Philosophy of mind in wetenschappelijke context: een beknopt overzicht

Marc Slors

1 Introductie

Psychiatrie, klinische psychologie, cognitieve psychologie en een scala aan andere wetenschappen houden zich bezig met de werking van wat we in het dagelijks taalgebruik aanduiden als 'de menselijke geest'. We lijken te weten wat we met die term bedoelen, in die zin dat we weten hoe we erover moeten praten. Maar weten we ook wat de menselijke geest *is*? Als we geloven dat de 'geest' wordt voortgebracht door hersenprocessen, zoals de wetenschap dat lijkt te hebben aangetoond, hebben we dan eigenlijk enig idee van *hoe* grijze massa en bijvoorbeeld bewustzijn hetzelfde kunnen zijn? Voor de dagelijkse praktijk is een antwoord op een dergelijke vraag overbodig. Voor de wetenschappelijke praktijk vaak ook. Maar voor een bevredigend wetenschappelijk wereldbeeld niet. Het is daarom dat er een bloeiend filosofisch vakgebied bestaat dat geheel aan deze vraag gewijd is: de *philosophy of mind*. Dit hoofdstuk is bedoeld om het vakgebied te introduceren en om aan te geven waar en hoe het vak raakt aan de verschillende wetenschappen die zich bezighouden met de menselijke geest.

Philosophy of mind is een belangrijke tak van de filosofie, met name in de Angelsaksische landen, maar tegenwoordig zeker ook in een land als Duitsland. In sommige faculteiten filosofie in Engeland en de Verenigde Staten is dit het vakgebied van minstens de helft van de daar werkende filosofen. Datgene waarmee deze filosofen zich bezighouden, de *mind*, is moeilijk eenduidig in het Nederlands te beschrijven. Een woord als 'geest' dat ik hierboven gebruikte, of een woord als 'ziel' heeft vaak meer religieuze connotaties dan wenselijk is voor onbevooroordeeld onderzoek, en in ieder geval meer dan *mind*, dat in het Engels verschilt van *soul* en *spirit*. Een neutraler begrip als 'het mentale' komt dichterbij, maar kan in sommige zinsconstructies voor problemen zorgen. Om deze redenen zal ik in dit hoofdstuk het woord '*mind*' blijven gebruiken, daarbij verwijzend naar ons vermogen te denken, waar te nemen, te willen, emoties te hebben, taal te gebruiken, bewuste ervaringen te hebben, et cetera.

De overgrote meerderheid van de *philosophers of mind* stelt zichzelf tot doel één of meer aspecten van de *mind* zodanig te analyseren dat duidelijk wordt hoe en waarom zo'n aspect past in het zich ontwikkelende wetenschappelijke wereldbeeld. Vaak neemt zo'n project de vorm aan van het uitleggen hoe alledaagse begrippen als bijvoorbeeld 'vrije wil' of 'subjectiviteit' ondanks de schijn van het tegendeel niet in tegenspraak zijn met geldende wetenschappelijke kennis.

In die zin raakt zeer veel *philosophy of mind* aan wetenschappelijke projecten die zich bezighouden met, bijvoorbeeld, bewustzijn of vrije wil. Er is dan ook in toenemende mate sprake van vermenging van *philosophy of mind* en wetenschap. Daarbij moet uiteraard worden gedacht aan de psychologie, psychiatrie en de cognitieve neurowetenschappen. Maar er kan ook worden gedacht aan, bijvoorbeeld, evolutionaire biologie, sociolinguïstiek, sociobiologie, en zelfs culturele antropologie. Vaak is de samenwerking tussen filosofie en wetenschap vruchtbaar. Soms is er echter ook sprake van onbegrip of gebrekkige kennis van elkaars vakgebied.

In dit hoofdstuk wil ik allereerst een zeer beknopte, deels historische schets geven van drie van de belangrijkste filosofische theorieën over de aard van de *mind* die zijn ontwikkeld als alternatief voor het idee dat de *mind* moet bestaan uit een immateriële ziel. Daarna zal ik, om beter inzicht in de aard van het vakgebied te geven, vier verschillende probleemgebieden beschrijven waarover binnen de *philosophy of mind* wordt gediscussieerd, veelal tegen de achtergrond van een van de eerder besproken theorieën. Dit zal het eerste gedeelte van het hoofdstuk beslaan.

In het tweede gedeelte wil ik de verhouding tussen *philosophy of mind* en de wetenschappen bespreken. Daartoe bespreek ik zeer beknopt drie gevallen waarin *philosophy of mind* en wetenschap betrekking op elkaar hebben: de ontwikkeling van psychofunctionalisme in het licht van de zich ontwikkelende kennis over de hersenen, recente claims van neurowetenschappers over de aard van de wil, en het zogenaamde *theories-of-mind*-debat. De eerste twee gevallen zijn bedoeld om aan te geven dat filosofen en wetenschappers beter niet te veel 'op elkaars stoel kunnen gaan zitten'. Vruchtbare interactie tussen *philosophy of mind* enerzijds en met name de cognitieve neurowetenschappen en psychologie anderzijds dient uit te gaan van de eigenheid van de verschillende vakgebieden. Het derde voorbeeld, het *theories- of-mind*-debat is mijns inziens een goed voorbeeld van een vruchtbare en belangrijke interactie tussen filosofie en wetenschappen die juist voortkomt uit de erkenning van de eigenheid van de verschillende disciplines. Tot slot zal ik trachten aan te geven welke visies op de interactie tussen filosofie en wetenschap ten grondslag liggen aan deze drie voorbeelden.

2 *Philosophy of mind*

2.1 Vier belangrijke theorieën over de mind

Hoewel de wortels van de *philosophy of mind* zeker teruggaan tot Plato en Aristoteles, is het redelijk het begin van het vakgebied zoals we dat nu kennen halverwege de twintigste eeuw te lokaliseren. Daarbij moeten we dan wel direct één uitzondering maken: veel, heel veel *philosophy of mind* is in feite een reactie op het zeventiende-eeuwse lichaam-zieldualisme van Descartes, of beter, een poging daaraan te ontkomen.

2.1.1 Cartesiaans dualisme

Een belangrijke reden om Descartes als ijkpunt te nemen voor veel theorieën ligt in het feit dat de cartesiaanse opvatting van de *mind* als los van het lichaam bestaande ziel gezien kan worden als het schoolvoorbeeld van een theorie die er niet in slaagt de *mind* in te passen in het huidige wetenschappelijke wereldbeeld (Descartes 1637). De *mind*/ziel was volgens Descartes van een volstrekt andere orde dan de materiële werkelijkheid. Waar de essentie van materie 'uitgebreidheid' was – het vermogen ruimte in te nemen – was de essentie van de ziel voor hem 'denken' (waarbij onder 'denken' een heel scala aan mentale activiteit valt en zeker niet slechts logisch nadenken). De ziel neemt bij Descartes geen ruimte in.

Om meerdere redenen is deze opvatting wetenschappelijk gezien problematisch. Methodologisch is ze problematisch omdat de *mind*/ziel geen enkele objectieve toegang toelaat (alle wetenschappelijke instrumenten zijn immers 'uitgebreid'). Introspectie is onze enige toegang tot de *mind*. En ondanks het feit dat Wilhelm Wundt (1907) deze methode een wetenschappelijke status heeft proberen te geven bleek het grootste probleem ervan onoverkomelijk: de resultaten zijn nooit te verifiëren door derden.

Vanuit theoretisch oogpunt lag het grootste probleem in de interactie die de *mind* met het lichaam moet hebben. Via onze zintuigen beïnvloedt het uitgebreide lichaam de ziel. En via wilsimpulsen beïnvloedt de ziel het lichaam. Het probleem met deze interactie ligt er vooral in dat het lichaam wetenschappelijk gezien werkt volgens de wetten der natuur. Ingrijpen door de ziel, die niet in de natuurwetten voorkomt, zou betekenen dat natuurwetten gebroken worden. En dat is in regelrechte tegenspraak met natuurwetenschappelijke inzichten.

2.1.2 Filosofisch behaviorisme

Hoewel de onwetenschappelijkheid van de ziel als aparte substantie een belangrijke reden was voor de psychologie om zich van het cartesianisme af te keren waren er ook filosofische redenen. Ik beperk me hier tot de Angelsaksische filosofie. Zonder meer een van de belangrijkste en invloedrijkste critici van het cartesianisme was Gilbert Ryle, die in 1949 zijn *The concept of mind* publiceerde. Steen des aanstoots voor Ryle was het idee dat de cartesiaanse visie op de mens neerkwam op een '*ghost in a machine*'. Hoewel dit in zekere zin een karikatuur van Descartes was (de ziel was nooit een *ghost*) had Ryle beslist een punt met zijn kritiek op enerzijds het volstrekt onlichamelijke karakter van de *mind* en anderzijds het machinale karakter van lichaam en gedrag. Staande in de destijds sterke (taal)analytische traditie in Engeland stelde Ryle zich de vraag onder welke omstandigheden we geneigd zijn mentale termen (zoals '*belief*' en '*desire*') te gebruiken. Die voorwaarden en alleen die moeten worden gezien als betekenisverlenend voor de betreffende termen – en daarmee voor de verzamelterm '*mind*'. Zijn conclusie was dat mentale termen verwijzen naar disposities (grofweg 'geneigdheden onder specifieke omstandigheden') tot specifiek gedrag. Zo zal een deel van wat het betekent als ik zeg dat ik geloof dat Amsterdam de hoofdstad van Nederland is, bestaan uit mijn geneigdheid 'Amsterdam' te antwoorden wanneer iemand mij vraagt wat de hoofdstad van Nederland is. Of om, gegeven wat basale achtergrondkennis, Amsterdam aan wijzen op de kaart van Nederland als reactie op een verzoek de hoofdstad van Nederland aan te wijzen. Of ... et cetera. Een volledige lijst van soortgelijke disposities zou de volledige betekenis van mijn geloofstoestand vatten, aldus Ryle.

Ryle is een filosofisch behaviorist. Het is belangrijk hier filosofisch behaviorisme te onderscheiden van het psychologische behaviorisme van wetenschappers als Watson (1913) en Skinner (1974). Watson en Skinner waren van mening dat de *mind* een onwetenschappelijke entiteit was, vooral omdat hij (zeker na het debacle van het introspectionisme) niet wetenschappelijk te onderzoeken is. De

mind moest daarom volgens hen uit het wetenschappelijke wereldbeeld verwijderd worden; het was onwetenschappelijk geworden erover te spreken. Ryle deed in feite het omgekeerde: het begrip '*mind*' en mentale termen als '*belief*' en '*desire*' werden als onelimineerbaar beschouwd en moesten daarom opnieuw worden beschreven als het cartesianisme onhoudbaar bleek. Waar Watson en Skinner de *mind* elimineerden, *definieerde* Ryle de *mind*.

En daarmee staat Ryle aan het begin van een traditie die in eerste instantie tracht de *mind* te analyseren en definiëren en in die zin te begrijpen. Het ging er hem daarbij met name om ons alledaagse begrip van '*mind*' te articuleren. Compatibiliteit met de wetenschap stond voor hem als doel niet voorop.

2.1.3 Identiteitstheorieën

Dat doel werd vooral nagestreefd door Australische filosofen als U.T. Place en J.J.C. Smart (Smart 1959). Daarbij hadden ze aanzienlijk minder oog voor analyse van het *begrip* '*mind*' dan Ryle. Onder invloed van de psycholoog Edward Boring stelden zij dat de *mind* (en dan met name het bewustzijn) identiek moet zijn met de hersenen.

Boring verkondigde in 1933 een vroegere versie van een dergelijke identiteitstheorie binnen de psychologie, maar kreeg daarmee weinig voet aan de wetenschappelijke grond, hoofdzakelijk vanwege het destijds dominante behaviorisme. Place en Smart werkten enerzijds in een andere traditie – de filosofische – en anderzijds in een latere tijd, de jaren 1950, waarin ze een beginnende cognitivistische (d.w.z. anti-behavioristische) tijdgeest mee hadden. De belangrijkste bijdrage van Place aan de identiteitstheorie is zijn uitleg geweest van hoe we het begrip 'identiteit' moeten begrijpen, dat wil zeggen wat het betekent te zeggen dat de *mind* ons brein *is*. Het ging er uitdrukkelijk niet om uit te leggen *waarom* de *mind* identiek is aan onze hersenen. Uiteindelijk is die identiteit net zo min uit te leggen als de identiteit van water met H_20 – het gaat om feiten van de natuur. Een belangrijke bijdrage van Smart aan de identiteitstheorie is zijn uitleg van het feit dat mentale termen zogenaamd '*topic neutral*' zijn, hetgeen wil zeggen dat de contra-intuïtiviteit van deze theorie berust op niet gerechtvaardigde dualistische denkgewoonten.

Later, in de jaren 1960, krijgt de identiteitstheorie een wat analytischer karakter bij David Armstrong (1968), wederom een Australiër. Anders dan zijn voorgangers trekt Armstrong zich wel iets aan van Ryles dispositionele analyse van mentale termen. Maar anders dan Ryle claimt Armstrong dat die disposities uiteindelijk geconstitueerd worden en daarmee bestaan uit hersenstructuren. Veel elementen van het analytische functionalisme (zie onder) zijn bij Armstrong al te vinden.

2.1.4 Functionalisme

De *philosophy of mind* die na Ryle en de Australische materialisten ontstaat, draagt duidelijk een stempel van beide kampen: van Ryle neemt ze het project over het begrip '*mind*' te analyseren; van Place, Smart en Armstrong neemt ze het idee over dat die analyse vooral als doel moet hebben compatibiliteit met het wetenschappelijke wereldbeeld na te streven. Dit tweeledige karakter van theorievorming over de *mind* (eerst analyse, dan implementatie) is dan ook een van de belangrijkste kenmerken van de theorie die, in zeer verschillende vormen overigens, tot op heden de meest dominante is geworden: het functionalisme.

De identiteitstheorie, die de *mind* identificeerde met de hersenen, leek oorspronkelijk een lang leven beschoren, juist omdat ze nauw aansloot bij het destijds nieuwe cognitivisme in de psychologie, het idee dat menselijk gedrag niet te begrijpen is zonder te kijken naar interne – mentale – processen. Niettemin maakte het functionalisme een tamelijk abrupt einde aan deze theorie. Hoofdzakelijk omdat het een veel grotere verklarende kracht heeft waar het gaat om de vraag hoe en waarom hersenstructuren mentale toestanden ('*beliefs*' en '*desires*') kunnen realiseren.

De eerste basale gedachtestap achter de vroege vormen van functionalisme is de vergelijking van de *mind* met een 'turingmachine', een niet werkelijk bestaande machine die door Alan Turing is bedacht en die de voorloper van de moderne computer is. Het gaat om een machine die pas functioneert nadat wat we nu een 'programma' zouden noemen is ingelezen. Al in 1950 was de claim van Turing dat zo'n machine in principe intelligent zou kunnen zijn. Intelligentie moet daarbij begrepen worden in termen van het hebben van bepaalde capaciteiten of functionaliteit – niet in termen van een bepaald soort innerlijk waarin afwegingen gemaakt worden. In 1967 bouwt Hilary Putnam voort op dit idee en stelt dat ook de menselijke *mind* in principe begrepen zou kunnen worden naar analogie van een turingmachine.

Om het basale idee achter het functionalisme te begrijpen kan een eenvoudig voorbeeld helpen. Het gaat om het 'programma' dat een fictieve coca-cola-automaat bestuurt. In dit voorbeeld kost een blikje cola 20 cent, en zijn er slechts twee soorten munten, die van 10 cent en die van 20 cent. Figuur 1 geeft in de vorm van een matrix aan hoe de automaat dient te werken:

	Input 10 cent	Input 20 cent
Interne toestand 1	Output: geen Interne toestand wijzigt in toestand 2	Output: Blikje cola Interne toestand blijft toestand 1
Interne toestand 2	Output: Blikje cola Interne toestand wijzigt in toestand 1	Output: Blikje cola + 10 cent Interne toestand wijzigt in toestand 1

Figuur 1

Beginnend in interne toestand 1, zal de machine als volgt te reageren op input: als er een munt van 10 cent in de machine wordt gestopt dient er geen blikje cola uitgegeven te worden. Maar de interne toestand van het systeem moet veranderen. Immers nu zal een volgende input van 10 cent moeten leiden tot de uitgifte van een blikje cola. De toestand waarin de automaat een blikje cola zal uitgeven na input van nog eens 10 cent noemen we interne toestand 2. Na uitgifte van een blikje zal het systeem weer terugkeren in toestand 1 (anders zou een blikje immers opeens 10 cent kosten, en dat is niet de bedoeling). Wanneer het systeem in toestand 1 is en er wordt 20 cent ingevoerd, dan zal er een blikje worden uitgegeven en het systeem zal in toestand 1 blijven. Als er nu 20 cent wordt ingevoerd wanneer het systeem in toestand 2 is, geeft de machine niet alleen een blikje uit, maar zal hij ook 10 cent wisselgeld moeten teruggeven.

We zouden nu kunnen zeggen dat toestand 1 is te kenmerken als een toestand waarin het systeem 'gelooft' dat er nog geen geld in de machine is gestopt, en toestand 2 als de toestand waarin de machine 'gelooft' dat er 10 cent is ingegeven. Maar, en dat is cruciaal, natuurlijk is dat alleen houdbaar als het systeem ook geacht wordt te 'weten' dat een blikje cola 20 cent kost en dat er niet te veel betaald mag worden. Kortom, de geloofstoestanden ('*beliefs*') van het systeem kunnen alleen worden gekarakteriseerd in relatie tot de gehele bovenstaande matrix. Anders gezegd: de geloofstoestanden van het systeem worden gekenmerkt of gedefinieerd door alle relaties tussen (1) inputs, (2) outputs en (3) interne toestanden van het systeem in relatie tot het totale programma.

Het toeschrijven van geloofstoestanden aan zo'n systeem lijkt metaforisch. Toch is de claim van het functionalisme dat het verschil tussen onze geloofstoestanden en dat van de coca-cola-automaat slechts gradueel is: de 'programma's' volgens welke onze hersenen werken zijn oneindig veel complexer dan die van een automaat, maar niet *principieel* verschillend. Ook onze mentale toestanden kunnen worden gekenmerkt in termen van inputs, outputs en een gigantisch complexe matrix van interne toestanden die zullen veranderen onder invloed van wisselende in- en outputs.

Het sterke idee van het functionalisme is dat mentale toestanden zoals geloofstoestanden worden gekarakteriseerd los van de 'hardware' die het programma laat 'draaien'. De bovenstaande matrix geeft het uiterst eenvoudige 'mentale' leven van de automaat weer, zonder dat er iets wordt gezegd over hoe de machine fysiek gezien in elkaar zit. Er zijn in principe oneindig veel mogelijkheden om de bovenstaande matrix fysiek te realiseren. In jargon heet dit dat mentale toestanden zogenaamde *causal-role*-toestanden zijn, die verschillende *realisers* kunnen hebben. Belangrijk is dat dezelfde *causal roles* verschillend gerealiseerd kunnen worden, een idee dat 'meervoudige realisatie' wordt genoemd, waardoor het mentale niet meer direct kan worden geïdentificeerd met het fysieke.

Om het verschil tussen *roles* en *realisers* duidelijker te maken kan de volgende parallel helpen: toen Mendel in de negentiende eeuw de wetten van de genetica ontdekte, postuleerde hij entiteiten, genen, die de dragers waren van overerfbare eigenschappen. Genen spelen *causal roles* – dat wil zeggen ze worden gedefinieerd in termen van wat ze *doen*, en wel in strikt wetmatige zin, en niet in termen van wat ze fysiek gezien *zijn*. Pas in de jaren 1960 werd bekend wat de *realisers* van genen zijn: DNA.

In de *role/realiser*-distinctie weerspiegelt zich de tweestappenstrategie waarmee deze paragraaf begon: *eerst* worden mentale toestanden gedefinieerd (in termen van *causal roles* en een ingewikkelde matrix), *vervolgens* kunnen we kijken hoe een machine, een computer, of onze hersenen die *roles* en die matrix realiseren. Op deze manier kan worden *uitgelegd* hoe en waarom bepaalde

fysische structuren tot een *mind* kunnen leiden. Daarin zit het belangrijkste verschil met eerdere identiteitstheorieën die het gegeven dat onze hersenen een *mind* voortbrengen als een niet verder uit te leggen feit moesten aannemen. Dit belangrijke verschil verklaart het feit dat het functionalisme ongelooflijk snel leidde tot het verval van identiteitstheorieën.

Het is belangrijk hier te noemen dat er veel soorten functionalisme zijn. Ik zal één belangrijk onderscheid in typen functionalisme noemen. Analytische functionalisten, zoals Lewis (1972) of Dennett (1982) houden zich voornamelijk bezig met onze alledaagse toeschrijvingen van psychologische toestanden. Wij beschrijven, verklaren en voorspellen ons gedrag dagelijks in termen van *beliefs* en *desires*, zonder dat we precies weten wat er zich in onze hersenen afspeelt. Analytisch functionalisten claimen dat we deze zogenaamde *folk-psychology* kunnen uitleggen in termen van ingewikkelde matrixen van *causal-role* toestanden. Over de hersenen valt veel te weten te komen, maar we moeten niet *in* de hersenen gaan zoeken naar mentale toestanden; mentale toestanden zijn toestanden van het systeem (de mens) als geheel. Psychofunctionalisten zoals Fodor (1975) daarentegen, stellen dat *folk-psychology* in principe fouten maakt waar het gaat om het beschrijven en voorspellen van ons gedrag. Ze zal daarom uiteindelijk moeten worden vervangen door een wetenschappelijke psychologie. Volgens Fodor zou die wel degelijk moeten kijken naar hoe het feitelijke 'programma' van onze hersenen eruitziet. Psychofunctionalisten zoeken de functionele beschrijving van de *mind* dus wel degelijk op neuraal niveau. Onze hersenen zouden letterlijk een programmeertaal bevatten waarin neuraal gerealiseerde symbolen worden gemanipuleerd. Die symbolen representeren de wereld en constitueren daarmee wat we over de wereld denken. De manipulaties daarvan zijn onze mentale processen.

2.2 Deelproblemen

De theorieën over de aard van de *mind* die in de voorgaande paragraaf zijn besproken gaan over de *mind* als geheel. Ze trachten de vraag naar de relatie tussen *mind* en lichaam, inclusief hersenen, het *mind-body*-probleem, te beantwoorden. Het zou kunnen lijken alsof het *mind-body*-probleem het enige echte probleem in de *philosophy of mind* is. Dat is beslist niet het geval. Hoewel het probleem en de voorgestelde antwoorden daarop altijd op de achtergrond meespelen is er een waslijst aan problemen en vragen die het begrip '*mind*' oproept. Het is onmogelijk om al deze vragen en problemen te behandelen of zelfs maar te noemen. In deze paragraaf noem ik een aantal belangrijke om een indruk te geven van de variëteit aan vragen en problemen binnen de *philosophy of mind*. Vanwege de complexiteit van de vele antwoorden zal ik me beperken tot de vragen zelf. De sectie over vrije wil zal iets uitgebreider zijn omdat ik later op dat onderwerp terug kom.

2.2.1 Mentale inhoud, representaties en intentionaliteit

Mentale toestanden, *beliefs* en *desires*, gaan ergens over. Dat wil zeggen, ze hebben inhoud. Ik geloof *dat* Amsterdam de hoofdstad van Nederland is. Ik wens *dat* ik snel vakantie zal hebben. In die zin zijn mentale toestanden anders dan andere systeemtoestanden: hun essentie is dat ze buiten zichzelf verwijzen. Het zijn, in aan Franz Brentano ontleend jargon, *intentionele* toestanden. Anders gezegd: ze hebben *inhoud*, dat wil zeggen, ze *representeren* de wereld of een in de wereld gewenste situatie.

Een van de hoofdvragen in de *philosophy of mind* is hoe intentionaliteit, mentale inhouden of representaties van de wereld kunnen worden begrepen. Zijn intentionele of representationele toestanden te definiëren als puur interne toestanden van het brein? Vanuit de oorspronkelijk cartesiaanse intuïtie dat de *mind* een 'binnenwereld' is die zich strikt onderscheidt van de materiële buitenwereld, bestaat een sterke neiging om deze vraag bevestigend te beantwoorden. Maar deze intuïtie is problematisch: welke mysterieuze eigenschap van interne toestanden in het brein zorgt ervoor dat bepaalde toestanden *over* de wereld gaan? Als een gedachte over de wereld een hersenproces is, wat zorgt er dan voor dat die gedachte verwijst naar iets dat zich ver buiten die hersenprocessen en in de wereld bevindt? Een meerderheid van hedendaagse analytische filosofen is van mening dat we in ieder geval een vorm van zogenaamd 'externalisme' moeten aanhangen: mentale inhouden en representatie van de wereld moeten worden begrepen in termen van causale *relaties* die de door het brein gerealiseerde *mind* heeft met de buitenwereld. Een consequentie daarvan, die rechtstreeks indruist tegen de cartesiaanse intuïtie, is dat wat wij denken over de wereld niet los kan worden gezien van hoe de wereld *is*. Het strikte onderscheid tussen de *mind* als 'binnenwereld' en de materiële buitenwereld is op losse schroeven komen te staan.[1]

Het denken over de aard van intentionaliteit kan consequenties hebben voor het denken over bijvoorbeeld hallucinaties en wanen. Enerzijds kunnen die begrepen worden als 'mentale plaatjes'. Daarmee behoren ze tot een ander debat in de *philosophy of mind*, het debat over *mental imagery*. Anderzijds, echter, gaan die wanen en hallucinaties over de wereld. Een hallucinatie betreft een foutief veronderstelde stand van zaken in de wereld. Als externalisten gelijk hebben, is het beter om niet zo maar te stellen dat hallucinaties begrepen moeten worden als door de hersenen veroorzaakte mentale toestanden. Het is dan beter ze te begrijpen in termen van een verstoorde causale interactie tussen hersenen en de wereld, een interactie waarin door de hersenen aangestuurd handelen niet aangepast is aan de feitelijke stand van zaken in de omgeving van de actor.

2.2.2 Eerstepersoonsautoriteit, zelfbewustzijn, het 'zelf' en persoonlijke identiteit

Het verdwijnen van een strikt onderscheid tussen de binnenwereld van de *mind* en de buitenwereld heeft consequenties. Een van de noties die gebaseerd is op dat strikte onderscheid is het alledaagse idee dat ieder van ons toegang heeft tot zijn eigen binnenwereld op een manier die niemand anders heeft. Ik weet wat ik zelf denk, zo lijkt het, op een directe, onbemiddelde manier. Anderen weten slechts wat ik denk op een bemiddelde manier: door te luisteren naar wat ik zeg en door te zien hoe ik me gedraag.

Maar wat nu als de inhouden van mijn gedachten deels bepaald zijn door de buitenwereld zelf? Wat als de een of andere vorm van externalisme klopt? Dan is het niet volledig aan mij om te zeggen wat ik werkelijk denk. Een van de debatten rond onze eerstepersoonstoegang tot onze eigen *minds* draait om de vraag hoe we binnen de context van het externalisme toch ruimte kunnen maken voor het idee dat we zelf op een autoritatieve manier kunnen praten over wat er omgaat in onze eigen hoofden (zie bv. Davidson 1984).

De vraag naar eerstepersoonsautoriteit moet echter niet verward worden met de vraag naar de aard van zelfbewustzijn. Bij deze laatste vraag gaat het er niet om met welke autoriteit we kunnen spreken over wat er in ons omgaat, maar hoe we kunnen uitleggen wat het betekent te zeggen dat wat er in ons omgaat, omgaat in *onszelf*, en wel met een zekerheid die we niet kennen van andere vormen van kennis. Hoe is het mogelijk dat terwijl ik me over vrijwel alles kan vergissen, ik me niet kan vergissen in het feit dat *ik* het ben die bepaalde gedachten heeft? Wat is die 'ik' dan waarover ik me niet kan vergissen? En op wat voor manier ben ik me van hem bewust?[2]

Op haar beurt moet deze vraag niet worden verward met de vraag naar de aard van het 'zelf'. Doorgaans wordt met het 'zelf' veel meer bedoeld het subject van ervaring dat centraal staat in de vraag naar zelfbewustzijn. Het zelf is naast subject vooral ook actor in de wereld. Veelal wordt het zelf bediscussieerd als een set waarden en levensdoelen waarmee wij ons identificeren.[3] In het overgrote deel van de filosofische literatuur over dit onderwerp is het zelf niet zozeer een objectief bestaande entiteit als wel een construct van onszelf waarmee we een plaats innemen in de sociale wereld.

De vraag naar de aard van het zelf is verweven met (maar niet identiek aan) de vraag naar persoonlijke identiteit. Die laatste vraag betreft het door de tijd heen identiek blijven van een persoon. Gegeven dat mensen gedurende hun leven continu veranderen, zowel in lichamelijk als in psychologisch opzicht, is het de vraag waarom wij onszelf desalniettemin als *dezelfde* persoon blijven zien. Het merendeel van de filosofen die zich met deze vraag bezighouden is van mening dat het bij persoonlijke identiteit niet zozeer gaat om lichamelijke continuïteit als wel om psychologische continuïteit. De vraag is dan wat psychologische continuïteit precies behelst en hoe ze gedefinieerd kan worden. Voor theorievorming over psychologische continuïteit, met name wanneer daarin het zelf zoals hierboven besproken een rol speelt, zijn psychiatrische ziektebeelden, met name dissociatieve persoonlijkheidsstoornissen, relevant (zie ook hoofdstuk 5 van dit boek). In hoeverre kan er van psychologische continuïteit en daarmee persoonlijke identiteit worden gesproken wanneer eenzelfde persoon (numeriek) zich op verschillende momenten als een ander zelf ervaart?

2.2.3 Subjectiviteit, qualia en fenomenaal bewustzijn

Het vermeende 'binnenwereld'-karakter van de eerstepersoonstoegang tot onze eigen *minds* wordt vaak gevangen in het begrip 'subjectiviteit'. Een deel van dat begrip wordt gedekt door de noties van 'eerstepersoonsautoriteit' en 'zelfbewustzijn'. Het gaat daarbij, zoals in de voorgaande paragraaf is besproken, om bewustzijn van onszelf als subject en om direct bewustzijn van de inhouden van onze gedachten. Maar een deel van het begrip 'subjectiviteit' wordt niet door deze noties gedekt: het kwalitatieve aspect van onze subjectieve *ervaringen*.

Een voorbeeld: aan iemand die nooit suiker heeft geproefd is niet precies uit te leggen hoe suiker smaakt. Het woord 'zoet' betekent weinig voor iemand die nooit *zelf* zoet heeft geproefd. Het kwalitatieve karakter van de smaak van suiker is puur subjectief en is niet uit te leggen in objectiverende termen of met behulp van een gemeenschappelijke intersubjectieve taal. Hetzelfde geldt eigenlijk voor alle kwalitatieve aspecten van onze ervaringen: het voelen van pijn, het waarnemen van kleuren, het ervaren van emoties, et cetera.

Deze aspecten van onze ervaringen worden vaak aangeduid met de term 'qualia' (enkelvoud: quale). Omdat een aanduiding van dit aspect van subjectiviteit met een zelfstandig naamwoord snel leidt tot het begrijpen van ervaringen of eigenschappen daarvan als een soort *dingen*, hetgeen op zijn minst leidt tot metafysische problemen, wordt tegenwoordig vaker over 'fenomenaal bewustzijn' gesproken.

Fenomenaal bewustzijn is een van de belangrijkste problemen in de *philosophy of mind*, en wel vanwege het feit dat het moeilijk te verenigen is met het gangbare fysicalisme[4] en, belangrijker nog, de functionalistische orthodoxie.[5] Het grote probleem is het volgende: functionalisme begrijpt mentale toestanden in termen van de (causale) *relaties* tussen interne toestanden van een systeem als de hersenen en de buitenwereld (denk aan het coca-colavoorbeeld). Maar qualia of toestanden van fenomenaal bewustzijn laten zich juist niet zo begrijpen, ze zijn puur *intrinsiek*, dat wil zeggen niet afhankelijk van de causale relaties tussen hersenen en buitenwereld. Het is volstrekt denkbaar, zeggen qualia-aanhangers, dat twee mensen functioneel identiek zijn en dus bijvoorbeeld exact hetzelfde reageren op de kleuren zie ze tegenkomen (ze hetzelfde noemen) terwijl de subjectieve *ervaring* van die kleuren per persoon verschilt. Als dit mogelijk is, is het begrip '*mind*' niet in termen van functionalisme te vangen.

2.2.4 Vrije wil en mentale veroorzaking

Als laatste cluster van problemen bespreek ik vragen rond de vrije wil en mentale veroorzaking. Hoewel beide begrippen soortgelijke associaties lijken op te roepen en daarom gemakkelijk met elkaar kunnen worden geïdentificeerd, gaat het in de analytische filosofie om twee volstrekt van elkaar gescheiden debatten.

Het debat over de vrijheid van de wil heeft zich in de loop van de twintigste eeuw, in de filosofie tenminste (zie paragraaf 3.2 hieronder), deels losgemaakt van vragen rond de aard van de *mind*. Tegenwoordig is het meer een onderdeel van handelingstheorie, meta-ethiek en debatten over persoonlijke autonomie dan van de *philosophy of mind*. Laat ik kort proberen aan te geven hoe deze ontwikkeling is verlopen.

Oorspronkelijk werd het probleem begrepen in termen van een conflict tussen onze vermeende keuzevrijheid en de gedetermineerdheid van ons gedrag die lijkt te worden geïmpliceerd door de aanname dat de *mind* gerealiseerd wordt door onze hersenen. Hersenen, immers, zijn onderdeel van de natuurlijke orde en daarom onderhevig aan deterministische natuurwetten. Als vrijheid van de wil betekent dat we op ieder moment van keuze iets anders zouden kunnen kiezen dan we feitelijk doen, dan is vrije wil niet te rijmen met het idee dat de *mind* fysiek gerealiseerd en dus feitelijk gedetermineerd is (Van Inwagen 1982). 'Oplossingen' voor dit bijtende probleem kunnen ofwel bestaan uit ontkenning van onze gedetermineerdheid, ofwel uit onkenning van onze keuzevrijheid.

Maar er is een derde mogelijkheid die door een meerderheid van filosofen is onderzocht: we kunnen proberen het begrijp 'vrije wil' zodanig te herinterpreteren dat het conflict tussen vrijheid en determinisme verdwijnt. Er is in de loop van de twintigste eeuw een scala aan voorstellen gedaan om dit zogenaamde 'compatibilisme' vorm te geven.[6] Een deel van deze voorstellen concentreert zich op een alternatieve lezing van de notie 'keuzevrijheid' (zie bv. Dennett 1984). Maar een belangrijker mogelijkheid is om de notie van 'vrijheid' geheel los te koppelen van de notie van 'keuze' en meer direct in verband te brengen met praktische noties als (1) morele verantwoordelijkheid of (2) persoonlijke autonomie. Daarmee verschuift het debat over vrijheid van *philosophy of mind* naar meta-ethiek en/of handelingstheorie en autonomie. Ik zal beide bewegingen kort schetsen.

(1) Een van de functies van het begrip 'vrijheid' is dat het de voorwaarde is voor de praktische notie van 'verantwoordelijkheid': we kunnen slechts dan iemand verantwoordelijk houden voor zijn daden wanneer hij die daden in vrijheid begaan heeft. Vanuit deze optiek zou het volgens sommigen goed zijn het begrip 'vrijheid' te analyseren in termen van de feitelijke voorwaarden waaronder we mensen verantwoordelijk houden voor hun daden (zie ook hoofdstuk 8 van dit boek).[7]

(2) Als vrijheid van handelen betekent dat we kunnen doen wat we (al dan niet gedetermineerd) willen, dan bestaat vrijheid van willen uit de mogelijkheid te willen wat we willen. Deze laatste notie is minder onzinnig dan ze klinkt. Denk aan een roker die wil stoppen met roken maar

desalniettemin graag een sigaret wil opsteken. In zo'n geval wil de betreffende persoon roken maar tegelijkertijd zou hij willen dat hij die wil, die drang naar roken, niet (meer) heeft. Er is dan sprake van een eerste-ordewil (drang naar roken) en een tweede-ordewil (de wil niet meer te willen roken). Als deze terminologie zinnig is, kunnen we vrijheid van de wil begrijpen als het in overeenstemming zijn van onze eerste-ordewil met de tweede-ordewil waarmee we onszelf identificeren (in het voorbeeld: de wil te stoppen met roken). Vrijheid van de wil heeft dan direct te maken met persoonlijke identiteit, met de doelen en waarden waarmee we onszelf verbinden, en met persoonlijke autonomie, dat wil zeggen de mate waarin we in staat zijn die doelen te realiseren en waarden na te leven.[8] De determinismekwestie lijkt irrelevant te zijn.

Binnen deze opvatting van de vrije wil wordt de vraag naar wilszwakte of akrasia belangrijk. Is het coherent te denken dat mensen werkelijk een bepaalde handeling in een bepaalde situatie kunnen beoordelen als de beste terwijl ze desalniettemin feitelijk anders handelen? Verschillende antwoorden op deze vraag kunnen leiden tot verschillende visies op behandeling van psychologische problemen die aan wilszwakte gerelateerd zijn: bij een negatief antwoord, bijvoorbeeld, zou een geval van ogenschijnlijke wilszwakte eerder met behulp van cognitief-therapeutische middelen behandeld moeten worden.

Waar de kwestie van de vrije wil begon met de mogelijke incompatibilitieit met determinisme en pogingen het determinisme buitenspel te zetten, gaat het in het debat over mentale veroorzaking om iets anders. De vraag die hier speelt is hoe we kunnen begrijpen dat mentale toestanden (intenties, bijvoorbeeld) ons gedrag veroorzaken *als* mentale toestanden. Uitgangspunt van vrijwel alle filosofen binnen het vakgebied is dat de *mind* gerealiseerd wordt door fysieke systemen als de hersenen. Al onze handelingen kunnen causaal worden verklaard door de werking van onze hersenen in combinatie met fysiologische gegevens over hoe de hersenen onze spieren aansturen. De grote vraag is welke rol de *mind* nog kan spelen in het verklaren van onze handelingen. Als mijn hersenen nu bepalen hoe mijn vingers deze tekst tikken, en als mijn wens deze tekst te tikken gerealiseerd wordt door precies die hersentoestanden die verantwoordelijk zijn voor mijn tik-gedrag, wat voeg ik dan nog toe aan de verklaring van mijn handelingen door te zeggen dat ik *wens* deze tekst te tikken. Dit is het probleem van mentale veroorzaking (een probleem, dus, waarin mijn gedetermineerdheid geen rol speelt).

De discussie over mentale veroorzaking speelt zich af tegen de achtergrond van de vraag of het feit dat de *mind gerealiseerd* wordt door de hersenen impliceert of de *mind* kan worden *gereduceerd* tot de hersenen. Als het antwoord op deze vraag positief is verdwijnt het probleem. Immers, als de *mind* zonder meer hetzelfde is als de hersenen, is er geen verschil tussen de hersentoestanden die mijn handelen veroorzaken en mijn intenties en worden mijn handelingen door mijn intenties veroorzaakt. Het punt is echter dat een meerderheid van filosofen een zogenaamd 'nonreductief fysicalisme' verdedigt dat voortkomt uit het functionalisme. Wanneer de *mind* zich tot de hersenen verhoudt zoals bijvoorbeeld software tot hardware is de *mind* niet tot de hersenen te reduceren: er is altijd de mogelijkheid om een programma anders te realiseren (denk aan Word op de PC versus Word op de Mac) en dus zijn software en hardware *niet* hetzelfde. Een groot deel van het debat over mentale veroorzaking draait er dus om de *mind*-als-software een rol te geven waar het gaat om een verklaring van handelen die (1) onmisbaar is, en (2) niet samenvalt met de causaal-verklarende kracht van onze hersenen-als-hardware (zie Heil en Mele 1993).

3 *Philosophy of mind* en wetenschap

De klassieke ontwikkelingen en probleemstellingen in de *philosophy of mind* zoals die tot nu toe besproken zijn, zijn hoofdzakelijk conceptueel van aard. De invloed van empirische wetenschap op theorievorming blijft in de ontwikkelingen en debatten die hierboven kort zijn aangeduid in feite beperkt tot het streven om filosofische theorieën ten minste niet strijdig te laten zijn met gangbare wetenschappelijke inzichten. In een enkel geval, zoals bij de ontwikkeling van het functionalisme, spelen wetenschappelijke en/of technologische ontwikkelingen een grotere rol.

Deze verhouding tussen filosofie en wetenschap is met name in de afgelopen twee decennia zeer snel aan het veranderen. Een scala aan ontwikkelingen in bijvoorbeeld de neurowetenschappen, de cognitiewetenschappen en de psychologie hebben directe gevolgen voor het specifieke soort theorievorming en conceptualisering dat voorheen een uitsluitend filosofische aangelegenheid was. De grenzen tussen met name de cognitieve neurowetenschappen en de *philosophy of mind* zijn snel aan het vervagen.

In dit tweede deel wil ik zeer beknopt een drietal ontwikkelingen aanstippen die een indruk geven van de mogelijkheden en moeilijkheden die deze gedeeltelijke integratie van vakgebieden met

zich meebrengt. Ik zal beginnen met het noemen van een tweetal ontwikkelingen in de cognitieve en neurowetenschappen die hebben geleid tot een problematisering van de theorie die lange tijd de orthodoxie in de *philosophy of mind* vormde: het psychofunctionalisme. Hier, zo lijkt het, zal de filosofie van de wetenschap moeten leren. Vervolgens zal ik kort stilstaan bij recente, invloedrijke, pogingen van wetenschappers zich te mengen in het debat over de vrije wil. In deze ontwikkeling lijkt een gebrek aan kennis van de filosofische traditie de wetenschappers ten minste gedeeltelijk parten te spelen. Ten slotte zal ik stilstaan bij een debat dat een schoolvoorbeeld lijkt van geslaagde integratie van wetenschappen en filosofie, het debat over de aard van toeschrijving van mentale toestanden aan anderen, ofwel het *theories-of-mind*-debat.

3.1 Ontwikkelingen in de cognitieve neurowetenschappen: psychofunctionalisme onder vuur

Functionalisme is sinds de jaren zeventig van de vorige eeuw dominant in de *philosophy of mind*. Met name het psychofunctionalisme van Jerry Fodor is zeer lange tijd, vanaf 1975 tot in de jaren 1990, van onvoorstelbaar grote invloed geweest. Volgens deze vorm van functionalisme bestaat er zoiets als een door hersenen gerealiseerde variant van wat in computers een programmeertaal heet. Deze hypothese is op veel punten vruchtbaar gebleken voor theorievorming in zowel de filosofie, als de artificiële intelligentie en cognitiewetenschappen. Niettemin lijken een aantal ontwikkelingen de theorie te ondergraven. Hiervan geef ik twee voorbeelden.

Psychofunctionalisme gaf aanleiding tot de ontwikkeling van het klassieke artificiële-intelligentieprogramma volgens welk een computer in principe (maar niet in de praktijk) in staat zou moeten zijn de programmeertaal van de hersenen te simuleren. Een computer zou onze 'denkwetten' moeten kunnen nadoen. Het aanvankelijke enthousiasme voor deze benadering nam af toen bleek dat zo'n 'denkwettenbenadering' leidde tot tamelijk starre cognitieve systemen, die met name onze capaciteit tot leren en flexibel omgaan met onverwachte situaties niet konden simuleren.

Een reactie op deze tegenvallende resultaten was de connectionistische neuralenetwerkenbenadering. In plaats van een computer 'top-down' te programmeren naar het beeld dat het psychofunctionalisme gaf van onze *mind* werd hij geprogrammeerd om netwerken van neuronen te simuleren zoals die in onze feitelijke hersens bestaan.[9] Op die manier, zo was de hoop, zouden 'bottom-up' eenvoudige cognitieve systemen ontstaan waarvan het gedrag meer op dat van ons lijkt in leervermogen en flexibiliteit. En inderdaad ontstonden er systemen die zichzelf dingen konden leren en die flexibel konden omgaan met onverwachte situaties. Het probleem met deze systemen is wel dat ze oneindig veel eenvoudiger zijn dan onze hersenen, zodat ze niet in de buurt komen van het simuleren van echte *minds*.

Een grote stap voorwaarts in deze ontwikkeling was de erkenning van het feit dat hersenen altijd belichaamd zijn; dat ze geëvolueerd zijn om lichamen in fysieke omgevingen te sturen (zie hoofdstuk 4 van dit boek).[10] Het koppelen van connectionisme aan robotica bleek enorm succesvol. Het gedrag van robots die worden aangestuurd door computerkopieën van eenvoudige dieren als kakkerlakken is verbluffend 'biologisch', adaptief, flexibel en in zekere mate 'intentioneel' te noemen. De suggestie die hieruit spreekt is dat het in principe mogelijk zou zijn werkelijk intentionele systemen, *minds*, op een 'bottom-up'-manier te construeren.

De consequenties van deze ontwikkelingen voor het psychofunctionalisme lijken groot. De 'denkwetten' en 'programmeertaal van onze hersenen' die door Fodor worden geïmpliceerd, worden gedegradeerd tot grove modellen of metaforische benaderingen van datgene wat er zich echt in onze hoofden afspeelt. Ons gedrag is in veel situaties misschien goed te interpreteren *alsof* het door een computerprogramma wordt aangestuurd, maar dat betekent niet dat zo'n programma ook echt bestaat.

Een dergelijke conclusie lijkt ook te worden ondersteund door de gedetailleerde kennis die momenteel beschikbaar komt over de feitelijke opbouw en werking van de hersenen. Een scala aan technieken, van verschillende typen hersenscanners tot zogenaamde '*single cell recordings*', maakt duidelijk hoe onvoorstelbaar complex de werking van de hersenen is. In steeds minder lijken de hersenen op de vaste hardware van een computer. Steeds meer blijkt dat individuele mentale vermogens (gezichtsherkenning, conceptueel denken, associëren, etc.) niet door gemarkeerde delen van de hersenen worden gerealiseerd, maar gedistribueerd over verschillende regio's. Steeds meer blijkt dat bij hersenbeschadiging mentale taken door andere delen van de hersenen kunnen worden overgenomen. Steeds meer blijkt dat de feitelijke werking van de hersenen bij gelijke taken per keer verschilt.

Een dergelijke verhouding tussen mentale vermogens en fysieke realisatie daarvan is moeilijk te vergelijken met de statische verhouding tussen hardware en software in een computer. Wederom, zo lijkt het, is de sofware-hardwarevergelijking zoals die door het psychofunctionalisme van Fodor wordt

voorgesteld veel meer een metafoor die *soms* opgaat dan een feitelijke beschrijving van de aard van onze *mind*. Er is niet zoiets als een programma dat onze *mind is* die 'draait' op de hardware van onze hersenen.

3.2 Misverstanden rond de wil

Kan de wetenschap aantonen dat zoiets als een vrije wil of een bewuste wil niet bestaat? Met paragraaf 2.2.4 in het achterhoofd zouden we hier moeten antwoorden: dat hangt ervan af wat we onder 'vrije' of 'bewuste wil' verstaan. Dat de vele verschillende definities die in de filosofie hiervoor zijn voorgesteld niet behoren tot wat er zowel in het dagelijkse taalgebruik als in de neurowetenschappen onder deze begrippen wordt verstaan bleek recentelijk uit de populariteit van een aantal publicaties waarin neurowetenschappelijke data werden ingezet om beweringen over de vrije of bewuste wil te doen.

Achter deze publicaties zitten een groot aantal experimenten. Uit een van de meest in het oog springende daarvan, oorspronkelijk uitgevoerd door Benjamin Libet,[11] blijkt dat wanneer ons gevraagd wordt uit vrije wil op een door ons gekozen moment een knop in te drukken, we ons pas bewust zijn van onze intentie *nadat* onze hersenen een proces in gang hebben gezet dat uiteindelijk leidt tot het indrukken van die knop. Uit dit experiment en een scala aan soortgelijke bevindingen concludeerde Daniel Wegner in 2002 dat zoiets als een bewuste wil een illusie is. Het is niet de intentie waarvan we ons bewust zijn die uiteindelijk de oorsprong is van onze handeling. Integendeel, die bewuste intentie is niet meer dan een bijverschijnsel van het feitelijke causaal effectieve proces.

Hoewel Wegner het heeft over de 'bewuste wil' is zijn stellingname veelal geïnterpreteerd als een positie over de aard van de vrije wil. Niet alleen geven sommige van zijn formuleringen daartoe aanleiding, ook is het zo dat het hierboven beschreven experiment in termen van 'vrije wil' gepubliceerd is. De details van Wegners positie en van de discussie die volgde op zijn publicatie laat ik hier achterwege. Van belang is dat er conclusies werden getrokken over de vrije of bewuste wil uitgaande van een alledaags begrip van die noties dat – hoewel niet helemaal onbereflecteerd – geheel losstaat van en onbeïnvloed is door een rijke traditie van nadenken over de moeilijkheden en valkuilen daarvan. Hoewel publicaties zoals die van Wegner steevast beginnen met een opmerking over de onvruchtbaarheid van filosofische discussie in vergelijking met empirische wetenschap geeft zijn tekst niet heel veel blijk van grondige kennis van die filosofische discussie.

En dat is jammer. Want in feite heeft Wegner het vaker over het vraagstuk van mentale veroorzaking dan over het vraagstuk van de wil. Hetzelfde geldt voor Libet. Voor zover hij het over mentale veroorzaking heeft lijkt zijn positie onterecht te impliceren dat we slechts in bewuste of vrije wil kunnen geloven als we dualist worden: dan zijn *wij* het, *en niet onze hersenen* die ons gedrag veroorzaken. Hier wordt voorbijgegaan aan een groot aantal binnen de filosofie ontwikkelde niet-dualistische theorieën die wel degelijk mentale veroorzaking kunnen verdisconteren. Voor zover hij het over de wil heeft is zijn positie veelal compatibel met een aantal goed uitgewerkte filosofische theorieën daarover die geen afstand doen van vrijheid.

Hoewel de in Wegners boek genoemde experimenten en zijn discussie daarover zonder meer belangwekkend zijn en zeker niet zonder filosofische consequenties, is zijn stellingname ongelukkig geformuleerd. Het debat dat op zijn publicatie volgde laat eens te meer zien hoe belangrijk het is concepten – zoals 'wil', 'intentie', 'bewust' en 'vrij' – die ten grondslag liggen aan experimentele *set-ups* te verhelderen. Een dergelijke verheldering is een taak voor de *philosophy of mind*.

3.3 Vruchtbare interdisciplinaire interactie: het theories-of-mind-*debat*

Wat beide voorgaande paragrafen laten zien is dat integratie van *philosophy of mind*, cognitieve neurowetenschappen en psychologie beter niet daaruit kan bestaan dat filosofen op de stoel van wetenschappers gaan zitten en omgekeerd. Het gaat erom dat de waarde en eigen functie van wetenschap en filosofie onderkend worden; dat het belang wordt gezien van enerzijds wetenschappelijke data voor filosofische theorievorming (wat zijn de feitelijke uitgangspunten voor ons theoretiseren?) en anderzijds filosofische conceptverheldering voor empirische wetenschap (wat is het precies dat we met een experiment onderzoeken?). Een voorbeeld van een samenwerking tussen de verschillende disciplines waarin zonder meer oog is voor de eigenheid van filosofie en wetenschap is het zogenaamde *theories-of-mind*-debat. Ik zal dit debat in iets meer detail behandelen.

Vreemd genoeg is dit debat begonnen noch in de filosofie, noch in de psychologie, maar in de ethologie. In 1978 vroegen Premack en Woodruff zich af of chimpansees in staat waren te herkennen dat andere soortgenoten een eigen *mind* hebben. De terminologie die ze hiervoor gebruikten was

ontleend aan een aantal ontwikkelingen uit de filosofie (aan het analytisch functionalisme van filosofen als David Lewis, maar ook aan het werk van Wilfrid Sellars) en geheel in lijn met het cognitivisme dat in die tijd hoogtij vierde (zie paragraaf 2.1.3). Wat ze zich afvroegen was of chimpansees een zogenaamde *theory of mind* hadden. Daarmee werd niet bedoeld dat apen bewust aan theorievorming zouden doen – de term 'theorie' moet eerder worden gelezen als verzameling 'hypothesen' over de wereld (zich uitend in gedrag) en in dit geval dus over de oorsprong van het gedrag van soortgenoten.

De bevindingen van Premack en Woodruff zijn hier niet relevant. Wat relevant is, is dat hun invloedrijke artikel direct leidde tot methodologische en conceptuele vragen onder filosofen, met name Daniel Dennett en Gilbert Harman.[12] Terwijl deze filosofen het hebben van *beliefs* als paradigmatische mentale toestand beschouwden, vroegen ze zich af hoe te testen is of iemand werkelijk *beliefs* aan een ander kan toeschrijven. Hoe is het begrip '*belief*' te conceptualiseren zodanig dat het binnen de vraagstelling van Premack en Woodruff te operationaliseren is? Het antwoord dat met name Harman op deze vraag gaf werd direct toegepast in de ontwikkelingspsychologie in de vorm van de zogenaamde *false belief test*, een test die tot op heden door velen (maar op dit moment zeker niet meer door iedereen) gezien wordt als bepalend voor het kunnen toeschrijven van een *theory of mind* (ToM) aan iemand. Ik beperk me tot de ontwikkelingspsychologische test (in een van de vele varianten en iets vereenvoudigd): de experimentator legt een snoepje in een van twee kastjes, voor de ogen van een kind. Hij gaat de kamer uit en iemand anders haalt, voor de ogen van het kind, het snoepje uit het ene kastje en stopt het in het andere. Het kind wordt nu gevraagd waar de experimentator zal zoeken als hij terugkomt en een snoepje wil. Bij normale ontwikkeling zal een kind voor ongeveer zijn vierde jaar het kastje aanwijzen waar het snoepje dan ligt; het kind kan nog geen onderscheid maken tussen wat hijzelf gelooft en wat een ander gelooft. Na het vierde jaar zal, bij normale ontwikkeling, het kind geneigd zijn het lege kastje aan te wijzen, hetgeen aangeeft dat het een *belief* aan een ander kan toeschrijven dat niet het zijne is. Het heeft dan een notie van de *mind* van de ander, een vroege ToM.

De term '*theory of mind'* draagt, zoals gezegd, een duidelijk cognitivistisch, functionalistisch karakter. De veronderstelling die in de term verpakt zit, is dat het begrijpen van andermans *mind* en daarmee het voorspellen en verklaren van diens gedrag een kwestie is van het gebruik maken van bepaalde impliciete of expliciete *kennis* over de ander. Die kennis kan zijn aangeboren of in de loop van de ontwikkeling worden verkregen, daarover lopen de meningen onder de aanhangers van deze zogenaamde *theorie-theorie*-benadering uiteen. De theorie-theorie (TT), dat wil zeggen de stelling dat begrip van andermans gedrag en het vermogen dat gedrag te voorspellen mogelijk wordt gemaakt door gebruik te maken van kennis met een hypothetisch of representationeel karakter, was tot halverwege de jaren tachtig zonder concurrentie. Dit zowel in de filosofie en de cognitiewetenschappen (vanwege de dominantie van het functionalisme, met name in de psychofunctionalistische variant) als in de psychologie (vanwege het toen dominante cognitivisme).

Vanaf 1986 komt er verandering in deze situatie met de publicatie van een drietal artikelen die, onafhankelijk van elkaar, een alternatief bieden voor TT: de simulatietheorie (ST) (Gordon 1986; Goldman 1989; Heal 1986). Kort gezegd is het idee achter dit alternatief dat we geen kennis gebruiken voor het begrijpen van andermans gedrag, maar de werking van onze eigen *minds*. Door ons te 'verplaatsen' in de ander, in zowel de situatie als in zo veel mogelijk van de mentale toestanden van de ander en te merken hoe wij zelf in die situatie zouden handelen, zijn we in staat het gedrag van de ander te begrijpen en voorspellen. Natuurlijk wordt door geen van de ST-varianten het proces zo expliciet en bewust voorgesteld als hier beschreven.

De simulatietheorie begint als filosofische theorie maar krijgt al snel varianten in de cognitiewetenschappen.[13] Daarbij moet gezegd worden dat er in de cognitiewetenschappen, vanwege het overheersende cognitivistisch functionalistische karakter daarvan, een duidelijke voorkeur voor TT-benaderingen blijft bestaan. Hetzelfde geldt, vanwege de invloed van de cognitiewetenschappen daarop, voor de psychologie.

In de jaren negentig wanneer ST binnen de filosofie sowieso meer aanhang krijgt komt er steun voor dit alternatief vanuit de neurowetenschappen. Aan het begin van de jaren negentig wordt in Parma het bestaan van zogenaamde spiegelneuronen ontdekt: neuronen die, kort en enigszins onprecies gezegd, op neuraal niveau intentioneel gedrag van soortgenoten spiegelen. De ontdekkers, en met name Vittorio Gallese, speculeren over een functie van deze neuronen bij het herkennen en toeschrijven van intentioneel gedrag aan anderen, en wellicht zelfs een evolutionaire functie bij het ontstaan van taal uit gebaren. In 1998 bundelt Gallese zijn krachten met een van de bedenkers van ST, Alvin Goldman. Spiegelneuronen, zo stellen zij, doen op laag niveau in feite precies wat ST zegt dat iemand doet die het gedrag van een ander probeert te begrijpen in termen van een achterliggende *mind*. De vondst van spiegelneuronen wordt tamelijk breed gezien als steun voor ST.

Er zijn meer neurowetenschappelijke gegevens die richting ST wijzen. Zo blijkt dat patiënten van wie die hersendelen beschadigd zijn die hen in staat stellen basale emoties als angst en walging te

ervaren, niet in staat zijn die emoties te herkennen in de gezichtsuitdrukkingen van anderen (Goldman & Sripada 2005). De suggestie die hieruit spreekt is dat eigen ervaring van deze emoties blijkbaar nodig is voor het toeschrijven daarvan aan anderen, iets dat door ST maar niet door TT wordt voorspeld. (Zie hoofdstuk 6 van dit boek voor gedetailleerde informatie over de rol van de hersenen bij het begrijpen van anderen.)

Deze vondsten ondersteunen ST weliswaar, maar zijn niet doorslaggevend in het debat over ToM. Belangrijk is te onderkennen dat het in deze neurowetenschappelijke data vooral gaat om het toeschrijven van zeer basale intenties en emoties, niet om volledige propositionele attitudes (*beliefs* en *desires*). De dominantie van TT in artificiële intelligentie en cognitiewetenschappen voor wat betreft het modelleren van de toeschrijving van propositionele attitudes is ook niet zonder reden: nog altijd is TT veel beter te modelleren dan ST.

Het ToM-debat is zo verdeeld over filosofie, neurowetenschappen, cognitiewetenschappen en ontwikkelingspsychologie. Maar er zijn meer relevante gebieden van wetenschappelijk onderzoek, zoals met name dat naar het psychiatrische ziektebeeld van autisme. Een belangrijk deel van het spectrum aan autistische stoornissen bestaat uit het onvermogen van de autist de *mind* van de ander te (her)kennen. Inzicht in datgene wat bij de autist ontbreekt, levert daarom ook inzicht in wat niet-autisten wel kunnen waar het gaat om het (her)kennen van de *mind* van anderen. Oorspronkelijk werd autisme gezien als het disfunctioneren van een ToM begrepen in TT-termen (bv. Baron-Cohen 1995). Daarnaast zijn er ST-verklaringen voor autisme waarbij een beroep wordt gedaan op het onderontwikkelde voorstellingsvermogen van de autist. Tegenwoordig wordt ook duidelijk dat het disfunctioneren van spiegelneuronen een zeer belangrijke rol speelt, hetgeen zou pleiten voor een ST-gerelateerde benadering van autisme (zie boven) (Williams e.a. 2001; Oberman e.a. 2005; Dapretto e.a. 2006).

Een andere tak van wetenschap die steeds belangrijker wordt, is de culturele antropologie. Zo zijn er aanwijzingen voor het feit dat de leeftijd van vier jaar niet universeel is voor het slagen voor de *false belief test*. Daaruit zou blijken dat een ToM niet zozeer een neuraal verankerd mechanisme is, zoals in de cognitiewetenschappen veelal wordt aangenomen, maar eerder een cultureel construct.

3.4 Twee visies op filosofisch-wetenschappelijke interactie

Dit hoofdstuk begon met de constatering dat er in takken van wetenschap als de psychiatrie en psychologie veelal een min of meer intuïtief begrip van het mentale, de geest of, in termen van dit hoofdstuk, de *mind* wordt gehanteerd (waarbij ik hier psychiatrie onder de wetenschappen schaar – zie daarover ook hoofdstuk 3 van dit boek). Explicitering van dat begrip is de hoofdtaak van de *philosophy of mind*, en de wijze waarop dat gebeurt is kort beschreven in het eerste gedeelte van dit hoofdstuk. In het tweede gedeelte zijn een aantal moeilijkheden en mogelijkheden van samenwerking tussen *philosophy of mind* en een aantal wetenschappen behandeld. In deze laatste paragraaf wil ik betogen dat er vanuit de traditionele door functionalisme gedomineerde *philosophy of mind* min of meer automatisch een visie op de verhouding wetenschap-filosofie ontstaat die leidt tot problemen zoals beschreven in de paragrafen 3.1 en 3.2. Voor het vruchtbare soort samenwerking tussen filosofie en wetenschappen zoals in het *theories-of-mind*-voorbeeld uit paragraaf 3.3 is een andere manier van kijken naar de verhouding vereist. Dit, zo wil ik verdedigen, zou consequenties kunnen hebben voor toekomst van de *philosophy of mind*.

Kenmerkend voor de *philosophy of mind* van de laatste vijfentwintig jaar van de twintigste eeuw is de functionalistische benadering. Zoals hierboven beschreven is, ontleent die benadering aan het filosofische behaviorisme het idee dat we het begrip ‘*mind*’ en mentale termen als ‘*belief*’ en ‘*desire*’ moeten analyseren. Aan de identiteitstheorieën van de jaren 1950 en 1960 ontleent ze het idee dat die analyse compatibel moet zijn met gangbare wetenschappelijke inzichten. Dat wil zeggen: het begrip ‘*mind*’ zoals de functionalistische analyse dat presenteert moet zodanig zijn dat wetenschappelijk valt uit te leggen hoe de hersenen, of eventueel de hersenen en het lichaam samen, dit implementeren.

Vanuit deze benadering ontstaat een min of meer heldere tweedeling in taken waar het gaat om het verklaren van het verschijnsel ‘*mind*’. De taak van de filosofie is met name begripsanalyse, in dit geval functionele analyse van het begrip ‘*mind*’, en die van de wetenschap verklaring van implementatie. Natuurlijk is deze voorstelling van zaken iets te schematisch (de cognitiewetenschappen en artificiële intelligentie bevinden zich bijvoorbeeld ergens in het midden: deels voortkomend uit de *philosophy of mind* en deels bestaande uit empirische studies verenigen ze beide stappen). Niettemin is ze kenmerkend voor een impliciete maar binnen de filosofie tot voor kort dominante benadering van de interactie tussen filosofie en wetenschap.

Kenmerkend aan deze benadering is dat analyse van het begrip '*mind*' of van mentale termen in principe niet als taak van de wetenschap wordt gezien. In het voor de filosofie gunstige geval betekent dit dat de taak van analyse door de wetenschap wordt uitbesteed. In het voor de filosofie ongunstige geval wordt aangenomen dat we min of meer intuïtief weten wat termen als 'bewustzijn', 'vrije wil' et cetera betekenen. Wetenschappelijke data worden niet of nauwelijks betrokken bij de analyse van mentale *concepten* maar worden slechts relevant geacht voor de *verklaring* van het bestaan van datgene waarnaar deze concepten verwijzen. De interactie tussen wetenschap en filosofie bestaat binnen deze benadering daaruit dat filosofie, mits de relevantie daarvan onderkend wordt, de explananda voor wetenschap levert. Wetenschap op haar beurt legt in zoverre beperkingen op aan filosofie dat wanneer blijkt dat de implementatie van filosofisch-theoretische concepten moeilijkheden oplevert, dit zal moeten leiden tot bijstelling van die concepten.

Deze traditionele visie op de interrelatie tussen *philosophy of mind* en wetenschap is niet onzinnig. Een deel van de cognitieve neurowetenschappen is er op gebaseerd, en het artificiële-intelligentieproject van de jaren 1980 was er grotendeels op gestoeld. Niettemin heeft deze benadering een serieuze tekortkoming, één die toegelicht kan worden aan de hand van de drie voorbeelden van de paragrafen 3.1, 3.2 en 3.3.

Het voorbeeld van paragraaf 3.1 is er een waarin de implementatie van het psychofunctionalistische model van de *mind* zodanige problemen oplevert dat het model herzien lijkt te moeten worden. Maar, en dat is cruciaal, vanuit de bovenbeschreven benadering van de interactie tussen filosofie en wetenschap houdt de wetenschappelijke feedback hier op. Wetenschappelijke data leveren zelf *geen* bijdrage aan de herziening van de functionalistische analyse van de *mind*.

Het voorbeeld van paragraaf 3.2 was er één waarin grotendeels voorbij werd gegaan aan analyse van het begrip 'vrije wil' of 'bewuste wil'. Desondanks waren de conclusies die ten aanzien van die begrippen werden getrokken verstrekkend tegen de achtergrond van de aanname dat de begrippen inhoudelijk vastliggen en niet kunnen worden genuanceerd of verder ingevuld door gebruik te maken van de uiterst relevante wetenschappelijke data. Maar die aanname is een artefact van de bovenbeschreven visie. Waarom zouden we praktisch relevante begrippen als 'vrije wil' (zonder welke ons juridisch systeem bijvoorbeeld niet zou kunnen fungeren) niet aan een nieuwe analyse onderwerpen juist vanuit de inzichten die door Wegner naar voren zijn gebracht?

De fundamentele fout in de hierboven beschreven traditionele benadering van de verhouding filosofie-wetenschap is mijns inziens dat de rol van de *mind*-wetenschappen te beperkt wordt gezien. Het overgrote deel van de wetenschappen die zich met de *mind* bezighouden neemt weliswaar een bepaalde visie op de *realisatie* of *implementatie* van de *mind* als uitgangspunt (meestal: de *mind* wordt door de hersenen gerealiseerd) maar is vervolgens geïnteresseerd in de *werking* ervan. En dat gegeven zou moeten leiden tot de filosofische erkenning van het feit dat ze een bijdrage kunnen en moeten leveren aan de *analyse* van het begrip '*mind*' en van mentale termen als 'bewustzijn' en 'vrije wil'.

Het ToM-debat, dat hierboven summier is beschreven, is een voorbeeld van interactie tussen filosofie en wetenschappen waarin onderkend wordt dat de rol van wetenschappen meer is dan louter de verklaring van implementatie. Omgekeerd krijgt filosofie een bredere rol dan louter die van analyse (denk aan de ontwikkeling van de *false-belief*-test). Het is deze meer liberale opvatting van de verhouding tussen filosofie en wetenschap die het *theories-of-mind*-debat een goed voorbeeld maakt van geslaagde integratie van de twee.

Een van de meest interessante aspecten van dat debat is dat het steeds meer het concept '*mind'* zelf ter discussie stelt: *wat* is het eigenlijk dat we aan anderen toeschrijven wanneer we hun gedrag in intentionele termen proberen te begrijpen? Gaat het eigenlijk wel om het postuleren van een verborgen set mentale toestanden? Of gaat het eerder om het interpreteren van gedrag in psychologische termen waarbij die termen eerder een sociale, talige oorsprong hebben dan dat ze verwijzen naar een verborgen innerlijk? Een deel van het debat verschuift van de vraag of we wel een '*theory* of mind' hebben naar of we wel een 'theory of *mind*' hebben.[14] Daarmee wordt uiteindelijk een deel van de basale aannames achter de traditionele *philosophy of mind* ter discussie gesteld.

3.5 Tot slot

Het is zonder meer mogelijk dat ook binnen de filosofie onder invloed van het ToM-debat de vraag naar de aard en implementatie van de *mind* ondergeschikt wordt gemaakt aan de gezamenlijke wetenschappelijk-filosofische vraag naar de voorwaarden waaronder we mentale toestanden aan anderen en onszelf toeschrijven. Die voorwaarden worden niet langer, zoals in het filosofisch behaviorisme, vanuit de leunstoel bepaald. Empirische data zoals die over autisme of over spiegelneuronen blijken uitermate relevant. Daarmee zou de toekomstige *philosophy of mind* heel goed

kunnen gaan verschillen van de traditionele. Niet alleen worden de grenzen tussen filosofie en cognitiewetenschappen poreus, dat waren ze al. Het zijn vooral de grenzen tussen filosofie en wetenschappen als ontwikkelingspsychologie, klinische psychologie en psychiatrie die meer dan tot nu toe het geval is zouden kunnen vervagen.

Literatuur

Armstrong, D. (1968). *A materialist theory of the mind*. Londen: Routledge & Kegan Paul.
Baron-Cohen, S. (1995). *Mindblindness: an essay on autism and theory of mind*. Cambridge MA: MIT Press.
Boring, E.G. (1933). *The physical dimensions of consciousness*. New York: Century.
Bermudez, J.L. (1998). *The paradox of self-consciousness*. Cambridge MA: MIT Press.
Block, N. (1978). Troubles with functionalism. In W. Savage (red.) *Minnesota Studies in the Philosophy of Science* vol. IX (pp. 261-325). Minneapolis: University of Minnesota Press.
Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy* 4, 73-122.
Carruthers, P. & Smith, P.K. (1996). *Theories of theories of mind*. Cambridge: Cambridge University Press.
Chalmers, D. (1996). *The conscious mind*. New York: Oxford University Press.
Churchland, P. (1982). *Matter and consciousness*. Cambridge MA: MIT Press.
Clark, A. (1997). *Being there*. Cambridge MA: MIT Press.
Dapretto, M., Davies, M.S., Pfeiffer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y. & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*, 28-30.
Davidson, D. (1984). First person authority. *Dialectica*, *38*, 101-112.
Dennett, D.C. (1984). *Elbow room: the varieties of free will worth wanting*. Cambridge MA: MIT Press
Dennett, D.C. (1987). *The intentional stance*. Cambridge MA: MIT Press.
Descartes, R. (1637). *Discourse on the methods of rightly directing one's reason and of seeking the truth in the sciences*, vert. E. Anscombe & P.T. Geach. In *Descartes: Philosophical writings*, 5-58**.** Edinburgh: Nelson (1954).
Fodor, J. (1975). *The language of thought*. New York: Crowell.
Frankfurt, H. (1971). Freedom of the will and the concept of a person. In Robert Kane, *Free will* (pp. 127-144). Malden: Blackwell (2002).
Gallagher, S. (2005). *How the body shapes the mind*. New York: Oxford University Press.
Gallagher, S. & Hutto, D. (in pers). Primary interaction and narrative practice. In Zlatev e.a. (red.), *The shared mind: perspectives on intersubjectivity*. Amsterdam: John Benjamins.
Goldman, A. (1989). Interpretation psychologized. *Mind and Language*, *4*, 161-185.
Goldman, A. & Sripada, C. (2005). Simulationist models of face-based emotion recognition. *Cognition*, *94*, 193-213.
Gordon, R. (1986). Folk-psychology as simulation. *Mind and Language*, *1*, 158-171.
Heal, J. (1986). Replication and functionalism. In Butterfield, J. (red.), *Language, mind and logic*, 135-150. Cambridge: Cambridge University Press.
Heil, J. & Mele, A. (1993). *Mental causation*. New York: Oxford University Press.
Hutto, D. (2008). *Folk-psychological narratives: The socio-cultural basis of understanding reasons*. Cambridge MA: MIT Press.
Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly* 32, 127-136.
Kane, R. (2005). *Free will*. Oxford: Oxford University Press.
Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, *50*, 249-258.
Libet, B. (2003). *Mind and time*. Harvard: Harvard University Press.
Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*, 435-450.
Oberman, L.M., Hubbard, E.M., McCleery, J.P., Altschuler, E.L., Ramachandran, V.S. & Pined, J.A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, *24*, 190-198.
Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, *1* (4), 515-526.
Putnam, H. (1967). Psychological predicates. In Capitan, W.H. & Merrill, D.D. (red.), *Art, mind and religion* (pp. 37-48). Pittsburgh: Pittsburgh University Press.

Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson (red.), *Language, mind and knowledge*, 215-271. Minneapolis: University of Minnesota.
Ryle, G. (1949). *The concept of mind*. Londen: Hutchinson.
Schechtman, M. (1996). *The constitution of selves*. Ithaca: Cornell University Press.
Shoemaker, S. (1975). Functionalism and qualia. *Philosophical Studies*, *49*, 291-315.
Skinner, B.F. (1974). *About behaviorism*. Londen: Jonathan Cape.
Smart, J.J.C. (1959). Sensations as brain processes. In W. Lyons (red.) (1995), *Modern philosophy of mind*, 62-74. Londen: J.M. Dent**.**
Stich, S., Nichols, S. Leslie, A & Klein, D. (1996). Varieties of off-line simulation. In Carruthers, P. & Strawson, P.F. (1962). Freedom and Resentment. In *Freedom and resentment and other essays*, 1-25. Londen: Methuen (1974).
Taylor, C. (1989). *Sources of the self*. Cambridge: Cambridge University Press.
Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*, 433-460.
Van Inwagen, P. (1982). The incompatibility of free will and determinism. In Gary Watson, *Free will* (pp. 46-58). Oxford: Oxford University Press, 1982.
Watson, J.B. (1913/1995). Psychology as the behaviorist views it. In W. Lyons (red.) (1995), *Modern philosophy of mind*, 14-27. Londen: J.M. Dent**.**
Watson, G. (1975). Free Agency. *The Journal of Philosophy*, *72*, 205-220.
Wegner, D. (2002). *The illusion of conscious will*. Cambridge MA: MIT Press.
Williams, J.H.G., Whiten, A, Suddendorf, T. & Perrett, D.I. (2001). Imitation, mirror neurons and autism. *Neuroscience and Biobehavioral Reviews*, *25*, 287-295.
Wundt, W. (1907). On the *Ausfrage* experiments and on the methods of the psychology of thinking, *Psychologische Studien*, *3*, 301-360.

---

[1] De twee belangrijkste teksten hier zijn Putnam (1975) en Burge (1979).
[2] Zie voor een goede inleiding in de problematiek de eerste twee hoofdstukken van Bermudez (1998).
[3] Zie bijvoorbeeld Taylor (1989), maar ook Frankfurt (1971), Watson (1975) en Schechtman (1996).
[4]. Zie Nagel (1974)
[5] Zie Block (1978), Shoemaker (1975), Jackson (1982) en Chalmers (1996).
[6] Van de vele inleidingen in deze problematiek is die van Kane (2005) aan te bevelen.
[7] Zie Strawson (1962).
[8] Locus classicus hier is Frankfurt (1971).
[9] Voor een toegankelijke beschrijving van deze connectionistische benadering, zie Churchland (1982) hoofdstuk 7.
[10] Zie met name Clark (1997).
[11] Zie Libet (2003).
[12] Zie de introductie van Carruthers & Smith (1996).
[13] Zie bijvoorbeeld Stich e.a. (1996).
[14] Zie Gallagher (2005); Gallagher & Hutto (in pers); Hutto (2008).