (Wittgenstein, 1953, §293). De onmogelijkheid onze 'kevers' te vergelijken betekent dat ze geen rol kunnen vervullen in het taalspel. Maar het betekent niet dat de sensatie “wat het voor mij is om groen te zien” als zodanig niet bestaat, maar slechts dat een dergelijke uitspraak niet zinnig gebruikt kan worden in een gesprek met een ander.

We kunnen nu ook zien waar de posities van Wittgenstein en Dennett van elkaar verschillen. Voor Wittgenstein bestaan de sensaties die ten grondslag liggen aan onze uitspraken hierover wel degelijk. Onze uitspraken hierover refereren echter niet aan een mentaal object dat hier ten grondslag aan ligt, maar zijn slechts expressies van onze psychologische situatie (Hutto, 1995). Wittgensteins conclusie bevindt zich dus op het niveau van de taal: we kunnen niets zinnigs zeggen over onze sensaties omdat er geen publieke standaard gemaakt kan worden op basis waarvan we het gebruik van dergelijke woorden kunnen vaststellen. Ons taalgebruik aangaande onze sensaties van kleur is niet referentieel, het verwijst niet naar een quale, maar is expressief, het vertelt de toehoorder iets over de toestand waarin we ons bevinden (Hutto, 1995; Mason, 2005). Dennetts conclusie gaat echter een stuk verder en bevindt zich op het ontologische vlak: omdat onze woorden met betrekking tot sensaties niet naar een mentaal object verwijzen bestaan de sensaties zelf ook niet.

Maar hoe ziet Dennett onze waarneming dan? Om te zien wat Dennett voorstelt als vervanging voor de ge-elimineerde qualia bekijk ik in het volgende hoofdstuk zijn reactie op 'Mary'.

# DENNETT'S ELIMINATIVISME

*There is no phenomenological manifold ... there is – so far as introspection concerned – darkness.*

DANIEL C. DENNETT (1979, p. 95)

Er zijn geen qualia, geen onuitspreekbare, private intrinsieke eigenschappen die direct toegankelijk zijn voor het bewustzijn volgens Dennett. In dit deel bekijk ik zijn visie op fenomenale ervaringen in termen van zijn analyse van 'Mary'.

## 5.1 Reactieve disposities en bierdrinkers

Het voorbeeld van een *acquired taste*<sup>23</sup> maakt volgens Dennett (1991) duidelijk waarom er geen qualia kunnen zijn. Stel, er zijn twee ervaren bierdrinkers die de smaak van een Duveltje wel kunnen waarderen. Maar beide zijn het er ook over eens dat de eerste keer dat ze een slok bier namen ze het veel minder lekker vonden. Bierdrinker A meent dat de smaak van bier voor hem veranderd is. Bier smaakt voor hem niet langer als de eerste slok die hij nam, want “niemand zou die smaak lekker vinden”. Hij meent dus dat een herhaalde blootstelling aan bier ertoe geleid heeft dat hij een andere quale heeft wanneer hij bier proeft. Bierdrinker B daarentegen meent dat de smaak onveranderd is voor hem, maar dat de herhaalde blootstelling ertoe geleid heeft dat hij deze smaak is gaan waarderen.

Bij de bierdrinker A is de 'vorm' van de kwaliteitenruimte veranderd, bij bierdrinker B is de kwaliteitenruimte ge-

<sup>23</sup> In zijn hoofdstuk *Qualia Gediskwalificeerd* geeft Dennett (1991/1999) nog enkele voorbeelden, zoals een nauwgezette analyse van het geïnverteerde spectrum argument. Omdat de andere voorbeelden tot dezelfde conclusie leiden volsta ik hier met bespreking het meest aansprekende.

lijk gebleven, maar is zijn evaluatie van deze ruimte veranderd. Dit voorbeeld toont aan dat we onze ruwe sensorische informatie, de veronderstelde qualia, nooit los kunnen beschouwen van onze reacties hierop. Er zijn dan volgens Dennett geen intrinsieke eigenschappen van ervaringen die direct toegankelijk zijn voor het bewustzijn op een wijze niet niet wordt beïnvloed door de reactieve disposities. Als de bierdrinkers proberen om de ervaring van de smaak van de eerste slok bier te vergelijken met de smaak van bier voor hen *op dit moment* is het conceptuele onderscheid tussen de smaak zelf en de reactie hierop niet te maken.

Volgens Dennett moet 'hoe het smaakt' dan gereduceerd worden tot het complex van reactieve disposities<sup>24</sup> (Dennett, 1988). “*We zouden de qualia moeten 'vernietigen' om ze te 'redden'.*” (Dennett, 1991, p. 396 mijn vertaling<sup>25</sup>). *Wat het is om een ervaring te hebben* kan volgens Dennett (1988; 1991) dus gereduceerd worden tot de conjunctie van reactieve disposities die een dergelijke ervaring kenmerkt. De bijbehorende fenomenale ervaring wordt ofwel glashard ontkend, ofwel gefictionaliseerd<sup>26</sup> in de joyceaanse machine. Wanneer iemand dus de uitspraak doet “wat het voor mij is om rood te zien” zou Dennett dit reduceren tot die

---

24 Dennett (1988; 1991) meent dat er geen *fact of the matter* is dat onderscheid kan maken tussen beide theorieën met betrekking tot de *acquired taste* wanneer we het Cartesiaanse Theater hebben verlaten. Hij treedt echter nergens in nader detail waarom dit betekent dat de quale slechts de reactieve disposities is.

25 De officiële Nederlandse vertaling (Dennett, 1991/1999, p. 434) is hier erg ongelukkig: “*Om de reactieve disposities te kunnen 'redden', zouden we de qualia moeten 'vernietigen'.*”

26 Het onderscheid dat normaal aangebracht wordt tussen *wat het is om een ervaring te hebben* en de reactieve disposities is dus precies het onderscheid dat Dennett ontkent. Het is niet zo dat je eerst een ervaring hebt en als gevolg van die ervaring bepaalde reacties hebt: “[*There is no*] *distinction [...] between knowing 'what one would say and how one would react' and 'knowing what it is like'. If there is such a distinction, it has not yet been articulated and defended, by [anyone], so far as I know*” (Dennett, 2006, voetnoot 3)

uitspraak zelf, samen met alle affectieve en gedragsmatige reacties, geassocieerde herinneringen en alle andere disposities die het *in een staat van rood-zien* zijn met zich meebrengt. Een dergelijke visie op qualia leidt tot een contra-intuïtief resultaat van 'Mary' zoals we in de volgende paragraaf zullen zien.

## 5.2 De blauwe banaan

In *Het bewustzijn verklaard* geeft Dennett een eerste antwoord op Mary. Hij stelt dat het gedachte-experiment van Jackson een verkeerde intuïtiepomp is, omdat het ons iets laat doen wat wij helemaal niet kunnen. Jackson stelt dat Mary over *alle* materialistische informatie beschikt maar dit kunnen wij ons volgens Dennett niet voorstellen. In plaats daarvan stellen we ons iets anders voor, namelijk dat Mary heel veel materialistische informatie heeft over kleurervaring. Hij stelt dan ook het volgende, verrassende, alternatief voor:

*En op een dag besloten de mannen die Mary gevangen hielden dat het hoog tijd was dat ze kleuren zou zien. Als grapje prepareerden ze een blauwe banaan, die haar allereerste ervaring met kleur zou zijn. Mary wierp er één blik op en zei toen: 'Hé! Jullie proberen me voor de gek te houden! Bananen zijn geel maar deze is blauw!' De mannen stonden met de mond vol tanden. Hoe wist ze dat? 'Heel eenvoudig,' antwoordde Mary. 'Zijn jullie vergeten dat alles – echt alles – weet wat er maar te weten valt over de materiële oorzaken en gevolgen van de kleurwaarneming? En dus had ik, nog voordat jullie met die banaan op de proppen kwamen, al uiterst gedetailleerd opgeschreven wat voor indruk een geel (of een blauw of een groen) voorwerp zou maken op mijn zenuwstelsel. Zodoende wist ik precies wat voor gedachten ik zou hebben (want louter de dispo-*

*sitie om over iets na te denken, behoort toch niet tot jullie qualia, is het wel?). Ik was niet in het minst verrast door mijn ervaring van blauw – het verbaasde me alleen dat jullie zo'n stom grapje wilden uithalen. Ik snap wel dat jullie je moeilijk voor kunnen stellen dat ik zo veel weet over mijn reactieve disposities dat ik niet verrast werd door wat blauw met me deed. Natuurlijk is dat moeilijk voor jullie. Niemand kan zich gemakkelijk voorstellen wat de gevolgen zijn als iemand echt alles weet van de materiële aspecten ergens van.*

Dennett (1991/1999, p 438)

In haar gevangenschap kan Mary dus al voorspellen welke reacties ze zal hebben als ze een bepaalde kleur zal zien. Uit alle voor haar beschikbare informatie moet ze een manier vinden om van 'binnenuit' haar reacties op het zien van de verschillende kleuren te herkennen. Maar heeft ze nu ook afgeleid wat het voor haar zal zijn om de kleur te zien?

### **5.3 EEG-Mary**

Stel dat Mary het zichzelf gemakkelijk wil maken. Op basis van haar informatie over de opbouw van het brein zal ze gemakkelijk kunnen afleiden welke gebieden in haar visuele cortex actief zullen zijn bij het zien van een bepaalde kleur. Deze activatie kan ze vervolgens meten met een gedetailleerde EEG-scan. In theorie is het dus mogelijk om van 'buitenaf' bepaalde neurofysiologische reacties te herkennen die duiden op het waarnemen van specifieke kleuren. We kunnen ons eenvoudig voorstellen dat Mary een apparaat maakt dat op basis van de via EEG gemeten activatie haar via een beeldschermje laat zien welke kleur ze op dat moment ziet.



Deze situatie is echter precies vergelijkbaar met de oplossing van Dennett, in de zin dat Mary slechts de verwachte effecten van haar waarneming van kleur monitort. Dennett moet dus een sterkere claim maken. Dit kunnen we ook afleiden uit het feit dat Dennetts Mary niet verbaasd is wanneer ze voor het eerst de blauwe banaan ziet. Niet alleen moet Mary haar reactieve disposities op het zien van kleuren kunnen afleiden, wat het voor haar is om die kleur te zien moet *gelijk zijn* aan de conjunctie van deze reactieve disposities.

#### **5.4 Twee verhalen, maar waar zijn de argumenten?**

Wanneer we de twee mogelijke uitkomsten van het gedachte-experiment naast elkaar houden – de verbazing die Jackson, Lewis en Churchland veronderstellen versus de reactie van Dennetts Mary – hebben we twee onbeargumenteerde intuïties naast elkaar, die naar voren worden gebracht door de specifieke situatie waarin Mary zich bevindt. Dennett heeft echter wel de intuïties van de meeste mensen tegen zich, en legt met zijn voorbeeld van de blauwe banaan nog niet adequaat uit op welke wijze Mary kan bedenken hoe het voor haar moet zijn om een kleur te zien.

In een later artikel van zijn hand (Dennett, 2006) doet hij een nieuwe poging om zijn visie uit te leggen aan de hand van RoboMary. Maar voordat ik in de verdere gedachte-experimenten duik die Dennett gebruikt om zijn theorie te onderbouwen wil ik eerst een aantal spelregels voor 'Mary' behandelen.

#### **5.5 Spelregels voor 'Mary'**

Het is van belang om vast te stellen wanneer Mary vals speelt. Lewis (1988) stelt bijvoorbeeld dat, hoewel ervaring de enige bron is van fenomenale kennis die bestaat uit vaardigheden, deze ervaring een bepaald spoor na zal laten in haar brein. In theorie zou het dus mogelijk zijn om Mary in haar gevangenschap de herinnering aan een kleur te ge-

ven door een zeer gecompliceerde neurochirurgische operatie uit te voeren, die haar in de toestand van de Mary brengt die wel een kleur heeft gezien. Maar ook in de visie van Churchland (1989) zou Mary op deze wijze vals kunnen spelen. Door welbepaalde delen van haar visuele cortex te stimuleren met bijvoorbeeld een zeer nauwkeurige vorm van transcraniële magnetische stimulatie zou Mary ook de ervaring van een kleur kunnen hebben, zonder dat ze daadwerkelijk een kleur heeft gezien.

Dergelijke voorbeelden kunnen worden gezien als bewijs tegen de stelling dat ervaring de *enige* bron is van fenomenale kennis, precies de reden waarom Lewis en Churchland menen dat Mary iets leert wanneer ze voor het eerst een kleur ziet. De hiervoor genoemde chirurgische ingrepen – ofwel het aanbrengen van een fictieve herinnering – en directe breinstimulatie – het oproepen van een hallucinatoire ervaringstoestand – zijn immers geen voorbeelden van een dergelijke reële ervaring. In het geval van Mary zal ik dergelijke oplossingen beschouwen als vals spelen, omdat Mary hier niet op basis van alle fysische informatie de ervaring van kleur afleidt, maar deze op een niet-normale wijze tot stand brengt. Anders dan een verklaring van de mentale toestand van rood-zien is dit immers slechts een instantiatie van een dergelijke toestand of een herinnering hieraan op niet-normale wijze. Deze voorbeelden zijn bovendien eenvoudig uit te sluiten door Mary geen toegang te verlenen tot hersenchirurgische apparatuur e.d.

Er is echter nog een vreemdsoortiger alternatief, Moeras-Mary<sup>27</sup> (Dennett, 2006). Stel dat Mary, enkele seconden voordat ze voor het eerst de rode roos zal zien, getroffen wordt door een bliksem, die, bij wijze van kosmisch wonder, haar brein in de toestand brengt waarin ze gekomen zou zijn indien ze de rode roos net gezien had. Wanneer ze nu voor het eerst de rode roos ziet zal het voor haar lijken dat het eigenlijk de tweede keer is, en zal ze dus ook niet verbaasd zijn. Ook hier is echter op niet-normale wijze de

---

27 Dennett attribueert het idee van 'SwampMary' aan Gabriel Love.

toestand bereikt, en dus zal dit voorbeeld ook als vals spelen moeten gelden. Bij de behandeling van Dennetts RoboMary zullen we dus goed moeten opletten dat Dennett Mary niet vals laat spelen, door haar op één of andere wijze die niet wordt toegestaan door de uitgangspunten van het originele gedachte-experiment direct in een toestand van fenomenale ervaring te brengen.

## 5.6 RoboMary

Dennett (2006) constateert dus dat er in de literatuur een tweetal mogelijkheden te vinden zijn waarop Mary te weten kan komen wat het is om een kleur te zien. De eerste is simpelweg geconfronteerd worden met een rood object. De tweede is een vorm van kosmisch wonder of specialistische neurochirurgie zoals hiervoor beschreven. Maar is er niet een derde vorm mogelijk, één die wel in staat is om de Mary-intuïtie te weerleggen? Volgens Dennett wel, en hij introduceert hiervoor het idee van RoboMary, een robotische versie van Mary die in exact dezelfde situatie als Mary verkeert. Hieronder de introductie van RoboMary door Dennett (2006, p. 25-26):

*1. RoboMary is a standard Mark 19 robot, except that she was brought online without color vision: her video cameras are black and white, but everything else in her hardware is equipped for color vision, which is standard in the Mark 19. [...]*

*2. While waiting for a pair of color cameras to replace her black-and-white cameras, RoboMary learns everything she can about the color vision of Mark 19s. She even brings colored objects into her prison cell along with normally color-sighted Mark 19s and compares their responses – internal and external – to hers. [...]*

3. *She learns all about the million-shade color-coding system that all Mark 19s have. [...]*

4. *Using her vast knowledge, she writes some code that enables her to colorize the input from her black-and-white cameras [...] according to luminous data she gathers about what colors things in the world are. and how Mark 19s normally encode these. So now when she looks with her black-and-white cameras at a ripe banana, she can first see it in black and white, as pale gray, and then imagine it as yellow (or any other color) by just engaging her colorizing prosthesis.*

In het licht van het functionalisme is de eerste premisse volledig te accepteren, zolang de Mark 19 in functioneel opzicht gelijk is aan een mens. De tweede en derde premisse zijn in lijn met de gewone Mary. Het probleem komt in de vierde premisse waarin Dennett RoboMary in staat stelt om direct haar input in te kleuren. Zij maakt hierbij gebruik van de kleuren die ze nu juist moet afleiden uit fysische informatie. Deze oplossing lijkt daarom meer op de directe breinstimulatie zoals in de voorgaande paragraaf besproken: Mary activeert simpelweg van binnenuit haar kleuren-waarneming. Dennett realiseert zich dit ook, en stelt daarom een tweede vorm voor: *locked* RoboMary.

De voorwaarden voor dit nieuwe gedachte-experiment zijn hetzelfde. Daarnaast is het *locked* RoboMary echter onmogelijk gemaakt om haar kleurenregisters direct te benaderen. Wanneer ze dus een rode roos ziet, krijgt ze slechts een beeld in grijstinten. De toestand waarin RoboMary is op het moment waarin ze een grijs beeld ziet van de rode roos noemen we toestand A. Ze kan het beeld dat ze heeft van de roos nog steeds inkleuren op basis van alle informatie die ze heeft over kleuren en kleureervaringen, maar omdat ze haar kleurenregister niet meer kan manipuleren, kan ze de kleuren niet direct ervaren.

Maar ook hiervoor heeft Dennett een oplossing: *locked* RoboMary maakt een simulatie van zichzelf. Deze simulatie verschilt echter op één punt van haarzelf: de simulatie heeft een werkend kleurenregister<sup>28</sup>. Het door RoboMary ingekleurde zwart/wit beeld is de input voor de simulatie. De simulatie heeft nu een waarachtige ervaring van het rood van de roos<sup>29</sup>, die we toestand B noemen.

RoboMary kan nu de verschillen tussen toestand A en toestand B uitrekenen, en op basis hiervan zichzelf in toestand B brengen. Volgens Dennett heeft RoboMary nu afgeleid wat het is om de rode roos te zien, zonder dat ze direct haar kleurenregisters heeft geactiveerd, en dus zonder vals te spelen. Ze brengt zichzelf immers niet in de waarnemingstoestand, omdat ze de kleurregisters zelf niet manipuleert, maar in de toestand die volgt uit de waarneming.

## 5.7 Dennetts herinterpretatie van de logische structuur van 'Mary'

Dennett lijkt met zijn RoboMary toe te geven aan de Mary-intuïtie. Wanneer we immers aan RoboMary vragen of ze weet wat het is om rood te zien zal ze het volgende moeten antwoorden: *“I don't know, but I can work it out. Hold on a minute [or a second, or a picosecond]... Ok, there we are! Now I know”* (Beaton, 2005, p. 11). Met andere woorden, ze moet werk verzetten om te komen tot de ervaring van blauw, los van alle fysische kennis die ze heeft. Laten we de regimentatie van 'Mary' die ik in deel III heb gebruikt nog eens in herinnering halen:

---

28 Hoewel *locked* RoboMary zelf geen werkend kleurenregister heeft kan ze op basis van alle informatie over kleurenervaring natuurlijk eenvoudig een kleurenregister simuleren.

29 De simulatie is immers functioneel isomorf aan RoboMary zelf. Op basis van de aanname van het functionalisme moeten we dus concluderen dat de gesimuleerde RoboMary een bewuste ervaring heeft van de rode roos op basis van de input die ze aangeleverd krijgt door RoboMary.

- |  |  |
|--|--|
| (i) $\forall x ( Fx \rightarrow Kmx )$ | Fx: x is fysisch;                            |
| (ii) $\exists x ( \neg Kmx )$          | Kyx: y weet dat x;                           |
| (iii) $\exists x ( \neg Fx )$          | m = Mary; en x zijn<br>alle kenbare objecten |

Dennetts *locked* RoboMary lijkt dus na het leren van alle fysische informatie, maar voor het draaien van de simulatie van zichzelf-die-rood-ervaart, iets nog niet te weten. Omdat Dennett de oplossing van Churchland en Lewis dat het in dit geval om andere kennis gaat afwijst<sup>30</sup>, kan hij niet stellen dat hetgeen RoboMary niet weet een andere vorm van kennis is. In plaats daarvan veronderstelt Dennetts behandeling van RoboMary dus dat het voor het behoud van het physicalisme noodzakelijk is dat een subject met het kennisniveau van Mary moet kunnen bedenken *wat het is om* iets te ervaren op basis van alle fysische kennis (Beaton, 2005):

- |  |  |
|--|--|
| (vi) $\forall x ( Kmx \rightarrow Amx )$ | Ayx: y kan afleiden wat<br>het is om x te ervaren. |
| (vii) $\forall x ( Fx \rightarrow Amx )$ | uit (i) en (vi)                                    |

En wanneer het voor RoboMary niet mogelijk zou zijn om af te leiden wat het is om rood te zien:

- |                               |                  |
|-------------------------------|------------------|
| (vi) $\exists x ( \neg Amx )$ |                  |
| (vii) $\exists x ( \neg Fx )$ | uit (i) en (vii) |

Dit betekent echter dat uit Dennetts RoboMary volgt dat elk voorbeeld van een subject dat niet kan afleiden wat het

---

<sup>30</sup> Dennett gaat nergens in zijn behandeling van Mary expliciet in op de mogelijkheid van meerdere vormen van kennis die niet tot elkaar te reduceren zijn. Hij verwijst in het hoofdstuk *Qualia Gediskwalificeerd* wel naar de artikelen van Churchland (1985, 1989) als één van de filosofen die enig tegenwicht lijkt te bieden tegen de verbaasde Mary, maar komt verder niet terug op de ideeën van Churchland of Lewis.

is om rood te ervaren tot een verwerping van het fysicalisme leidt. Dennett is dus niet klaar wanneer hij (vi) heeft verworpen, maar moet ook de veel sterkere claim maken dat elk subject met de kennis van Mary noodzakelijkerwijs de ervaring van groen (en blauw, en geel etc.) kan afleiden: (viii)  $\forall y[\forall x(Kmx) \rightarrow \neg\exists x(\neg Ayx)]$ . Elk tegenvoorbeeld op deze these bedreigt nu het fysicalisme van Dennett.

## 5.8 Terug naar de menselijke vermogens: RoboDennett

Een dergelijk tegenvoorbeeld denkt Beaton (2005) te hebben gevonden in RoboDennett. Net als RoboMary (en Mary zelf natuurlijk) heeft RoboDennett toegang tot alle fysische informatie. Maar in tegenstelling tot RoboMary heeft hij geen ongelimiteerde mogelijkheden om zijn eigen toestand te manipuleren. Om de situatie van Mary goed op waarde te schatten is het immers van belang om de dichtstbijzijnde mogelijke wereld te selecteren waarin Mary wel de vaardigheden heeft die het gedachte-experiment veronderstelt – het verkrijgen en vrijelijk bewerken van alle fysische informatie – zonder haar extra vaardigheden toe te kennen die niet verondersteld worden. Beaton geeft hierbij RoboDennett wel het vermogen om een simulatie van zichzelf te draaien, net als RoboMary kan. Wanneer we RoboDennett immers in staat stellen om *alle* fysische informatie te overzien, moet zijn intellect van zo'n andere aard zijn dan het onze dat het draaien van simulaties niet logisch uitgesloten kan worden.

Het cruciale verschil tussen RoboDennett en RoboMary is dat eerstgenoemde niet in staat is om zijn lage niveau visuele systeem<sup>31</sup> direct te manipuleren, hierbij stuiten we

31 Beaton bedoelt hiermee, in lijn met Churchland (1989), het hersengebied V4. Een directe manipulatie van dit gebied zou het voor mensen mogelijk maken om terwijl ze naar dingen kijken de kleur hiervan te veranderen, de specifieke activatie in V4 is immers verantwoordelijk voor welke kleur we waarnemen. Dit is evident een vermogen waar normale mensen niet over beschikken, en dat niet

immers op dezelfde bezwaren als bij de originele, niet *locked*, RoboMary. Ook is RoboDennett niet in staat een functioneel equivalent te simuleren van zijn lage niveau visuele systeem en deze simulatie in dezelfde causale relatie te plaatsen tot de rest van zijn brein als zijn echte lage niveau visuele systeem zou staan als het wel operationeel was. Dennett (2006) stelt namelijk dat indien een representatie van een laag niveau visueel systeem van belang mocht blijken bij wat het is om een kleur te ervaren, zijn RoboMary simpelweg dit visuele systeem kan simuleren. Beaton stelt echter dat het herstructureren van zijn eigen cognitieve opbouw waarin het gesimuleerde visuele systeem de causale rol van het echte, uitgeschakelde, visuele systeem overneemt, simpelweg buiten de mogelijkheden van RoboDennett ligt<sup>32</sup>. Als blijkt dat dit lage niveau visuele systeem een belangrijke rol speelt in wat het is om kleuren te ervaren, zoals bij Churchlands vertrouwdehypothese<sup>33</sup> het geval

---

voorondersteld wordt door 'Mary'.

- 32 Deze oplossing van Dennett betekent namelijk dat RoboDennett de input van zijn camera zou moeten aftappen en direct naar de simulatie zou moeten leiden. De uitkomst van de simulatie moet dan op dezelfde wijze verbonden moeten zijn met de rest van RoboDennetts brein als zijn normale kleurenregister zou zijn indien het zou werken. Dit betekent dat RoboDennett van 'binnenuit' grote wijzigingen zou moeten aanbrengen in zijn hersenarchitectuur. Dit vermogen ontzegt Beaton RoboDennett.
- 33 Beaton zelf introduceert een uitbreiding op Churchlands positie. Hij stelt de representatie in V4 los van de mogelijkheid van Mary om hier een uitspraak over te doen slechts een voorbeeld is van *blindsight*. In plaats daarvan stelt hij dat “*The state of knowing what it is like to perceive X is the state in which one's more abstract capacity to reason about X is supported by lower level sensory processing apparatus which discriminates and responds to X*” (Beaton, 2005, p. 9). Omdat Churchland (1989) expliciet stelt dat de representatie in V4 slechts de mogelijksvoorwaarde is voor de vaardigheden die kleurwaarnemende organismen hebben, zoals herkenning, discriminatie, verbeelding etc. en hiernaast de mogelijkheid van introspectie benadrukt, denk ik echter dat deze uitbreiding van Beaton – in ieder geval grotendeels – overlapt met de positie van Churchland. Voor het geval RoboDennett volstaat



is, kan RoboDennett dus niet afleiden wat het is om kleuren te zien uit alle beschikbare fysische informatie.

## 5.9 Een heterofenomenologische analyse van RoboDennett

Maar in Dennetts heterofenomenologie is een verschil dat geen verschil maakt, geen verschil (Dennett, 1991; Dennett, 2004). Dus waarom kan RoboDennett niet gewoon doen *alsof* hij ervaart wat het is om kleuren te zien, door simpelweg alle gedragsdisposities van zijn simulatie over te nemen (Beaton, 2005)? In deze afgezwakte vorm blijft de ervaring van RoboDennett buiten beschouwing en telt alleen zijn gedrag, precies in lijn met de heterofenomenologie van Dennett. Als vervolgens een verschil optreedt in gedrag tussen RoboDennett en Dennett zelf dan zal zelfs laatstgenoemde moeten toegeven dat RoboDennett niet weet wat het is om een kleur te zien.

Een tweetal verschillen zullen optreden in gedrag volgens Beaton (2005) tussen RoboDennett – de robot die functioneel isomorf is aan Dennett maar zijn handelingen baseert op wat de simulatie in zijn hersenen doet – en Dennett zelf. Ten eerste zal RoboDennett geen onbewuste reacties geven op het zien van een kleur. Als voorbeeld noemt Beaton dat rood licht leidt tot een verhoogde staat van alertheid in resusaapjes (Humphrey and Keeble, 1978). Aangezien dit soort processen zich allemaal buiten onze bewuste controle om voltrekken is het onwaarschijnlijk dat RoboDennett deze processen kan beïnvloeden. Dennett zelf kan deze onbewuste processen natuurlijk ook niet beïnvloeden, maar zijn intacte lage niveau visuele systeem veroorzaakt deze processen wel. Een tweede verschilpunt is de reactietijd. In mensen zijn bewuste reacties altijd langzamer dan onbewuste reacties. Als RoboDennett probeert om

---

in dit geval de noodzakelijkheid van de lage niveau visuele representatie.

de onbewuste reacties zo goed mogelijk na te bootsen zal zijn reactietijd dus langer zijn dan die van Dennett.

Het is echter de vraag of deze onbewuste processen zo interessant zijn voor een definitie van wat het is om een ervaring te hebben. We kunnen ons immers goed voorstellen dat iemand in coma, die dus geen fenomenaal bewustzijn heeft, deze automatische reacties (ten dele) nog wel vertoont. Deze automatische reacties staan dan dus los van onze bewuste ervaringen als gevolg van onze kleurwaarneming.

## 5.10 De behavioristische trekken van de heterofenomenologie

RoboDennett zou in de theorie van Dennett dus wel degelijk het volkopsychologisch concept *wat het is om* een ervaring te hebben, toegeschreven kunnen worden, zelfs indien hij al het onbewuste gedrag niet vertoont. Heterofenomenologisch is er, in tegenstelling tot wat Beaton (2005) beweert, volgens mij geen onderscheid te maken tussen RoboDennett en Dennett. Puur gedragsmatig<sup>34</sup> opereren beide immers op dezelfde manier. Deze conclusie is echter radicaal behavioristisch: zolang je je gedraagt als taalgebruiker, ben je bewust (Kirk, 1993). Als dit Dennetts conclusie is dan is het onduidelijk waarom hij een heel boek heeft geschreven over het bewustzijn, en welke waarde we moeten hechten aan zijn metaforen.

---

34 Voor zover dit gedrag van belang is om de ervaring van een kleur toe te schrijven. Over het toeschrijven van toestanden aan entiteiten heeft Dennett uitvoering gesproken in zijn boek *The intentional stance* (Dennett, 1987). Het is volgens Dennett logisch om intentionaliteit toe te schrijven aan die entiteiten die zich intentioneel (lijken te) gedragen. Hier volgt dezelfde behavioristische conclusie. Intentionaliteit is niet meer dan het gedrag waaruit we dit kunnen afleiden. Eenzelfde attributie lijkt Dennett te willen invoeren voor het bewustzijn (Dennett, 1991), zie ook zijn behandeling van filosofische zombies (Dennett, 1995).

Echter, als we een theorie hebben voor fenomenale ervaringen waarin de representatie in V4 constitutief is voor deze fenomenale ervaringen volstaat een eenvoudige breinscan van RoboDennett om te bepalen dat er geen activatie is in zijn V4 (of het functioneel equivalente RobotV4), op basis waarvan we kunnen concluderen dat hij geen waarachtige fenomenale ervaringen heeft. Dennett geeft een dergelijke indicator echter niet binnen zijn theorie, waardoor dit onderscheid weg zou vallen.

### **5.11 De onderdeterminatie van 'RoboMary'**

Bij RoboMary stelt Dennett echter dat wanneer zou blijken dat het op één of andere manier noodzakelijk is voor haar uiterlijke gedragingen dat de overtuigingen van RoboMary gebaseerd zijn op een representatie in haar kleurenregister dat zij eenvoudig een dergelijk kleurenregister kan simuleren. De vraag is echter wat dit aantoont, los van de hierboven beargumenteerde onmogelijkheid van het van 'binnenuit' aanpassen van de eigen breinarchitectuur. Als *locked* RoboMary op deze wijze in een toestand komt die volledig functioneel equivalent is aan een toestand waarin haar kleurwaarneming gewoon zou werken vormt zij een ondersteuning voor willekeurige welke functionalistische theorie.

Echter, om aan te nemen dat RoboMary überhaupt een werkzaam gedachte-experiment is, moeten we het functionalisme al als aanname hanteren. Onze theoretische aannames met betrekking tot bewustzijn mogen dan immers geen verschil maken tussen mensen en functioneel equivalente robots. Met andere woorden: het gedachte-experiment RoboMary van Dennett kan gebruikt worden om elke willekeurige functionalistische theorie van fenomenale ervaringen te aanvaarden. Maar als je dit toch al als premisse moet accepteren is de toegevoegde waarde van RoboMary zeer beperkt. Ze vormt hoogstens een goed model om functionalisme toe te lichten.

Dennetts poging zijn eliminativisme met betrekking tot qualia te onderbouwen blijkt dus weinig succesvol te zijn. In plaats van een duidelijk bewijs voor zijn stelling dat qualia te reduceren zijn tot reactieve disposities laat het gedachte-experiment ruimte voor alle functionalistische theorieën. Als we dus andere argumenten hebben om de vertrouwdheidshypothese te verkiezen boven Dennetts eliminativisme worden we hierbij niet gehinderd door RoboMary. In het volgende deel zal ik daarom op zoek gaan naar de voordelen die Churchlands theorie heeft op die van Dennett.

# QUALIA: VAN REACTIEVE DISPOSITIES NAAR DE VERTROUWDEHEIDSHYPOTHESE

*Precies! Er lijkt een fenomenologie te zijn. Dat is een feit dat de heterofenomenoloog van harte onderschrijft. Maar uit dit onuitwisbare, universeel geldige feit volgt nog niet dat fenomenologie echt bestaat. Daar draait het om.*

DANIEL C. DENNETT (1991/1999, p. 402-403)

In het vorige deel heb ik laten zien dat Dennetts verdediging van zijn eliminatie van qualia – gedefinieerd als een complex van reactieve disposities – ondergedetermineerd is. 'RoboMary' staat alternatieve functionalistische interpretaties toe. In deel III hebben we gezien dat Churchland een dergelijke visie heeft ontwikkeld. In dit deel vergelijk ik daarom de visies van Dennett en Churchland, om te concluderen dat Churchland beter in staat is om onze fenomenale ervaringen te verklaren. Vervolgens bekijk ik of er in de vertrouwdheidshypothese ruimte is voor de directe introspectie van hersentoestanden, zoals Churchland beweert, een fundamentele voorwaarde voor een derde-persoonstheorie van bewustzijn volgens Churchland.

## **6.1 Defictionaliseren: de grondslag van de fenomenologie in het brein**

In zijn poging om de goochelvoorstelling van het bewustzijn te *reverse-engineeren* en daarbij tegelijkertijd het Cartesianes Theater te vermijden lijkt Dennett elke vorm van hiërarchie in het brein te willen uitsluiten. Parallele processen werken naast elkaar en door elkaar heen zonder dat

er een centrale plaats is waar het allemaal samenkomt. Het bewustzijn is in Dennetts theorie geen uitdrukking van wat er in de hersenen gebeurt, maar vormt een fictionele laag, afgesloten van de hersenprocessen die eraan ten grondslag liggen. De joyceaanse machine van Dennett is bovendien zelf een instantiatie van het Cartesiaanse Theater, een plaats waarin alles samenkomt tot één *stream of consciousness*. De meervoudige conceptversies veronderstellen immers al één hoofdredacteur die het geheel aan elkaar breit (Kirk, 1993).

Het lijkt dan alleen maar alsof er een fenomenale, introspectief toegankelijke wereld is. Maar onze fenomenologische ervaringen zijn simpelweg hoe de dingen ons voorkomen, hoe ze voor ons *lijken* te zijn. Het onderscheid dat Dennett hier maakt tussen de manifeste fenomenologie en de wetenschappelijk-ontologische realiteit hiervan is niet te maken (Rudd, 1999). Dit 'lijken-alsof' moet immers op zijn minst bestaan, of lijkt het slechts alsof dit 'lijken-alsof' bestaat (McGinn, 1995)?

Churchland (1985; 1989) geeft ons een visie die deze oneindige regressie waar Dennett in verzeild raakt door zijn heterofenomenologische methode, een halt toe kan roepen. Een ervaring van rood wordt veroorzaakt door een representatie van 'rood' in de hersenen. Deze representatie is volledig reëel, alleen niet linguïstisch te beschrijven. Door de vertrouwdheidshypothese aan te nemen, ontstaat er weer ruimte voor de fenomenale ervaringen. Dit zijn geen hersenschimmen, zoals Dennett ons wil doen geloven, maar zijn direct gefundeerd in de werking van onze hersenen. Onze interne wereld wordt zo dus niet langer ontkend, maar opnieuw opgebouwd uit reële representaties van de wereld om ons heen.

In deel III hebben we gezien dat Churchland de introspectie van deze representaties analoog aan perceptie definieert. Dit betekent wellicht dat we alsnog een derde-persoonstheorie van bewustzijn kunnen formuleren, wanneer het mogelijk zou zijn om onze kennis op basis van ver-

trouwdheid te reduceren tot propositionele kennis. In de rest van dit deel bekijk ik Churchlands poging om dit te doen.

## **6.2 Hoe theoretische kennis onze perceptie niet verandert**

Volgens Churchland kan introspectie gezien worden als een vergelijkbaar proces als perceptie. Dit is van belang omdat hij meent dat de wetenschap onze perceptie fundamenteel veranderd heeft en dat hetzelfde zal gebeuren voor introspectie. Door onze volkspychologische concepten op wetenschappelijk begrip te brengen, kunnen we volgens Churchland de manier waarop we naar de wereld en naar onszelf kijken drastisch veranderen zonder onze aangeboren perceptuele apparaten te veranderen. Om te begrijpen hoe introspectie getransformeerd kan worden, bekijk ik eerst Churchlands kijk op perceptie.

Churchland (1985) geeft een aantal voorbeelden van hoe een theoretische onderbouwing kan bijdragen aan het uitbouwen van perceptuele discriminatie. Neem een wijnkenner die op basis van zijn jarenlange ervaring met het proeven van wijn vijftien tot twintig verschillende elementen kan detecteren op basis van zijn smaakvermogen: ethanol, glycol, fructose etc.

Laten we een ander voorbeeld nemen. Stel dat je je hand boven een kaars houdt, waardoor je een ervaring van warmte hebt. Er zijn nu twee mogelijkheden om deze ervaring te beschrijven. De eerste is dat door de verbranding flogiston vrijkomt uit de kaars (Priestley, 1796), die je tastzin waarneemt als warmte. De tweede mogelijkheid is de meer moderne verklaring dat de kaars verbrandt door de continue aanvoer van zuurstof en dat de warmte die je voelt de warmtestraling is die vrijkomt door dit proces. Is er puur op basis van de ervaring van warmte een onderscheid te maken tussen deze twee theorieën? Natuurlijk niet, wat

je waarneemt is immers niet direct de warmtestraling of flogiston, maar de prikkeling van zenuwen.

Maar dit heeft implicaties voor de wijnkenner. Natuurlijk is het zo dat de verschillende stoffen in wijn door een sommelier te detecteren zijn en dat we uit chemisch onderzoek weten dat deze stoffen bestaan en hoe ze zijn opgebouwd. Het is echter niet zo dat de perceptie van de wijnkenner verandert op het moment dat hij kennis neemt van de achterliggende chemie. Stel dat de wijnkenner per ongeluk de namen van de verschillende elementen door elkaar gooit, hij noemt de smaak van fructose opeens ethanol en vice versa. Dit weerhoudt hem er niet van om de smaak van wijn te beschrijven, alleen zal zijn beschrijving niet overeen komen met de feitelijke compositie. Het is bovendien goed denkbaar dat we een wijnkenner vanaf zijn geboorte voor de gek kunnen houden en hem de verkeerde namen voor de verschillende elementen aan kunnen leren. Dit zou voor zover ik kan zien geen enkel effect hebben op zijn succes als wijnkenner in bijvoorbeeld het leren herkennen van specifieke wijnen. De wijnkenner kan dus de verschillende onderdelen waaruit de smaak van een wijn is opgebouwd prima labelen met begrippen uit de chemie, maar dat is dan ook precies wat er gebeurt: het labelen van een smaakonderscheid dat er al was.<sup>35</sup>

### **6.3 Een wittgensteiniaanse analyse van Churchlands neurofysiologisch geïnspireerde introspectie**

Introspectie is volgens Churchland (1985) goed te vergelijken met perceptie: er is sprake van een soort interne waarneming van perceptuele representaties. Door onze introspectie te funderen in onze kennis van de hersenen zullen we toestanden van onze hersenen kunnen waarnemen als nooit te voren:

---

35 Voor een uitgebreidere verwerping van Churchlands idee dat we via introspectie onze eigen hersentoestanden kunnen introspecteren zie Shoemaker (1984).



*Dopamine levels in the limbic system, the spiking frequencies in specific neural pathways, resonances in the nth layer of the occipital cortex, inhibitory feedback from the lateral geniculate nucleus, and countless other neurophysiological niceties could be moved into the objective focus of our introspection.*

Churchland (1985, p. 16)

Maar op dezelfde wijze als we hierboven gezien hebben bij perceptie is het ernstig de vraag of deze mogelijkheid echt open ligt voor ons. Natuurlijk staat het ons vrij om onze ervaringen opnieuw te labelen. In plaats van “ik zie rood” kan ik ook zeggen dat ik een distincte representatie in mijn hersengebied V4 heb, die op zijn beurt weer veroorzaakt is door de electromagnetische straling van een bepaalde golflengte. De ervaring zelf zal hierdoor echter onveranderd blijven. Net als dat we op basis van onze ervaring van warmte geen onderscheid kunnen maken tussen de flogiston theorie en de moderne verbrandingstheorie, kunnen we ook niets zinnigs afleiden uit onze introspectie over de opbouw van onze hersenen.

Bovendien heeft Churchland teveel toegegeven aan de qualiafreaks. Hij ziet de qualia nog steeds als een mentaal object – weliswaar te reduceren tot fysische processen – dat door het geestesoog wordt waargenomen. Dit zou immers betekenen dat de quale als een soort geest-onafhankelijke entiteit wordt gepositioneerd (Rudd, 1999). Er is een representatie van rood in V4, die onafhankelijk van de eventuele introspectie hiervan bestaat in de visie van Churchland. Pas de hogere orde perceptuele representatie van de lage niveau visuele representatie is causaal effectief in de mentale processen. Het grote probleem van Churchland is dat hij Dennett niet goed heeft gelezen. Er is geen interne observator die neer kan kijken op de hersentoestanden en de buitenwereld hiervan op de hoogte kan brengen. Wanneer Mary voor het eerst de rode roos ziet en zegt: “Dus zo ziet

rood eruit!” refereert ze niet aan een mentaal object “rood”, dat wetenschappelijk geherinterpreteerd kan worden in neurofysiologische termen, maar geeft ze expressie aan een nieuwe sensatie die ze heeft.

Wanneer we terugdenken aan Wittgenstein kunnen we begrijpen wat hier gebeurt. Ons spreken over onze ervaring is niet-referentieel, maar expressief. “A child has hurt himself and cries: and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain behaviour.” (Wittgenstein, 1953, §244). Zo is het goed mogelijk dat in de toekomst het kind aangeleerd zal worden dat zijn *c-vezels vuren* wanneer het zichzelf pijn heeft gedaan. Maar dit zal niets veranderen aan de ervaring hiervan.

#### **6.4 De onreduceerbaarheid van het eerste-persoonsperspectief: de vertrouwdheidshypothese geherformuleerd**

Anders dan Churchland meent, hebben we geen introspectieve toegang tot onze representatie van kleuren (en smaken, geuren etc.) in onze hersenen, zeker niet als zijnde hersenprocessen. Onze uitspraken over ons mentale leven zijn in de geest van Wittgenstein expressief in plaats van referentieel. Dit betekent overigens niet dat deze expressie de enige wijze is waarop onze sensaties kunnen participeren in het taalspel.

De vertrouwdheidshypothese maakt het mogelijk om in te zien dat er verschillende vormen van kennis naast elkaar bestaan. Zelfs in een materialistische wereld, zoals degene waar wij in leven, zijn er meerdere toegangen tot dezelfde informatie. De eerste is het voor ons allen bekende eerste-persoonsperspectief, het *wat het is om* een ervaring te hebben. De processen die ten grondslag liggen aan dit eerste-persoonsperspectief zijn vanuit het derde-persoonsperspectief te beschrijven en te begrijpen. Een complete neurofysi-

ologie zal uitputtend onze visuele perceptie kunnen beschrijven, zonder dat er één of ander niet-fysische eigenschap is die over het hoofd gezien wordt. Wat zij echter niet kan beschrijven is het fenomenale karakter van de ervaring. Het eerste-persoonsperspectief op dezelfde fysische processen is dus niet te reduceren tot een derde-persoonsbeschrijving. Onze hersenen zijn gewoon niet zodanig geëvolueerd dat we elke ervaring die we hebben op linguïstische wijze kunnen beschrijven.



# DE ONMOGELIJKHEID VAN EEN DERDE-PERSOONS- THEORIE VAN BEWUSTZIJN

In deze scriptie heb ik laten zien dat Dennetts gedachte-experimenten geen aanleiding geven om het fenomenale karakter van onze ervaringen te elimineren. Het door hem aangedragen gedachte-experiment RoboMary is onderhevig aan een circulaire redenering: zolang je een functionalistisch materialisme aanneemt krijg je dit er aan het eind van het gedachte-experiment, ook weer uit. De poging van Dennett om de *black box* van zijn behavioristische leermeester Ryle functionalistisch in te vullen laat diezelfde *black box* zo angstvallig leeg. De reductie van de fenomenale ervaring tot een complex van reactieve disposities lijkt welhaast een terugkeer naar de klassieke stimulus-respons relaties. De verschuiving van het explanandum van bewuste ervaringen naar de overtuigingen die mensen hebben over deze overtuigingen in de heterofenomenologie vormt de methodische legitimatie voor Dennetts eliminativisme.

Dennetts 'radicale' versie van Wittgensteins kritiek op de reïficatie van mentale objecten leidt tot zijn overbodige fictionalisering van het bewustzijn, die bovendien stuit op onoverkomelijke op bezwaren. Hoe kan het zijn dat het slechts lijkt alsof er een fenomenologie is? Een heldere interpretatie van zijn grote voorbeeld Wittgenstein laat zien dat zelfs wanneer onze uitspraken over onze ervaringen niet referentiëel zijn dit niet hoeft te betekenen dat deze ervaringen niet bestaan.

Een goed alternatief voor Dennetts al te behavioristische conclusies vormt een uitbreiding op de vertrouwde hypothese van Churchland. De door Dennett gefictionaliseerde fenomenale ervaringen blijken wel degelijk reële repre-

sentaties te zijn in de hersenen. Maar anders dan Churchland meent, is geen interne observator nodig die naar deze representatie kijkt om hem causaal effectief te maken; we moeten oppassen dat we onszelf niet opeens alsnog in het Cartesiaans Theater begeven. In plaats daarvan is de veraste uitspraak van Mary “Dus zo ziet rood eruit” een expressie van een nieuwe ervaring, een ervaring die ze niet kon afleiden in haar zwart/witte kamer.

Er zijn dus wel degelijk qualia. De complexe representaties in onze hersenen zijn privaat en letterlijk onuitspreekbaar, ofwel niet linguïstisch te beschrijven. Ze vormen bovendien een intrinsieke eigenschap van onze waarneming die niet kan worden teruggevoerd op de wereld van elektromagnetische golven en luchttrillingen. Er is alleen geen extra element, iets niet-fysisch, waardoor een betere kennis van onze hersenen wel degelijk kan bijdragen aan een beter begrip van onze fenomenale ervaringen. Het enige element dat we uit de definitie van Dennett moeten laten vallen is de directe toegankelijkheid vanuit het bewustzijn. Dit veronderstelt immers een mentaal object, dat door het bewustzijn vanuit een centrale plaats geïnterpreteerd en verwerkt wordt. In een fysicalistische herinterpretatie van qualia die Wittgenstein serieus neemt kunnen we qualia dus niet langer zien als mentale objecten waaraan gerefereerd wordt in ons spreken over onze mentale wereld, maar als sensaties waaraan we op expressieve wijze uiting kunnen geven.

Het bewustzijn is dus, anders dan Dennett meent, nog lang niet verklaard. Een degelijke analyse van onze fenomenale ervaringen, waarbij zowel het eerste- als het derdepersoonsperspectief serieus genomen wordt kan ons echter wel een belangrijk houvast geven om een fysicalistische theorie voor het bewustzijn te ontwikkelen, maar ook om de grenzen van de verklaarbaarheid van bewustzijn vanuit een derde-persoonsperspectief te verkennen.







# REFERENTIELIJST

- Beaton, M. (2005). What RoboDennett Still Doesn't Know. *Journal of Consciousness Studies*, 12, 3-25.
- Block, N. (1978). Troubles with functionalism. *Minnesota Studies in The Philosophy of Science*, 9, 261–325.
- Block, N. (1990). Inverted Earth. *Philosophical Perspectives*, 4, 52-79.
- Block, N. (1993). Consciousness Ignored. Review of Daniel Dennett's *Consciousness Explained*. *The Journal of Philosophy*, 181-193.
- Chalmers, D. J. (2007). Consiousness and its Place in Nature. In S. Stich and F. Warfield (Eds.) *The Blackwell Guide to Philosophy of Mind* (pp. 102-142). Oxford: John Wiley & Sons.
- Churchland, P.M. & Churchland, P.S. (1981). Functionalism, Qualia, and Intentionality. *Philosophical Topics*, 12:1, 121-145.
- Churchland, P.M. (1985). Reduction, Qualia, and the Direct Introspection of Brain States. *Journal of Philosophy*, 82:1, 8-28.

- Churchland, P.M. (1989). Knowing Qualia: A Reply to Jackson. In P. Ludlow, Y. Nagasawa & D. Stoljar (Eds.) *There's something about Mary* (pp. 163-172). Cambridge, MA: MIT Press.
- Cottrel, A. (1999). Sniffing the Camembert: On the Conceivability of Zombies. *Journal of Consciousness Studies*, 6, 4-12.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.
- Degenaar, M. en Lokhorst, G. (2011). Molyneux's Problem, *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*, E.N. Zalta (Ed.). URL = <<http://plato.stanford.edu/archives/fall2011/entries/molyneux-problem/>>.
- Dennett, D. (1979). On the Absence of Phenomenology. In Gustafson & Tapscott (Eds.) *Body, Mind and Method*. Dordrecht: Kluwer.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge, Ma: The MIT Press.

- Dennett, D. (1988). Quining Qualia. In A. Marcel and E. Bisiach (Eds.), *Consciousness in Modern Science*. Oxford: Oxford University Press.
- Dennett, D. (1991). *Consciousness Explained*. London: Little, Brown and Co.
- Dennett, D. (1993). The Message is: There is no Medium (reply to Jackson, Rosenthal, Shoemaker & Tye). *Philosophy & Phenomenological Research*, 53:4, 889-931.
- Dennett, D. (1994). Tiptoeing Past the Covered Wagons. In U. Neisser & D. Jopling (Eds.) *The Conceptual Self in Context: Culture, Experience, Self-Understanding*. Cambridge: Cambridge University Press.
- Dennett, D.C. (1995). The Unimagined Preposterousness of Zombies. *Journal of Consciousness Studies*, 2, 322-6.
- Dennett, D. (1999). *Het bewustzijn verklaard* (T. Maas & F. Smeets, Vert.). Amsterdam: Olympus. (Originele werk gepubliceerd in 1991).
- Dennett, D. (2001). Are we explaining consciousness yet? *Cognition*, 79:1-2, 221-237.

- Dennett, D. (2003). Who's On First? Heterophenomenology Explained, *Journal of Consciousness Studies*, 10(9-10), 19-30.
- Dennett, D. (2003b). Explaining the "Magic" of Consciousness. *Journal of Cultural and Evolutionary Psychology*, 1(1), 7-19.
- Dennett, D. (2004). Consciousness: How much is that in real Money? In R. L. Gregory (Ed.), *Oxford Companion to the Mind*. Oxford: Oxford University Press.
- Dennett, D. (2006). What RoboMary Knows. In T. Alter en S. Walter (Eds.) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press.
- Dennett, D. (2007). Heterophenomenology Reconsidered, *Phenomenology and Cognitive Science*, 6, 247-270.
- Dennett, D.C. (2008). Daniel Dennett: Autobiography (Part 1). *Philosophy Now*, 68, niet gepagineerd, opgehaald van <[http://philosophynow.org/issues/68/Daniel\\_Dennett\\_Autobiography\\_Part\\_1](http://philosophynow.org/issues/68/Daniel_Dennett_Autobiography_Part_1)>.

- Graham, G. en Horgan, T. (2000). Mary, Mary, Quite Contrary. *Philosophical Studies*, Vol 99, 59-78.
- Graham, G. (2010). Behaviorism, *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2010/entries/behaviorism/>>.
- Held, R., and Y. Ostrovsky, B. Degelder, T. Gandhi, S. Ganes, U. Mathur, and P. Sinha, 2011, "The newly sighted fail to match seen with felt", *Nature Neuroscience*, 14: 551-553.
- Humphrey, N. K. en Keeble, G. R. (1978). Effects of red light and loud noise on the rate at which monkeys sample the sensory environment. *Perception*, 7, 343-348.
- Hutto, D. D. (1995). Consciousness Demystified: A Wittgensteinian Critique of Dennett's Project. *The Monist*, 78:4, 464-479.
- Jackson, F. (1982). Epiphenomenal Qualia. In P. Ludlow, Y. Nagasawa & D. Stoljar (Eds.) *There's something about Mary* (pp. 39-50). Cambridge: MIT Press.

- Jackson, F. (1986). What Mary didn't know. In P. Ludlow, Y. Nagasawa & D. Stoljar (Eds.) *There's something about Mary* (pp. 51-56). Cambridge, MA: MIT Press.
- Jackson, F. (1998). Postscript on Qualia. In P. Ludlow, Y. Nagasawa & D. Stoljar (Eds.) *There's something about Mary* (pp. 417-420). Cambridge: MIT Press.
- Jackson, F. (1998b). *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Kim, J. (1993). *Supervenience and Mind*. Cambridge: Cambridge University Press.
- Kirk, R. (1993). 'The Best Set of Tools'? Dennett's Metaphors and the Mind-Body Problem. *The Philosophical Quarterly*, 43:172, 335-343.
- Kirk, R. (2012). Zombies. *The Stanford Encyclopedia of Philosophy (Summer 2012 Edition)*. E.N. Zalta (ed.), opgehaald van = <<http://plato.stanford.edu/archives/sum2012/entries/zom-bies>>.
- Levin, J. (2010). Functionalism. *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*, E. N. Zalta (ed.), opgehaald van

- <http://plato.stanford.edu/archives/sum2010/entries/functionalism/>.
- Lewis, D. (1983). Postscript to "Mad Pain and Martian Pain". In *Philosophical Papers Volume I* (pp. 130-2). Oxford: Oxford University Press.
- Lewis, D. (1983b). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61: 343-77.
- Lewis, D. (1988). What experience teaches. In P. Ludlow, Y. Nagasawa & D. Stoljar (Eds.) *There's something about Mary* (pp. 77-103). Cambridge, MA: MIT Press.
- Locke, J. (1690). *An Essay Concerning Human Understanding*.
- Mandik, P. (2010). Swamp Mary's revenge: deviant phenomenal knowledge and physicalism. *Philosophical Studies*, 148:231-247.
- Mason, D. (2005). Demystifying Without Quining: Wittgenstein and Dennett on Qualitative States. *South African Journal of Philosophy*, 24:1, 33-43.
- McGinn, C. (1995). Consciousness Evaded: Comments on Dennett. *Philosophical Perspectives*, 9, pp. 241-249.

- Nemirow, L. (1990). Physicalism and the Cognitive Role of Acquaintance. In W. Lycan (Ed.) *Mind and Cognition: A Reader*. Oxford: Blackwells
- Nagel, Th. (1974). What Is It Like to Be a Bat?, *The Philosophical Review*, 83: 435–450.
- Priestley, J. (1796). *Considerations on the doctrine of phlogiston, and the decomposition of water*. Philadelphia: Thomas Dobson.
- Shoemaker, S. (1984). Churchland on Reduction, Qualia, and Introspection. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol 2, 799-809.
- Schrödinger, E. (1958). The Mystery of Sensual Qualities, in *What is Life? with Mind and Matter and Autobiographical Sketches*. Cambridge: Cambridge University Press
- Stoljar, D. en Nagasawa, Y. (2004). Introduction. In P. Ludlow, Y. Nagasawa & D. Stoljar (Eds.) *There's something about Mary* (pp. 1-36). Cambridge, MA: MIT Press.
- Ramachandran, V.S. & Hubbard, E.M. (2001). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society of London, B*, 268, 979-983.



- Ramsey, W. (2011). Eliminative Materialism, *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, E. N. Zalta (Ed.), opgehaald van <<http://plato.stanford.edu/archives/spr-2011/entries/materialism-eliminative/>>.
- Robinson, H. (1993). Dennett on the Knowledge Argument. In P. Ludlow, Y. Nagasawa & D. Stoljar (Eds.) *There's something about Mary* (pp. 69-74). Cambridge, MA: MIT Press.
- Rudd, A.J. (1999). What It's Like and What's Really Wrong with Physicalism: A Wittgensteinian Perspective. *Journal of Consciousness Studies*, 5:4, 454-63.
- Tye, M. (2009). Qualia. In Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy (Summer 2009 Edition)*, opgehaald op 12 april 2012 van <http://plato.stanford.edu/archives/sum2009/entries/qualia/>
- Vallor, S. (2009). The fantasy of third-person science: Phenomenology, ontology and evidence. *Phenomenology and Cognitive Science*, 8, 1-15.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

# BRONVERMELDING FIGUREN

Figuur 1: Daniel Dennett, at the Second World Conference on the Future of Science, in Venice, 2006, foto gemaakt door David Orban. Opgehaald van [http://commons.wikimedia.org/wiki/File:Daniel\\_Dennett\\_in\\_Venice\\_2006.png](http://commons.wikimedia.org/wiki/File:Daniel_Dennett_in_Venice_2006.png).

Figuur 2: Mary's Room, ongedateerd, niet geattribueerd. Opgehaald van de site van Jesse Prinz <http://subcortex.com/pictures/>.

Figuur 3: Voorbeelden van stimuli, uit Held et al. (2011).

Figuur 4: Paul M. Churchland, ongedateerd, niet geattribueerd. Opgehaald van de website van Paul Churchland <http://philosophyfaculty.ucsd.edu/faculty/pchurchland/>.

Figuur 5: Uit: Compton's Pictured Encyclopedia, 1932.

Figuur 6: Ludwig Wittgenstein, ongedateerd, niet geattribueerd. Opgehaald van <http://www.solutions-centre.org/Conversations/dialogues/intro.htm>



In zijn boek *Het bewustzijn verklaard* stelt Dennett dat onze fenomenologie een illusie is. Onze hersenen laten ons slechts geloven dat er een innerlijke wereld is, een eerste-persoonsperspectief op de wereld om ons heen. In deze scriptie betoog ik dat deze fictionalisering van het bewustzijn overbodig is. Dennetts behandeling van het gedachte-experiment Mary onderbouwt zijn theorie niet, maar is ondergedetermineerd en laat ruimte voor andere interpretaties.

Aan de hand van het werk van Paul Churchland laat ik zien dat een physicalistische theorie van bewustzijn wel degelijk een plaats kan geven aan qualia, door verschillende vormen van kennis aan te nemen voor het eerste- en derde-persoonsperspectief op bewustzijn. Een niet-ontologische interpretatie van Dennetts belangrijkste inspiratiebron Ludwig Wittgenstein maakt duidelijk dat deze materialistische qualia niet gereïficeerd kunnen worden; ons spreken over onze innerlijke wereld is expressief in plaats van referentiëel. Dit betekent dat een theorie over bewustzijn vanuit een derde-persoonsperspectief - zoals Dennett probeert te formuleren - fundamenteel onmogelijk is.