

ZAKJAPANNERS, SUPERCHINEZEN EN HERSENSCHIMMEN

in de hedendaagse filosofie van de cognitiewetenschappen

Jan Sleutels

1 Inleiding. De geest volgens Gulliver

Op een van zijn beroemde reizen in verschillende verre landen brengt Gulliver een bezoek aan de Groote Academie van Lagado, de hoofdstad van Balnibarbië. De vriendelijke bevolking leidt hem onder meer rond door de afdeling voor theoretische en bespiegelende wetenschappen, waar een merkwaardige vinding wordt gedemonstreerd die naar verluidt “de grootste ignoramus tegen kleine kosten en een weinig lichamelijke inspanning in staat [zal stellen] om zonder de geringste aanleg of studie, boeken te schrijven over wijsbegeerte, poëzie, politiek, recht, wiskunde en godgeleerdheid.” Het is een apparaat van zes meter in het vierkant, waarvan de bovenkant bestaat uit een groot aantal kleine houten kubusjes, onderling verbonden met dunne draden en beplakt met papertjes waarop “al de woorden van hun taal geschreven [staan] met al hun verscheidene verbuigingen, wijzen, tijden en vervoegingen, zonder dat er echter eenige regel of orde in de rangschikking van die woorden gevolgd was.” Voorts zijn de meest “nauwkeurige berekeningen ... gemaakt van de algemeene verhoudingen tusschen de aantallen der partikels, zelfstandige naamwoorden, werkwoorden en andere rededeelen, zooals die in boeken gebruikt worden.” Om het apparaat staan veertig studenten opgesteld. Op bevel van de professor, zo bericht Gulliver ons,

“greep ieder van zijn leerlingen een ijzeren handeltje, waarvan er veertig rondom de bak in de zij-kanten aangebracht waren, en draaide het met één slag om, zoodat de schikking der woorden een algeheele verandering onderging. Toen liet hij zes en dertig leerlingen zachtjes de regels lezen, zooals die in de bak naar boven waren gekomen en waar er groepen van drie of vier woorden gevonden werden, die mogelijk in zinnen konden voorkomen, dan dicteerden de vinders die aan de vier overige leerlingen, die als schrijvers fungeerden. Dit werk werd een keer of vier herhaald en met iedere draai aan de handeltjes werkte het apparaat dusdanig, dat de woorden van plaats verwisselden en de houtblokjes met de andere zijden naar boven kwamen.”¹

1

Deze procedure wordt eindeloos herhaald en heeft inmiddels al verschillende boekdelen vol gebroken zinnen opgeleverd die door de professor nog moeten worden samengevoegd tot volledige zinnen en verhandelingen, uit welk rijk materiaal hij de wereld “een allesomvattende bewerking van de kunsten en wetenschappen” denkt te kunnen presenteren.

Swift liet er weinig twijfel over bestaan wat hijzelf van het idee vond. De stakker! Hoe had hij ook kunnen weten dat zijn verhaal nog eens werkelijkheid zou worden. Heden ten dage worden niet alleen werkelijk zulke apparaten gebouwd, maar zijn wij er zelfs in geslaagd het hele geval in het menselijk hoofd te plaatsen, met de veertig knapen en de professor en al! Ik doel hier natuurlijk op de moderne cognitiewetenschap en het zogenaamde computationele paradigma. Al stelt de cognitiewetenschap zich niet ten doel de wereld “een allesomvattende

* Eerder verschenen in C. Brown, P. Hagoort & T.C. Meijering, red., *Vensters op de geest*, Grafiet, Utrecht, 1990, pp. 252-290.

¹ Jonathan Swift, *Gulliver's reizen naar verschillende verre landen*. Nederlandse bewerking door G. Blom, geïllustreerd door Rein van Looy. J.H. Gottmer & Co., Haarlem, 1940, p. 215.

bewerking van de kunsten en wetenschappen" te presenteren, zij bestudeert wel degelijk wat daarbij allemaal komt kijken, namelijk wat er zich afspeelt in het hoofd van ons mensen, die zelf die kunsten en wetenschappen voortbrengen.

Het vergt een vrij eigenzinnig filosofisch fundament om een project als dit mogelijk te maken, een fundament dat bepaald niet van kritiek gevrijwaard is gebleven. Welke ontologie van de geest hebben wij nodig om een kennismachine in ons hoofd te kunnen postuleren? Kan het postuleren van zo'n machine ooit een afdoende verklaring van mentale processen geven? Bestaat er werkelijk een kenmachine in ons hoofd, of is onze hele geest een hersenschim? Werkt het apparaat met zinnen in een mentale taal? Is er één zo'n apparaat, of zijn er eigenlijk meer van? Wij zullen in het navolgende een aantal van de belangrijkste discussies over de grondslagen van de moderne cognitiewetenschappen de revue laten passeren.

Een van de constantes door de diverse discussies heen, een thema dat hier voortdurend aanwezig zal zijn zonder expliciet te worden besproken (het wordt niet zozeer 'gezegd' als wel 'getoond'), betreft het *naturalistisch schouwtoneel* van de cognitiewetenschappen en haar filosofie. Vanuit een traditioneel-filosofisch (veelal Kantiaans genoemd) perspectief bezien is het een rommeltje. Filosofen moeten zich eigenlijk bezighouden met begripsanalyse, wetenschappers met empirie; filosofen doen het voorbereidende grondwerk en ruimen vervolgens het veld voor het empirisch onderzoek. In de cognitiewetenschap ziet de situatie er echter heel anders uit. Telkens weer zullen wij kunnen vaststellen dat wetenschappelijke en filosofische argumenten en motieven daar bijzonder nauw met elkaar verweven zijn. Filosofische posities ten aanzien van de aard van het mentale hebben niet alleen een conceptueel en metafysisch aspect, maar blijken tevens tastbare methodologische en empirische consequenties te hebben. Zij impliceren een welbepaalde manier van werken in de wetenschappelijke praktijk, en kunnen mede worden getoetst aan het empirisch succes van hun wetenschappelijke implicaties. Evenzo hebben empirisch-wetenschappelijke posities niet alleen een methodologisch en empirisch aspect, maar blijken zij tevens verstrekkende conceptuele en metafysische implicaties te hebben. Wetenschap en filosofie lopen vrijwel vloeiend in elkaar over, empirie en begripsanalyse staan in voortdurende wisselwerking met elkaar.

Voor menig filosoof en wetenschapper geldt dit wellicht als een teken van de onvolwassenheid van de discipline; de cognitiewetenschap bevindt zich nog in een revolutionaire, pre-paradigmatische fase, zo zou men (met een knipoog naar Kuhn) kunnen redeneren, een fase die wordt gekenmerkt door begripsverwarring en *category mistakes*. De filosoof moet er nog zijn filosofisch grondwerk verrichten; daarna zal hij zich kunnen terugtrekken en de wetenschap de wetenschap laten. Misschien is deze Kantiaanse visie op de werkverdeling tussen filosofen en wetenschappers correct, misschien ook niet; er is een alternatief. De schrijver persoonlijk deelt de mening van andere auteurs in deze bundel en opteert voor deze alternatieve, positievere benadering van het naturalistisch schouwtoneel. Vanuit dit oogpunt is de interactie tussen empirie en analyse niet zozeer een noodzakelijk kwaad, als wel een goed dat node gemist kan worden. Laat ik mij hier beperken tot een persuasief argument. Een filosofie die zichzelf situeert in een cognitief vacuüm, in een ivoren toren, lijkt mij bepaald steriel. Hoe zou zij ooit van belang kunnen zijn voor onze normale, alledaagse en wetenschappelijke denkwereld wanneer zij zich niet bedient van argumenten, begrippen en ideeën ontleend aan wetenschap en denken van alledag? Hoe zou zij ooit kunnen overtuigen? Omgekeerd komt het mij voor dat een wetenschap die wars is van filosofie bepaald inert en blind zou zijn. Zij mist conceptuele *body* en dreigt door haar slaafse binding aan blinde feiten de kans op grootschalige theoretische innovaties mis te lopen. Vanuit dit perspectief bezien is de wisselwerking tussen filosofische analyse en empirisch onderzoek, zoals wij die onder meer in de cognitiewetenschappen tegenkomen, een uiterst wenselijk en vruchtbaar goed.

Het naturalistisch schouwtoneel is zelf een empirisch gegeven uit de praktijk van de cognitiewetenschappen. Wellicht zal het Kantiaans anti-naturalisme uiteindelijk zichzelf blijken te logenstraffen. Als (meta-)filosofische visie op de verhouding tussen filosofie en wetenschap

doet het tevens een (het zij toegegeven: moeilijk toetsbare) voorspelling over de toekomst van de cognitiewetenschappen: wanneer het filosofisch grondwerk gedaan zal zijn, zullen de filosofen zich kunnen terugtrekken om plaats te maken voor het echte empirisch-wetenschappelijke werk. Blijft de interactie voortbestaan, zo zou men kunnen redeneren, dan is het anti-naturalisme gefalsifieerd; komt de interactie inderdaad tot een einde, zoals voorspeld, dan is het anti-naturalisme zelf een soort van (meta?-)naturalisme, een filosofie met empirische consequenties! Zolang de oplossing van dit raadsel nog in de boezem der empirie besloten ligt, laten wij het oordeel over het naturalisme in de cognitiewetenschappen vooralsnog graag aan de lezer over.

2 *Functionalisme. Opkomst van het kenapparaat*

Het menselijk hoofd biedt niet zo een-twee-drie plaats aan een kenapparaat à la Gulliver; het vergt een heel bepaalde filosofie om zo'n ding daar onder te brengen. Swift vond het een belachelijk idee; de moderne cognitiewetenschap is op dat idee gebouwd. De onderliggende gedachte wordt gevormd door een stroming in de *philosophy of mind* die luistert naar de naam *functionalisme*. Het functionalisme kan worden gezien als een reactie op twee van zijn minder in-schikkelijke voorgangers in de filosofie van de geest, het behaviorisme en het reductief materialisme. Kort samengevat komt het neer op het inzicht dat psychologische termen zoals mening, verlangen, pijn, geheugen, aandacht en betekenis niet hoeven te worden beschouwd als een soort van steno voor de 'eigenlijke' beschrijving van het mentale in termen van hetzij neurofysiologie, hetzij van uitwendig waarneembare gedragingen.

Om met het laatste te beginnen, onze alledaagse manier van spreken over het mentale is volgens het *behaviorisme* niet meer dan een (voor alledaagse doeleinden heel nuttige) steno-grafische afkorting voor veel langere (en onhandige en saaie) beschrijvingen van gedragsdisposities. Wanneer de leek zegt dat U en ik allebei van postmoderne kunst houden, dan betekent dat volgens de behaviorist eigenlijk dat wij een bepaalde wetenschappelijk articuleerbare verzameling gedragsdisposities gemeen hebben, waaronder een zekere neiging tot het kopen van Italiaanse peper- en zoutstelen, het opzetten van theewater in ketels van Aldo Rossi, alsmede een neiging om daar vooral ook veel over te discussiëren.

Dat het niet goed afliep met het behaviorisme zal de meeste lezers bekend zijn. Het intuïtief aannemelijke inwendige aspect van mentale toestanden en processen blijkt node te kunnen worden gemist. Neem het bovenstaande voorbeeld. Stel dat U zwijgzaam van aard bent en ingetogen leeft; U koopt geen Italiaans design en U praat nooit over peper- en zoutstelen. Volgens het behaviorisme zouden wij met geen mogelijkheid een gemeenschappelijke liefde voor postmoderne kunst kunnen hebben. Een tweede bezwaar is nog ernstiger. Het werd de behavioristen al spoedig duidelijk (en ook hun critici ontging het niet) dat het onmogelijk is om mentale toestanden en processen te specificeren in termen van een eindige, niet-circulaire verzameling objectief waarneembare omstandigheden en gedragingen. De reeks van conditionele uitspraken die nodig zou zijn om genoemde voorliefde voor postmoderne kunst te analyseren, bijvoorbeeld, zou vermoedelijk oneindig lang moeten zijn, aangezien er geen eindige, niet-circulaire manier is om de talloze omstandigheden waaronder en wijzen waarop de dispositie kan worden gerealiseerd van tevoren te specificeren of samen te vatten (anders dan onder die mentalistische noemer 'houdt van postmoderne kunst' zelf, die juist moet worden geanalyseerd). Alle pogingen die in deze richting worden ondernomen blijken bij nader inzien ofwel onvolledig te zijn (de lijst bevat een eindig aantal conditionele uitspraken, maar gaat aan een oneindig aantal onvoorziene omstandigheden en/of gedragingen voorbij), ofwel circulair (de lijst bestrijkt weliswaar alle mogelijke gevallen, maar moet daartoe zelf weer gebruik maken van ongeanalyseerde mentalistische termen).

Volledigheidshalve zij hierbij opgemerkt dat het hier bekritiseerde zogenaamd 'logisch' behaviorisme onderscheiden dient te worden van het zogenaamd 'methodologisch' behaviorisme. Het eerste is een boude metafysische (of zo men wil: conceptuele) hypothese over de aard van mentale processen (respectievelijk de analyse van mentalistische categorieën). Het tweede is

een prudent methodologisch voorschrift voor de introductie van *nieuwe* mentalistische categorieën in de wetenschap: introduceer geen nieuwe begrippen tenzij je ze operationeel weet te definiëren in termen van gedrag en omstandigheden. Dit voorschrift staat nog steeds met grote letters boven de ingang van de cognitiewetenschap. Het geldt er als een gezonde rem op elke vorm van drieste maar oncontroleerbare speculatie die het uiteindelijke doel van de psychologie, namelijk de etiologie van het menselijk gedrag, uit het oog dreigt te verliezen. Overigens is de oorsprong of motivatie van de sterke en de zwakke vorm van behaviorisme dezelfde. Beide zijn geboren uit een afkeer van de traditionele subjectivistische en introspectionistische psychologie. In een poging de psychologie tot een 'harde' wetenschap te maken en haar object van het imago van wetenschappelijke ongrijpbaarheid te bevrijden, werd een alternatief gezocht voor de klassieke methode van subjectieve, oncontroleerbare introspectierapporten. Deze verstrengeling en onderlinge beïnvloeding van wetenschappelijke en filosofische motieven zullen wij in de cognitiewetenschap nog meermalen tegenkomen.

De andere voorganger van het functionalisme staat bekend als *reductief materialisme*, *identiteitstheorie*, of ook *'central state' materialisme*. Volgens deze theorie zijn mentale toestanden in feite bepaalde fysische toestanden van de hersens, of iets preciezer gezegd, is elke soort van mentale toestand of mentaal proces identiek met een zekere soort van fysische (en met name neurofysiologische) toestand of proces in het centrale zenuwstelsel. De identiteitstheorie voorziet een reductie van de psychologie tot de neurofysiologie, net zoals het in de geschiedenis van de wetenschap al meermalen is voorgekomen dat de ene wetenschap door een andere, meer elementaire werd gereduceerd. Als klassiek voorbeeld hiervan geldt de reductie van de fenomenologische thermodynamica tot de statistische thermodynamica, waarbij het sleutelbegrip 'temperatuur' van de eerste wetenschap werd gereduceerd tot het begrip 'gemiddelde kinetische energie van moleculen' in de tweede. Als het ons nu inderdaad menens is met de gedachte dat de mens een biologisch organisme is net als alle andere, lijkt het voor de hand te liggen om te zoeken naar een reductie van het geestelijke tot het biologische. Deze filosofische theorie over het mentale is al net zomin uit de lucht komen vallen als het behaviorisme; ook de identiteitstheorie heeft een duidelijke wetenschappelijke inspiratie. Zij is geënt op het onmiskenbare succes van de materialistische benadering van het verschijnsel *Homo sapiens* in de biologische wetenschappen. Meer in het bijzonder is het de neuropsychologie geweest die in de eerste helft van deze eeuw, op basis van het rijke materiaal dat twee wereldoorlogen hebben voortgebracht, baanbrekende ontdekkingen heeft kunnen doen met betrekking tot het uitvallen van specifieke cognitieve functies bij lokaal hersenletsel. Deze bevindingen gaven rond het midden van de eeuw aanleiding tot een zeker neurofysiologisch optimisme; er bleek een zeer nauwe samenhang te bestaan tussen hersens en geest, zodat het identificeren van de specifieke neurale correlaten van mentale processen en vermogens alleen maar een kwestie van tijd leek te zijn.

Toch bleek deze vorm van identiteitstheorie alras te sterk te zijn. Wil ik mij in dezelfde mentale toestand bevinden als bijvoorbeeld Gorbatsjov, dan zou dat per definitie betekenen dat mijn hersens precies dezelfde configuratie van neurale activiteit moeten vertonen als de zijne. Maar in ons alledaags spreken over het mentale lijken wij niet zulke strikte eisen te stellen. En onze hersens lijken dat al evenmin te doen: uit onderzoek is gebleken dat ons zenuwstelsel bijzonder 'plastisch' is, d.w.z. dat bij uitval van een bepaald deel van de hersens, bijvoorbeeld als gevolg van een beroerte, de functie van het beschadigde weefsel althans ten dele kan worden overgenomen door ander weefsel elders in het brein. Berucht in dit verband is de zogenaamde *law of mass action*, in de jaren '40-'50 geformuleerd door de neuropsycholoog Karl Lashley. Op zoek naar de specifieke neurale localisatie van bepaalde geheugeninhouden kwam Lashley tot de ontdekking dat er geen aantoonbare samenhang bestaat tussen de *plaats* van een hersenletsel en het verlies van bepaalde herinneringen. Hij stelde experimenteel vast dat (schrik niet, hij werkte met ratten) de enige toelaatbare conclusie hier moet luiden dat de omvang van het geheugenverlies evenredig is met de massa van het beschadigde weefsel (vandaar: *mass action*). Elke portie hersenweefsel, aldus Lashley, kan in beginsel verantwoordelijk worden gesteld voor

elk van de diverse cognitieve vermogens en gedragsrepertoires, een eigenschap waarvoor hij de term *equipotentiteit* introduceerde. Het is zonneklaar dat de filosofische identiteitstheorie in het licht van deze bevindingen nauwelijks nog houdbaar is. Empirisch gezien blijkt de cognitieve organisatie van het zenuwstelsel plastisch en equipotentieel te zijn; dezelfde mentale toestanden en vermogens kunnen worden gerealiseerd in verschillende neurofysiologische toestanden en structuren.²

Een laatste en meer filosofisch georiënteerde tegenwerping tegen het reductief materialisme is dat het bij voorbaat uitsluit dat bepaalde wezens een geest kunnen bezitten, terwijl wij toch intuïtief geneigd zijn om die mogelijkheid open te houden. Wij kunnen hier denken aan mensen (of andere wezens) met beschadigde hersens, met abnormale hersens, met niet-menselijke hersens, of zelfs aan wezens *zonder* hersens (bijvoorbeeld computers en Marsmannetjes). Voor zover het zenuwstelsel van elk van deze wezens afwijkt van dat van een normale mens, zou het volgens (de boven beschreven vorm van) het reductief materialisme onmogelijk zijn dat zij een geest hebben zoals die van de mens, respectievelijk dat zij dezelfde mentale processen en toestanden kennen als een normale mens. Deze vorm van materialisme houdt er duidelijk een veel te bekrompen en chauvinistische opvatting over het mentale op na, die de cognitieve organisatie van de eigen soort (een dwarsdoorsnee van *Homo sapiens*) tot absolute norm van het mentale verheft.

Tegen de achtergrond van deze problemen met het behaviorisme en de identiteitstheorie werd in de zestiger jaren het idee geboren dat mentale toestanden niet moeten worden geïdentificeerd met ofwel gedragstoestanden ofwel toestanden van het zenuwstelsel, maar met *functionele* toestanden van het zenuwstelsel. Het *functionalisme* geeft eigenlijk allebei een beetje gelijk. Enerzijds neemt het van het behaviorisme de gedachte over dat mentale toestanden een soort van verbinding zijn tussen de input en de output van een organisme, waarbij het er niet toe doet hoe die verbinding fysisch gezien precies gerealiseerd wordt. En anderzijds stemt het functionalisme in met de identiteitstheorie dat mentale toestanden een inwendige realiteit hebben en inwendig met elkaar samenhangen.

Het functionalisme definieert mentale toestanden en processen in termen van hun functie, d.w.z. in termen van hun causale rol als bemiddelaar tussen stimulus, respons en (andere) mentale toestanden en processen. In wezen is het functionalisme ontologisch gezien neutraal: het is even goed verenigbaar met een dualistische of een idealistische ontologie als met een fysicistische. In de woorden van Hilary Putnam, een van de vroege vaders van het functionalisme, is het psychologisch gezien volkomen irrelevant van welk spul wij gemaakt zijn, of het nu zielestof is of Zwitserse kaas (een van zijn pikantste suggesties), protoplasma of siliciumchips.³ Het punt waar het hierbij natuurlijk om draait is dat een en dezelfde functie op oneindig veel manieren kan worden gerealiseerd in oneindig veel substraten; psychologisch gezien is de aard van het substraat onbelangrijk, zolang het maar de goede functionele eigenschappen heeft. Dit aspect van *meervoudige realiseerbaarheid* wordt gezien als een van de grootste winstpunten van het functionalisme. Het spreekt voor zich dat het niet alleen geldt voor silicium en Zwitserse kaas, maar ook voor verschillen in hersenstructuur van individu tot individu.

Een ander fort van het functionalisme, en ten nauwste samenhangend met het aspect van meervoudige realiseerbaarheid, is zijn enorme conceptuele flexibiliteit in de beschrijving van systemen. De functionele analyse van toestanden en processen kan aangrijpen op een spec-

² Lashley's krasse theorie van *mass action* en equipotentiteit wordt heden ten dage in de neurowetenschap niet meer in deze sterke vorm verdedigd. Onze inzichten in de localisatie van neuroanatomische structuren zijn, Lashley's pessimistische bevindingen ten spijt, sterk toegenomen (zie bijvoorbeeld Luria 1973). Dat betekent overigens nog geen weerlegging van een betrekkelijke mate van plasticiteit van het zenuwstelsel; het bovengenoemde empirisch-wetenschappelijk argument tegen (een sterke vorm van) de identiteitstheorie moet in dit licht weliswaar enigszins worden gematigd, maar blijft in beginsel onverlet.

³ Putnam (1975), p. 302

trum van niveaus van beschrijving en verklaring. Een koffiezetapparaat, bijvoorbeeld, zou op het meest globale niveau van analyse kunnen worden gekarakteriseerd als een functie die gemalen koffiebonen en water als input neemt en koffie als output geeft. De nadere implementatie van deze functie wordt in deze grove karakteristiek nog open gelaten, zodat percoleermachines, filtreermachines en espressoapparaten, hoe verschillend zij ook zijn, allemaal deze zelfde functie implementeren. De functionele structuur van deze machines kan echter ook op een 'lager', meer fijngegraasd niveau van analyse worden beschreven, zodanig dat kan worden onderscheiden tussen de genoemde drie types van koffiezetapparaten. Bovengenoemde functie van bonen en water naar koffie wordt daarbij geanalyseerd in een aantal subfuncties, elk met een eigen input/output-karakteristiek (filters, percolators, drukvaten, verwarmingselementen, en zo voort). Op het niveau van deze constellaties van subfuncties beschouwd zijn de drie soorten koffiezetters niet langer equivalent. Het spreekt voor zich dat wij op deze manier doorredenerend, door de subfuncties op hun beurt te analyseren in sub-subfuncties, steeds meer bijzondere soorten van functioneel equivalente machines kunnen onderscheiden, tot op het niveau van de functies van de allerkleinste onderdelen toe.

Net zoals het functionalisme een spectrum van verklarings- en beschrijvingsniveaus voor koffiezetapparaten toelaat, introduceert het ook eenzelfde spectrum van niveaus voor de analyse van de werking van computers, cognitieve functies en zenuwstelsels, reikend van de meest globale functionele karakteristieken (het vertonen van 'intelligent' gedrag, of het slagen voor de 'Turing-test'), via allerlei tussenliggende niveaus (procedures in computerprogramma's, globale functies van delen van het zenuwstelsel, specifieke deelvermogens van de geest, bijvoorbeeld voor gezichtswaarneming en taal), tot op het niveau van de functionele eigenschappen van de kleinste onderdelen (neuronen in het zenuwstelsel, flipflops in de computer). Op elk van deze niveaus kan een functionalistische karakteristiek van het bestudeerde systeem worden gegeven; hoe fijner de beschrijving van de functionele organisatie, des te kleiner de groep van functioneel equivalente systemen. Een computer en een mens die op een betrekkelijk grof niveau kunnen worden beschreven als instanties van dezelfde functie, bijvoorbeeld als natuurlijke taal-verwerkers, hoeven op een lager niveau van beschrijving niet functioneel equivalent te zijn; zij maken wellicht gebruik van verschillende procedures, bijvoorbeeld voor lexicale verwerking, en zij zullen ongetwijfeld op het allerlaagste niveau van hardware en wetware volkomen verschillend georganiseerd zijn. Welk niveau van analyse cognitief gezien het interessantst is, is een controversiële kwestie waar wij nog op zullen terugkomen. Hier kan worden volstaan met de vaststelling dat de noties van *functionele structuur* en *functionele equivalentie* rekbare begrippen zijn. De boven besproken identiteitstheorie van lichaam en geest is in wezen een randgeval van het functionalisme: de extreme hypothese dat alle denkende wezens op het allerlaagste niveau van functionele beschrijving, namelijk dat van de eigenschappen van afzonderlijke neuronen, functioneel equivalent zijn.

Al moge het functionalisme op zich ontologisch neutraal zijn, in de moderne cognitiewetenschap en haar filosofie wordt het (uiteraard) als regel fysicistisch geïnterpreteerd. Functionalisten verwerpen weliswaar de traditionele *soort-identiteitstheorie* die stelt dat iedere soort van psychologische toestand identiek is met een bepaalde soort van neurofysiologische toestand, zij blijven niettemin vasthouden aan een zwakkere vorm van materialisme, de zogenaamde *teken-identiteitstheorie*.⁴ Deze zwakke identiteitstheorie stelt dat elke instantie van een mentale

⁴ Het onderscheid tussen teken- en soortidentiteitstheorie wordt hier gebruikt als vertaling van het Engelstalige onderscheid tussen theorieën van *token-token identity* respectievelijk *type-type identity*. Tokens en types verhouden zich als individuen en predicaten of categorieën van individuen. Tot het type *kat* behoren de individuele token-katten Minos, Smintheus en Pietermel. Zo behoren ook concrete, gedateerde instanties (d.w.z. tokens) van een pijngewaarwording of van een (aangename) herinnering aan de *Bains Douches* te Parijs tot het type of de categorie van pijngewaarwording of herinnering aan de *Bains Douches* te Parijs; concrete, gedateerde token-toestanden van het zenuwstelsel behoren op dezelfde wijze tot diverse neurofysiologische types, bijvoorbeeld tot activiteit van C-vezels, of tot een bepaald soort configuratie van activiteit in de linker parietale cortex.

soort identiek is met een instantie van een neurale soort, maar niet noodzakelijk steeds van dezelfde neurale soort. Aangezien reductie van theorieën een verschijnsel is dat zich afspeelt op het niveau van de algemene categorieën van de theorieën, is de teken-identiteitstheorie een vorm van niet-reductief materialisme. Met de afwijzing van soort-identiteiten wordt doorgaans de stelling verbonden dat de cognitiewetenschap zich afspeelt op een zelfstandig beschrijvingsniveau, onderscheiden van en onherleidbaar tot dat van de diverse fysische wetenschappen, inclusief de neurofysiologie. Cognitiewetenschap en neurofysiologie houden zich weliswaar bezig met een en dezelfde werkelijkheid, die van onze geest/hersens, maar delen die als het ware op verschillende manieren in. De verklarende en beschrijvende categorieën van de cognitiewetenschap corresponderen daarom niet met die van de neurowetenschappen. Als de wetenschap die zich met de *functies* van substraten bezighoudt, beschikt de cognitiewetenschap over een eigen object en eigen wetten die niet herleid kunnen worden tot die van de fysische wetenschappen waarin de vele uiteenlopende *substraten* van deze functies worden bestudeerd. Deze 'autonomie'-these over de relatie van psychologie en neurofysiologie is een van de meest opvallende en meest ingrijpende voorbeelden van de nauwe interactie tussen filosofie en praktijk van de cognitiewetenschappen.⁵

3 Computationalisme. Demonologie van de zakjapanner

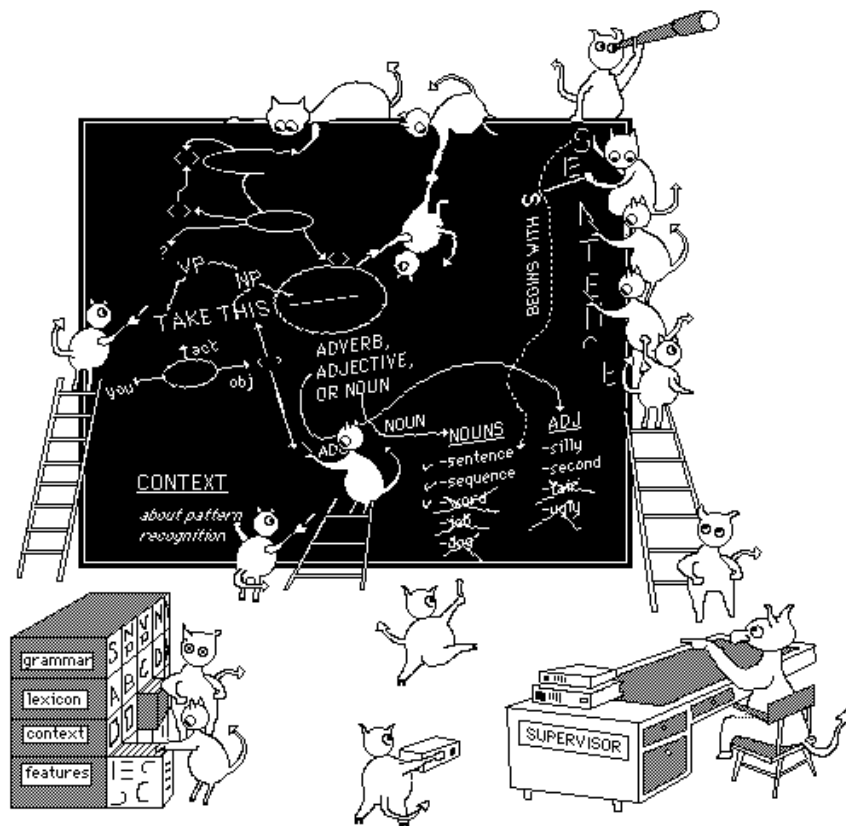
Waar blijft nu die machine van Gulliver? Het functionalisme leert ons dat het mentale gedefinieerd wordt door een bepaalde functionele organisatie, d.w.z. door zijn functie als bemiddelaar tussen een bepaalde input (informatie uit de buitenwereld en/of gegeven mentale toestanden) en een bepaalde output (gedrag en/of nieuwe mentale toestanden). Dit betekent dat de cognitiewetenschapper een aanmerkelijk grotere vrijheid van theoretiseren over het mentale heeft dan onder het oude regime van behaviorisme of reductief materialisme werd toegelaten. Hij hoeft zijn speculaties immers noch te beperken tot uitwendig waarneembare stimuli en responsen, noch tot neurofysiologisch geïdentificeerde hersenstructuren, al vinden wij beide restricties ook onder het functionalisme terug als een soort van marginale constraints op het onderzoek; de cognitiewetenschap moet uiteraard relevant zijn voor de verklaring van uitwendig waarneembaar gedrag, en mag geen entiteiten postuleren die neurofysiologisch gezien onmogelijk zijn. Binnen die grenzen is de moderne psychologie echter volkomen vrij in haar theorievorming, en kan zij alle mentale entiteiten postuleren die zij denkt nodig te hebben voor de verklaring van onze cognitieve vermogens. Hoe die entiteiten precies worden voorgesteld is onbelangrijk, zolang hun functionele eigenschappen maar vaststaan. Kort en goed: in de cognitiewetenschap wordt een functioneel gedefinieerd kenapparaat gepostuleerd.

Laten wij de moderne psycholinguïstiek als voorbeeld nemen. Als specialisme binnen de cognitiewetenschap probeert zij een speciaal deel van onze cognitieve vermogens te verklaren, namelijk ons taalvermogen. Daartoe postuleert zij een speciaal soort apparaat in het hoofd, gedefinieerd door een aantal functionele eigenschappen die te zamen onze linguïstische vermogens moeten verklaren. Zo'n taalmachine omvat in het typische geval onder meer een eenheid voor het ontleden van de zintuiglijke invoer (patroonherkenning, spraakanalyse), die haar resultaten doorgeeft aan diverse andere verwerkingseenheden voor de analyse van de morfologische, lexicale, syntactische en semantische eigenschappen van de input. Al deze onderdelen zijn met elkaar doorverbonden en wisselen informatie uit. Bovendien staan zij in verbinding met diverse geheugenbanken waarin de linguïstische kennis ligt opgeslagen die deze eenheden nodig hebben voor het volbrengen van hun taak. De bediening van dit geheel komt op rekening van een bende kleine duiveltjes of homunculi, die ervoor zorgen dat er op het juiste moment op de knoppen wordt gedrukt, dat de hendels en schakelaars goed staan afgesteld, die de wijzerstand opnemen en de geheugenbanken raadplegen, en zo voort.

⁵ Voor een klassieke verwoording van deze methodologische en metafysische autonomie, zie bijvoorbeeld Fodor (1981), pp. 127vv.

Laten wij ter illustratie van de overeenkomst tussen Gullivers oude taalmachine en het moderne kenapparaat een voorbeeld nemen uit een moderne inleiding in de cognitieve psychologie, het nog steeds populaire handboek *Human information processing* van Lindsay en Norman.⁶ In figuur 1 zien wij een deel van de menselijke taalprocessor voorgesteld als een soort kantoor met een groot schoolbord, een bende duiveltjes, archiefkasten en nog enkele parafernalia meer.⁷ Een verrekijkende verkenner boven op het bord geeft de binnenkomende visuele informatie door aan een handvol patroonherkenners die elk gespecialiseerd zijn in het herkennen van bepaalde onderdelen of aspecten van geschreven letters. Andere duiveltjes proberen inmiddels te achterhalen welk woord er op het ogenblik wordt gelezen. Weer andere duiveltjes zijn druk in de weer met de uitwerking van de syntactische structuur van de (tot dusver onvoltooide) zin; aan de hand van hun tussenresultaten wordt er al een selectie gemaakt van mogelijke kandidaten voor het volgende woord. Bij het verrichten van hun specifieke taken kunnen de duiveltjes niet alleen met elkaar overleggen, maar ook door middel van boodschappenduiveltjes informatie betrekken uit de diverse afdelingen van het lange termijn-geheugen, waarin zich archieven bevinden voor morfologische, syntactische, semantische en andere linguïstische kennis. Het hele gebeuren wordt gadeslagen en gecoördineerd door een centrale opzichter die sterk doet denken aan Gullivers professor in de bespiegelende wetenschap.

Hiermee is de kous evenwel nog niet af. Of het functionalisme nu wel of niet *toestaat* het bestaan van zo'n cognitieve machine in het hoofd te postuleren, dat zegt nog niets over de



8

Afbeelding 1

Deel van de taalprocessor, verantwoordelijk voor de verwerking van geschreven zinnen

⁶ Lindsay & Norman (1977) (2nd ed.).

⁷ Naar Lindsay & Norman (1977), figuur 7-7, p. 288.

vraag of dat ook een goede *verklaring* van onze linguïstische vermogens zou zijn. Op het eerste gezicht lijken wij met het postuleren van zo'n machien inderdaad weinig op te schieten, omdat wij in een apert circulaire verklaring dreigen terecht te komen. De processen en vermogens die moeten worden verklaard worden simpelweg toegeschreven aan een of ander inwendig mechanisme of subsysteem, zodat de vraag zich herhaalt ten aanzien van dit subsysteem (*petitio questionis*, in dit geval ook bekend als *homunculus fallacy*). Voorts zou de manoeuvre ervan kunnen worden beticht dat zij een 'category mistake' maakt, omdat zij probeert om processen en vermogens die eigenlijk thuishoren op een hoger verklarings- en beschrijvingsniveau (namelijk dat van de volledige persoon) toe te schrijven aan een onderdeel of subsysteem op een lager niveau.

Of het beroep op een kenapparaat in de cognitieve psychologie verklarende waarde heeft dan wel een drogredenering inhoudt, hangt ervan af hoe dat apparaat verder geanalyseerd wordt. Als wordt volstaan met de blote bepaling dat het systeem als geheel, of in een van zijn delen, de te verklaren eigenschappen bezit, en daarmee basta, dan is er natuurlijk geen sprake van een echte verklaring. Maar als het systeem wordt geanalyseerd in subsystemen, die elk een deelfunctie van het moedersysteem vervullen, en als de werking van deze subsystemen, langs dezelfde weg dooranalyserend, uiteindelijk kan worden begrepen in termen van uiterst elementaire mechanismen zoals flipflop-schakelaars, transistors of neuronen, dan schuilt er een niet te onderschatten verklarende waarde in het idee. Zoals de filosoof Daniel Dennett het onder woorden bracht, "homunculi are only bogeymen if they duplicate entire the talents they are rung in to explain... If one can get a team or committee of relatively ignorant, narrow-minded, blind homunculi to produce the intelligent behavior of the whole, this is progress."⁸

Welnu, die analyse van functies en subfuncties in subfuncties en subsubfuncties is precies het oogmerk van de cognitieve psychologie. De filosofische theorie achter deze aanpak staat bekend als het *computationalisme* en kan worden beschouwd als een bijzondere versie van het functionalisme, toegespitst op de eisen waaraan een adequate oorzakelijke *verklaring* van mentale vermogens en processen zou moeten voldoen. Een vermogen wordt natuurlijk niet verklaard door gewoon te postuleren dat er een machine voor bestaat; dat is circulair. Maar wanneer wij ook precies kunnen specificeren *hoe* die machine werkt, dan is er wel degelijk sprake van verklaring. De computationele benadering probeert nu juist de details daarvan te achterhalen. Zij stelt zich tot doel de werking van het kenapparaat als het ware te herleiden tot die van een heleboel aaneengeschakelde en samenwerkende zakjapanners (of andere apparaten die niets mysterieus meer hebben). Op grond van een 'top-down' analyse van de functies van onze cognitieve organisatie wordt geprobeerd de werking van het kenapparaat uiteindelijk te begrijpen in termen van de meest elementaire schakelingen, computationele procedures voor het verwerken van formeel gedefinieerde symbolen.⁹

De benadering van het computationalisme wordt gevoed door een inspirerende analogie met de moderne (digitale) computer. Net zoals de werking van een computer op gebruikersniveau (bijvoorbeeld schaakspelen, of het besturen van een lasrobot) uiteindelijk kan worden gespecificeerd in termen van in machinetaal geschreven procedures voor het verwerken van louter enen en nullen, zo zou ook de werking van de menselijke geest in laatste instantie moeten kunnen worden begrepen in termen van procedures voor het manipuleren van de enen en nullen in de machinetaal van de geest. De analogie kan zelfs vrij letterlijk worden genomen. In het vakgebied van de Artificiële Intelligentie (AI) wordt daadwerkelijk onderzocht hoe een computer kan worden geprogrammeerd die bepaalde menselijke cognitieve vermogens 'simuleert' of 'emuleert' (bijvoorbeeld taalproductie en -verwerking, patroonherkenning, interpretatie van 3D-beelden, enzovoort). En op het gebied van de neuropsychologie wordt onderzocht hoe die 'machinetaal van de geest' zich verhoudt tot de bouw en werking van het concrete menselijke brein.

⁸ Dennett (1978), p. 123.

⁹ Voor een meer gedetailleerde schets van het computationalisme in de klassieke cognitiewetenschap en in het nieuwe connectionisme, zie de bijdrage van Sleutels & Geurts, *Knopen en connecties*.

Het computationalisme is een van de onbetwiste hoekstenen van de moderne cognitiewetenschap.¹⁰ Zonder deze grondslag zou de functionalistische ontologie van de geest veel van haar aantrekkelijkheid verliezen, juist vanwege de bovengenoemde problemen met betrekking tot haar verklarend vermogen. Hiermee zijn wij er echter nog niet. Het computationalisme heeft weliswaar de *richting* aangegeven waarin verklaringen gezocht kunnen worden, maar daarmee zijn die *verklaringen zelf* nog niet gegeven. Het computationalisme legt als het ware een agenda voor verder onderzoek vast, en belooft dat dit onderzoek vruchtbaar zal zijn. Of die belofte ook werkelijk waargemaakt kan worden, is een zaak die enkel door empirisch onderzoek zal kunnen worden beslist. Niet alleen de wortels van de filosofische theorie van het computationalisme liggen derhalve op het terrein van wetenschap en techniek (namelijk bij de ontwikkeling van de digitale computer), maar ook de vruchten die wij van haar mogen verwachten.

4 *Van zakjapanner tot superchinese?*

Merkwaardig genoeg zijn het vooral de gedoodverfde winstpunten van het functionalisme die kritiek hebben uitgelokt. Wij hebben eerder gezien dat het functionalisme werd geprezen om zijn bevrijding van het mentale uit de houdgreep van behaviorisme en identiteitstheorie. Met name de versoepeling van de eisen die de identiteitstheorie aan het mentale stelde werd als een belangrijke verbetering aangemerkt. Waarom zou je per se bepaalde zenuwcellen moeten hebben om pijn te kunnen voelen? Waarom zou je überhaupt (een bepaald soort) hersens moeten hebben om te kunnen denken? Nu zijn er echter critici die betogen dat deze versoepeling het functionalisme uiteindelijk noodlottig wordt. De strakke eisen voor de identificatie van mentale toestanden en processen worden door het functionalisme vervangen door nieuwe, minder strakke criteria. Ergens wordt een grens getrokken tussen wat wel en wat niet als een geest zal gelden. Maar welke criteria het functionalisme ook kiest, zo betogen de critici, steeds zullen wij kunnen aandringen dat die criteria op hun beurt ofwel te soepel zijn ofwel te strak. In de woorden van *Urheber* Ned Block, zal elke vorm van functionalisme ofwel te 'liberaal' zijn ofwel te 'chauvinistisch', door ofwel een geest toe te dichten aan dingen die dat niet verdienen, ofwel een geest te onthouden aan dingen die er wel een hebben. En of het functionalisme nu te gul is met de geest of te krenterig, in beide gevallen worden onze spontane intuïties over wat mentaal is met voeten getreden.

De redenering achter deze tegenwerping loopt ongeveer als volgt. Stel dat (soorten van) psychologische toestanden worden geïndividueerd in termen van hun input/output-functies, zoals het functionalisme voorstelt. Twee systemen zullen zich dan per definitie in dezelfde soort psychologische toestand bevinden indien zij op een bepaald niveau van beschrijving kunnen worden beschouwd als functioneel equivalente systemen, d.w.z. als instanties van een en dezelfde functionele organisatie. Nu is echter onmiddellijk duidelijk dat deze formulering veel te ruim is; op de keper beschouwd schrijft zij gewoon aan *teveel* dingen een geest toe — niet alleen aan mensen, maar ook aan voldoende complexe telefooncentrales, aan een reusachtig ingewikkelde mierenhoop, of zelfs aan de bevolking van China in een subtiel gecoördineerde gemeenschappelijke inspanning van haar één miljard leden. Het is immers niet ondenkbeeldig dat elk van deze systemen op een bepaalde niveau van beschrijving dezelfde functionele organisatie zou kunnen bezitten als de menselijke geest, terwijl wij het toch als bepaald tegenintuïtief ervaren in deze voorbeelden te spreken van mentale wezens.

Hoe kunnen dergelijke tegenvoorbeelden worden vermeden? Kennelijk zijn de eisen die worden gesteld aan het hebben van 'dezelfde functionele organisatie' niet streng genoeg. Wij zouden ze moeten aanscherpen met meer specifieke eisen, bijvoorbeeld door te eisen dat de systemen over bepaalde zintuigen moeten beschikken of over nader te bepalen gedragsmoge-

¹⁰ Zie bijvoorbeeld Fodor (1981), Pylyshyn (1984), en het overzicht van de geschiedenis van de cognitiewetenschappen in Gardner (1985). In de onderhavige bundel, onder meer Levelt en Sleutels & Geurts, *Knopen en connecties*.

lijkheden. Deze remedie lijkt echter al net zo erg te zijn als de kwaal. Ditmaal stelt het functionalisme zich weliswaar niet te liberaal op, maar valt het ten prooi aan een omgekeerd chauvinisme: tegelijk met de nepgeesten zoals de Superchinees zullen immers ook mogelijkwijs echte geesten worden geëlimineerd, zoals bepaalde diersoorten en Martianen en computers, omdat zij niet voldoende op mensen lijken, of verminkte, gehandicapte of anderszins abnormale mensen, omdat zij niet voldoende op normale mensen lijken.

Een mogelijk antwoord op deze objectie is dat het moeilijk zou zijn om die vermeende tegenvoorbeelden tot in detail uit te werken, en dat als dat *wel* gedaan zou worden, dan ofwel aan het licht zou komen dat de voorbeelden eigenlijk onmogelijk zijn, omdat zij lijden aan interne tegenstrijdigheden, ofwel dat wij niet meer met zo grote stelligheid zouden durven beweren dat er wel of niet sprake is van een echte geest in die voorbeelden. Ofwel de voorbeelden zelf zouden problematisch worden, ofwel onze spontane intuïties over wat geestelijk is en wat niet zouden ons in de steek laten. Zo is erop gewezen dat de tegenwerping zich beweegt op een vrij grofgemaasd niveau van analyse, en dat zij aanmerkelijk aan overtuigingskracht zou inboeten als wij zouden proberen haar te formuleren op het niveau van meer fijngeemaasde functionele eigenschappen.

Deze repliek laat zich verduidelijken aan de hand van een door Daniel Dennett gemaakt onderscheid tussen drie houdingen (*stances*) die men kan innemen om het gedrag van een functioneel systeem, bijvoorbeeld een mens of een machine, te voorspellen en te begrijpen, namelijk de intentionele houding (*intentional stance*), de ontwerphouding (*design stance*) en de fysische houding (*physical stance*).¹¹ De zetten van een schaakspelende computer, bijvoorbeeld, kunnen wij op drie manieren voorspellen en verklaren. Allereerst is het mogelijk (maar bijzonder omslachtig) om een zet te verklaren in termen van de elektronische schakelingen van de hardware van de computer. Deze fysische houding levert ongetwijfeld de meest exacte beschrijving van het gedrag van het systeem, tot op het exacte voltage in het kleinste onderdeel, maar zegt op het eerste gezicht weinig of niets over wat er nu precies zo interessant is aan het gedrag, bijvoorbeeld waarom de computer nu juist voor een korte rokade kiest. Vanuit de ontwerphouding is over dat laatste al meer te zeggen; zij concentreert zich op het doelmatig ontwerp van de diverse procedures in het computerprogramma, waaronder (vermoedelijk) een of meer procedures voor het bepalen van de opportuniteit van rokades. Bezien vanuit de intentionele houding, ten slotte, wordt het gedrag van de machine verklaard in termen van haar onderliggende 'bedoelingen', 'strategieën' en andere (quasi?-)intenties. Het computerontwerp en de hardware van de machine komen hierbij niet meer ter sprake en zijn, vanuit de intentionele houding bezien, volkomen irrelevant.

Welnu, in termen van dit onderscheid lijkt Blocks objectie gekluisterd te zijn aan het innemen van de intentionele houding ten opzichte van een systeem, en zou zij veel aan geloofwaardigheid verliezen wanneer wij het systeem zouden beschouwen vanuit het oogpunt van de fijnere details van zijn ontwerp, of zelfs vanuit dat van zijn fysische details. Van een systeem dat alleen op het meest globale niveau een functionele architectuur bezit zoals die van de menselijke geest, en waaraan wij alleen 'gemakshalve' intenties toeschrijven, kunnen wij ons inderdaad goed voorstellen dat het bij nader inzien niet *echt* een mentaal systeem is. Wanneer echter een systeem ook vanuit de meer fijngeemaasde ontwerphouding functioneel equivalent zou blijken te zijn met de menselijke cognitieve organisatie, of zelfs vanuit de fysische houding, worden onze intuïties over het mentale allens troebeler. Tot welke graad van precisie, zo zouden wij ons kunnen afvragen, moet de functionele structuur van een systeem overeenkomen met die van de menselijke geest, vooraleer wij er *echt* niet meer onderuit zouden kunnen om het desbetreffende systeem mentaal te noemen? Uiteraard kan deze repliek niet als een sluitend tegenargument worden opgevat, aangezien zij op de keper beschouwd niet meer is dan een *argumentum ad ignorantiam*.

¹¹ Dennett (1978). Een nadere uitwerking en verduidelijking van Dennetts ideeën over de intentionele houding geeft zijn bundel deels nieuwe artikelen *The Intentional Stance* (1987).

Daargelaten of Blocks eigen argumentatie wel zo sluitend als hij ons wil doen geloven, vast staat dat zijn redeneertrant school heeft gemaakt. In de afgelopen tien jaar heeft de wijsgerige wereld opmerkelijk veel energie gestoken in het verzinnen van steeds subtielere gedachtenexperimenten pro en contra het functionalisme. Ook het tweede tegenargument tegen het functionalisme dat wij hier bespreken is een loot van deze stam. Het is meer in het bijzonder gericht tegen de computationalistische uitwerking van het functionalisme en draait om de zogenaamde 'intentionaliteit' van mentale toestanden. Wij hebben zoëven gezien dat het computationalisme mentale processen beschouwd als algoritmische bewerkingen op interne symbolen. Deze mentale symbolen hebben een bepaalde betekenis; zij staan voor bepaalde objecten die zij op grond van hun symbool-karakter representeren. Althans, zo luidt de theorie. Deze theorie is aangevallen in een roemrucht artikel van John Searle uit 1980, getiteld 'Minds, brains, and programs'. Searle betoogt daar dat het functionalisme geen recht kan doen aan die voor mentale toestanden zo wezenlijke intentionaliteit; bijgevolg kan het functionalisme onmogelijk een juiste theorie van het mentale zijn.

De objectie loopt ongeveer als volgt. Stel dat wij mij opsluiten in een kamer met een volledige set Chinese karakters, keurig geordend en van nummers voorzien. Ik ken absoluut geen Chinees. Voor mij zijn die krabbels en kriebels allemaal eender; ik kan ze zo op het eerste oog alleen aan hun volgnummers uit elkaar houden. Nu krijg ik echter ook nog een zakjapanner en een handboek mee, met daarin alle regels die nodig zijn om de nummers van die karakters met elkaar in verband te brengen, in de trant van: *Als iemand je de serie Chinese karakters met volgnummers F_1, F_2, \dots, F_m aanbiedt, bereken dan als volgt welke serie Y_1, Y_2, \dots, Y_n je als 'antwoord' moet geven.* Deze regels, die in duidelijk Nederlands geschreven zijn, stellen mij in staat om de ene serie symbolen met de andere in verband te brengen. Stel nu dat ik mij hierin zo intensief oefen dat ik feilloos elke tekst in het Chinees kan 'beantwoorden' met de door mijn boekje en zakjapanner voorgeschreven juiste andere Chinese tekst. Voor de vloeiend Chinees sprekende Chinezen die zich nieuwsgierig buiten mijn kamer gereed houden, lijkt het nu alsof zij werkelijk in het Chinees met mij (of met de kamer) kunnen communiceren. En toch begrijp ik nog steeds niets van het Chinees, omdat ik *ex hypothesi* alleen maar berekeningen maak volgens de voorschriften in mijn boekje.

Dit beroemde gedachtenexperiment zou moeten aantonen dat intentionele cognitieve verschijnselen, in dit geval het spreken van een taal, weliswaar knap kunnen worden *gesimuleerd* met behulp van algoritmen, maar dat zij nooit volledig kunnen worden *verklaard* in termen van functionele eigenschappen of zuiver syntactische operaties, d.w.z. in termen van algoritmen voor het verwerken van formele symbolen. Bovenstaand voorbeeld laat zien dat formeel vlekkeloze symboolmanipulatie geen garantie biedt voor het feit dat het desbetreffende systeem zich ook bewust is van de *betekenis* van de symbolen. Eigenlijk kan er al nauwelijks van 'symbolen' worden gesproken. De Chinese symbolen hebben op zichzelf immers geen betekenis; alle betekenis die zij uiteindelijk blijken te bezitten is een afgeleide van de intentionaliteit van de vloeiend Chinees sprekende omstanders die de vragen stellen en de antwoorden interpreteren.¹²

Searles argumentatie heeft een stortvloed van vaak verrassend emotionele discussie losgemaakt.¹³ Sommige commentatoren vallen Searle bij omdat zij bang zijn dat het functionalisme het verschil tussen mensen en bijvoorbeeld thermostaten dreigt te reduceren tot een gradatiekwestie. Anderen beschuldigen hem juist van mystificatie van het mentale, van onwetenschappelijkheid, of van op hol geslagen filosofische fantasie. Weer anderen proberen het gedachten-

¹² Overigens kan een variant op dit argument reeds in 1970 worden aangetroffen bij Norman Kretzmann ('Medieval logicians on the meaning of the *propositio*', *Journal of Philosophy* 67, 1970, p. 787). Kretzmann slaagt erin een geheel eigen bijdrage tot het veld van ethnische stereotypen te leveren; hij voert niet een Chinees ten tonele, maar een "cooperative Turk who knows no English and German or chemistry."

¹³ Zie bijvoorbeeld het lange commentaar op zijn artikel in *Behavioral and Brain Sciences* 3, 1980, pp. 424-457.

experiment zodanig bij te sturen dat onze spontane intuïtie (namelijk dat er geen sprake is van echte kennis van het Chinees) op losse schroeven komt te staan. Zo is onder meer voorgesteld dat de kamer waarin ik mij bevind moet worden uitgerust met bepaalde randapparatuur (een TV-camera als oog, een motorisch systeem voor de output), en dat zij in haar geheel in het hoofd van een robot moet worden gemonteerd — *dan pas* zou het geheel op een echte Chinees lijken... Het merendeel van deze replieken was door Searle zelf al geanticipeerd in zijn oorspronkelijke artikel. Tot op de dag van vandaag houdt deze kwestie de gemoederen bezig.

Volgens Searles eigen visie op intentionaliteit is het mentale een doodgewoon biologisch verschijnsel, te vergelijken met melkvorming, fotosynthese, mitose en spijsvertering.¹⁴ Dit 'biologisch naturalisme', zoals hij het noemt, ziet het mentale als een product van de hersens, net zoals melk een product is van de melkklieren. Intentionaliteit is als het ware een soort van afscheiding of secreet van het brein; het is de taak van de neurobiologie om te onderzoeken hoe dit proces precies in zijn werk gaat. Searle stelt met grote nadruk dat hij zeker niet wil ontkennen dat onze geest een soort machine is; integendeel, hij zegt dat volgens hem *alleen* machines kunnen denken, en wel alleen een *heel bepaald soort* machine, "namely brains and machines that had *the same causal powers* as the brain."¹⁵ Jammer genoeg laat Searle in het midden wat wij hier moeten verstaan onder dergelijke 'causal powers'. Hij maakt met name niet duidelijk waarin zijn positie eigenlijk zou verschillen van het functionalisme. Als wij bedenken dat het functionalisme zich uitdrukkelijk richt tot de 'causale vermogens' van hersens en machines, namelijk tot hun vermogen om (computationeel) te bemiddelen tussen input en output, lijkt Searles biologisch naturalisme eerder zelf een vorm van functionalisme te zijn.

Laten wij dit hoofdstuk besluiten met een meer algemene opmerking over de betoogtrant van Block en Searle. Het regime van surrealistische gedachtenexperimenten heeft ons in een decennium meer filosofische fantasieën opgeleverd dan elk van de twee millennia daarvoor. Er zijn nu geesten van Zwitserse kaas, automaten van zielstof en dioden, van ectoplasma, ether en toiletpapier. Er zijn gehalveerde breinen, gevierendeelde breinen, ontlijfde breinen *in vitro*, kosmische culturen met geavanceerde cerebroscopie, denkende mierenkoloniën en hersenloze Marsmannetjes. Wij hebben computers met *Weltschmerz* en computers zonder *Weltschmerz*, wij hebben superchinezen en zakjapanners. Een rijke sheik richt de economie van Bolivia in naar de microstructuur van het menselijk brein; kan Bolivia nu denken? Hoe voelt het aan om een vlemuis te zijn, of de Verenigde Staten van Amerika?¹⁶

Mij persoonlijk bekruipt allengs sterker het vermoeden dat deze gedachtenexperimenten maar één ding duidelijk maken, namelijk dat voor elk vernietigend argument tegen het functionalisme een tegenargument kan worden bedacht dat intuïtief niet minder overtuigend is. Zo'n benadering moet onherroepelijk tot een impasse leiden; de balans van intuïties blijft immers steeds in evenwicht en er wordt geen milligram aan inzicht geproduceerd. Wetenschapshistorisch gezien lijkt zo'n situatie karakteristiek te zijn voor een discipline waarvan het nog onduidelijk is welke richting het onderzoek dient in te slaan. Men tast als het ware in het duister om zich heen en klampt zich vast aan iedere strohalm, zelfs aan troebele intuïties omtrent Martianen en *Weltschmerz*. Een vergelijkbaar beroep op intuïties treft men aan in andere periodes van de geschiedenis van de wetenschap waarin sprake is van ingrijpende conceptuele innovaties. Galileo's theorie van een draaiende aarde stuitte op enorme weerstand vanwege haar tegenintuïtieve consequenties, waaronder het spontane idee dat wij van de aarde zouden worden weggeslingerd, dat verticaal omhooggeworpen stenen mijlenver van ons vandaan zouden landen, en dat de wolken razendsnel langs de hemel zouden trekken. Als Copernicus gelijk had zouden wij intuïtief gesproken van de aarde moeten vallen naar het middelpunt van de kosmos toe. En wat te denken van de schijnbaar absurde consequenties van Einsteins relati-

¹⁴ Zie onder meer Searle (1983), pp. 262vv.

¹⁵ Searle (1980), p. 424; cursivering JS.

¹⁶ Een aardige dwarsdoorsnee van deze science fiction is verzameld in de populaire bindel *The mind's I* van Hofstadter & Dennett (1981).

teitstheorie? De tweelingenparadox, kromme lichtstralen en een gebogen ruimte, een kosmische maximumsnelheid. Toch zijn in al die gevallen de intuïties gezwicht voor de theorieën en niet omgekeerd. Wellicht moet de moraal van dit hoofdstuk dan ook luiden dat het niet zozeer de nieuwe functionalistische theorie van de geest is die nog verder verfijnd moet worden, maar veeleer onze spontane intuïties over wat het mentale eigenlijk is. Wellicht moet hier niet de hond met de staart kwispelen, om met Quine te spreken, maar veeleer de staart met de hond.¹⁷

5 Eliminatief materialisme. De geest een hersenschim?

Op een heel ander front wordt het functionalisme onder vuur genomen door het zogenaamd 'eliminatief materialisme', vooral bekend geworden door het werk van Paul Churchland.¹⁸ Net als de functionalist verwerpt ook de eliminatief materialist de soort-identiteitstheorieën van het oudere reductionistisch materialisme. Maar daarmee houdt dan ook alle overeenkomst tussen beide op. De reductie van psychologie tot het neurobiologie wordt ditmaal niet afgewezen omdat het mentale op de een of andere manier abstract is ten opzichte van zijn materieel substraat, maar omdat het überhaupt *niet bestaat*. Wil je iets reduceren, dan moet er natuurlijk eerst iets *zijn* om gereduceerd te worden. Welnu, volgens het eliminatief materialisme is er niets in de werkelijkheid dat correspondeert met onze traditionele psychologische begrippen en categorieën, en kunnen die dus ook niet worden herleid tot neurobiologische begrippen en categorieën. De geest is niet meer dan een spook, een fictie, een hersenschim die uiteindelijk door de neurowetenschap uit ons hoofd zal worden uitgebannen.¹⁹

De kritiek van het eliminatief materialisme richt zich meer in het bijzonder op onze zogenaamde 'volkpsychologie'. Daaronder wordt verstaan het geheel van begrippen en vuistregels die wij in het alledaagse leven gebruiken om onze eigen geest en die van anderen te beschrijven. De begrippen en (veelal impliciete) regels van deze volkpsychologie vormen te zamen een soort van vóórwetenschappelijke theorie over het mentale, een theorie in termen waarvan wij de oorzaken en gevolgen van ons eigen en andermans gedrag beschrijven, voorspellen en verklaren. Begrippen als 'oordeel', 'verlangen', 'herinnering' en 'mening', alsmede het daaraan gekoppelde idee van een afzonderlijk oordeels- en redeneervermogen, een wil, een geheugen, enzovoort, zijn evenzovele onderdelen van de volkpsychologie. Haar 'wetten' of regels omvatten vele honderden *common sense* generaliseringen ten aanzien van de werking en onderlinge samenhang van deze mentale faculteiten en begrippen, in de trant van 'Als X wil dat P, en meent dat (indien Q dan P), dan zal X proberen dat Q (als er verder niets tussenkomt)'.¹⁹

Een belangrijk en veelbesproken aspect van de volkpsychologie is dat zij mentale processen voorstelt als een soort linguïstische aangelegenheid, als het verwerken van inwendige 'zinnen' in een *language of thought*. Haar mentale toestanden zoals 'menen dat' en 'zich herinneren dat' worden geanalyseerd als *propositionele attitudes*, d.w.z. als een bepaalde verhouding of attitude van een persoon (het cognitief subject) tot een interne representatie (de inhoud van de mening of herinnering). Wanneer Bianca Castafiore meent dat *perestrojka* een schone zaak is, zeggen wij in feite dat een bepaald subject (Bianca C.) zich op een bepaalde manier (namelijk opiniërend) verhoudt tot een intern symbool met een bepaalde propositionele inhoud (namelijk dat *perestrojka* een schone zaak is). Mentale processen worden nu geanalyseerd als de verwerking van dergelijke interne representaties. Net als zinnen in andere (natuurlijke en formele) talen bezitten deze interne symbolen een complexe structuur en onderlinge samenhang, die van

¹⁷ Quine (1960), pp. 18-19.

¹⁸ Zie onder meer Churchland (1979), (1981), en (1988), pp. 43vv, alsmede het encyclopedische werk *Neurophilosophy* van Paul Churchlands echtgenote Patricia (1986).

¹⁹ De theorie van Churchland is weliswaar niet de enige vorm van eliminativisme ten aanzien van het mentale, maar wel de duidelijkste en de meest uitgewerkte versie ervan. Daarnaast heeft bijvoorbeeld ook het instrumentalisme van Daniel Dennett (1978 en 1987) onmiskenbaar eliminativistische trekjes, terwijl een 'syntactische' (als onderscheiden van 'neurobiologische') versie van het eliminatief materialisme wordt verdedigd door Stephen Stich (1983).

cruciaal belang is bij onze alledaagse verklaringen van gedrag. Wanneer wij bijvoorbeeld willen verklaren waarom Bianca een portret van Gorbatsjov op haar vleugel heeft staan, zullen wij haar bovengenoemde mening over *perestrojka* in verband proberen te brengen met andere meningen en wensen van haar waarin *perestrojka* een rol speelt, onder meer dat Gorbatsjov een van *perestrojka*'s voormannen is, dat je voormannen van schone zaken in de regel bewondert, dat het past om wie je bewondert op je vleugel te vereeuwigen, en zo voort. Deze meningen en wensen vertonen een welbepaalde samenhang op grond van de elementen van hun inhoud (Gorbatsjov, *perestrojka*, voorman, schone zaak). Deze elementen van de mentale representaties, zo wordt in de volkspychologie verondersteld, zijn expliciet aanwezig in de structuur van Bianca's kenapparaat en kunnen daarom ook causaal verantwoordelijk zijn voor Bianca's gedrag.

Dit voorwetenschappelijk beeld van het mentale wordt in ruwe lijnen overgenomen door de cognitieve psychologie. Ook zij postuleert het bestaan van interne faculteiten en analyseert mentale processen in termen van procedures voor de verwerking van expliciet gerepresenteerde symbolen in een soort van mentale taal. Het eerder gegeven voorbeeld uit de psycholinguïstiek is hiervan een voor de hand liggende illustratie. Maar ook op tal van andere terreinen worden mentale processen geanalyseerd naar het model van een computerprogramma, d.w.z. in termen van procedures voor de verwerking van symbolen. Cognitieve theorieën van kennisrepresentatie bijvoorbeeld hanteren als regel het model van een inwendige archiefruimte waarin keninhouden inderdaad als een soort zinnen liggen opgeslagen. Met behulp van bepaalde procedures kunnen deze zinnen naar hun diverse elementen met elkaar in verband worden gebracht, zodanig dat wanneer 'Gorbatsjov' wordt gevraagd ook 'perestrojka' wordt gereedgehouden. Gezien deze grote overeenkomsten tussen de theoretische structuur van de volkspychologie en die van de cognitieve psychologie worden zij doorgaans tot hetzelfde zogenaamd 'sententialistische paradigma' gerekend, met als centrale stelling dat mentale processen een vorm van *sentence crunching* zijn.²⁰

Churchland tekent op vier punten bezwaar aan tegen het sententialistische paradigma, d.w.z. tegen de gangbare cognitieve psychologie, het functionalisme en de volkspychologie. In de eerste plaats merkt hij op dat het merendeel van onze volkstheorieën onwaar is gebleken in het licht van wetenschappelijke ontwikkelingen, en dat zowel de theorieën als hun respectievelijke ontologieën uiteindelijk door deze laatste zijn geëlimineerd. Zo bijvoorbeeld de kristallen sferen van de oude volksskosmologie, de heksen en duivels, de ether en de phlogiston, en de naïeve, oude ideeën over het levende en het levenloze. In de tweede plaats beschuldigt Churchland de volkspychologie ervan in haar 2000-jarig bestaan geen verklaring te hebben gevonden voor zelfs de meest elementaire cognitieve vermogens zoals geheugen, leervermogen en intelligentie. Erger nog, de volkspychologie bedient zich van intentionele begrippen en verklaringen die een vlekkeloze inpassing binnen het succesvolle wereldbeeld van de natuurwetenschappen in de weg lijken te staan. En ten slotte bestempelt Churchland het functionalisme als een laakbare immuniseringsstrategie, met behulp waarvan *iedere* theorie zou kunnen worden gesauveerd, zelfs (zoals Churchland laat zien) de oude alchemistische vier-geesten theorie. Wat let ons immers om bijvoorbeeld de 'geest van kwik' te construeren als een bepaalde functionele eigenschap van materiële substraten, gedefinieerd als een zekere dispositie om licht te weerkaatsen, vloeibaar te worden bij verhitting, en zo voort voor elk van de andere eigenschappen waarvoor de geest van kwik verantwoordelijk werd geacht. Een theorie die zelfs de alchemie weet te sparen kan niet worden vertrouwd, aldus Churchland.

Tegenover deze negatieve kritiek op de volkspychologie stelt het eliminatief materialisme ook iets positiefs, en wel de belofte dat het ook anders en beter kan. In de volkspychologie, aldus Churchland, ligt ons spontane mens- en zelfbeeld bevat, het theoretisch raamwerk in termen waarvan wij ons zelfbewustzijn en onze cognitieve interactie met de buitenwereld conceptualiseren. Aangezien dit raamwerk slechts een van de vele mogelijke theorieën is, en nog een slechte, onware theorie bovendien), moet het ook kunnen worden vervangen door andere,

²⁰ Zie bijvoorbeeld Patricia Churchland (1986), pp. 386vv.

beteren theorieën. De oude volkpsychologie hebben wij gedurende onze jeugd verworven toen wij ons een bepaalde manier van spreken over wat er zich in ons eigen hoofd en in dat van anderen afspeelt aanleerden; evenzo moet het mogelijk zijn om een nieuwe, meer wetenschappelijke 'volkpsychologie' te leren, eenvoudigweg door van jongs af te worden ondergedompeld in een wetenschappelijke (neurofysiologische) beschrijving van het inwendig gebeuren. Het vooruitzicht dat Churchland ons schetst is fascinerend.²¹ Wij zouden ons kunnen aanleren niet langer een beperkt aantal kleuren te zien, maar onze kleurwaarnemingen rechtstreeks in termen van lichtfrequenties te conceptualiseren; wij zouden ons bewustzijn niet langer hoeven te conceptualiseren in termen van inwendige zinnen in een mentale taal, maar direct kunnen denken in trillingen, banen en vormveranderingen van meerdimensionale lichamen in neuronale vectorruimten, waarvan de oude propositionele attitudes maar zwakke ééndimensionale afspiegelingen zouden zijn. Ons bewustzijn zou werkelijk en letterlijk kunnen worden verruimd. Het belang van deze conceptuele omwenteling zou vele malen groter zijn dan dat van de Copernicaanse revolutie.

Het eliminatief materialisme is niet alleen de meest recente, maar ontegenzeggelijk ook de meest dierste vorm van kritiek die het functionalisme heeft uitgelokt. Hoe ernstig de situatie eigenlijk is begint nu langzaam door te dringen. Wij zullen hier in het kort enkele mogelijke antwoorden op de bovenstaande argumenten bespreken.²²

In de eerste plaats mag worden betwijfeld of inderdaad het merendeel van onze 'volkstheorieën' is geëlimineerd (in plaats van gereduceerd) door wetenschappelijke theorieën. Het tegendeel lijkt juist het geval te zijn. Volkstheorieën lijken veeleer de meeste wetenschappelijke theorieën *overleefd* te hebben; die laatste komen en gaan, maar de 'volks'-ideeën blijven. Dat hoeft ook geen wonder te heten. Volkstheorieën lijken veel meer 'volks' te zijn dan 'theorie'. Zij bieden niet zozeer een *theorie* in de volle zin des woords, als wel een eerste, betrekkelijk observationele inventarisering van een bepaald domein van verschijnselen. Zij verklaren die verschijnselen niet zozeer, maar beschrijven veeleer wat er überhaupt te verklaren valt. Vanuit dit gezichtspunt beschouwd zouden wij moeten zeggen dat het eliminatief materialisme zijn doel voorbij schiet. Aan de ene kant spreekt het immers de verwachting uit dat al onze cognitieve vermogens te zijner tijd zullen worden verklaard door de neurowetenschappen, maar tegelijkertijd verwerpt het ons standaard waarnemingsvocabulaire (de volkpsychologie) in termen waarvan deze vermogens worden gespecificeerd. Hoe zouden wij dan ooit kunnen bepalen wat er eigenlijk verklaard moet worden? Vanuit deze optiek lijkt het bovendien op zijn zachtst gezegd incorrect om de volkpsychologie aan te wrijven dat zij geen verklaringen heeft opgeleverd; wanneer zij ons al meer dan twintig eeuwen voorhoudt wat de *explananda* van het mentale zijn, en als er nog steeds geen *explanans* is, dan is dat toch zeker de fout van de desbetreffende wetenschap, d.w.z. van de cognitieve psychologie, of zelfs van de door Churchland zo hooggeschatte neurobiologie!

Wat het alchemistische argument tegen het functionalisme betreft, zo moge duidelijk zijn dat dit staat of valt met de diagnose die over de volkpsychologie wordt gesteld. Wanneer het onjuist is om een bepaalde onderzoeksstrategie zoals het functionalisme toe te passen op een foute theorie zoals de alchemie, dan wil dat nog niet zeggen dat het eveneens onjuist is die strategie toe te passen op een *goede* theorie. Per slot van rekening zal iedere procedure die de verkeerde input krijgt ingevoerd met de verkeerde output voor de dag komen, maar dat kan nauwelijks de procedure worden aangewreven.

Als er echter *iets* is waar het eliminatief materialisme *echt* een hekel aan heeft, dan is het wel intentionaliteit. Zoals wij eerder zagen verwijt Searle het functionalisme dat het zogezegd niet intentioneel genoeg is; paradoxaal genoeg vreest de eliminativist het omgekeerde: dat het veel te intentioneel is! Deze intentionaliteit van het mentale zou op gespannen voet staan met het

²¹ Zie onder meer Churchland (1979), pp. 116vv, (1981), pp. 84vv.

²² Voor een meer gedetailleerde bespreking van het eliminatief materialisme, zie onder meer Stich (1983), Goldman (1986), Patricia Churchland (1986), alsmede Sleutels (1988).

zo succesvolle materialistische wereld- en mensbeeld van de andere wetenschappen. Volgens Churchland is de intentionaliteit van het mentale net zo'n vreemde eend in de bijt als eertijds het 'vitaal principe' in de negentiende-eeuwse biologie; net zoals dat laatste is geëlimineerd door de organische scheikunde, zou ook het eerste moeten worden geëlimineerd door de neuro-wetenschap.

Wij zouden echter met enig recht kunnen betwijfelen of deze vergelijking wel helemaal zuiver op de graat is. Moet de volkpsychologie niet eigenlijk worden vergeleken met zoiets als de *volksbiologie*, in plaats van met (een bepaalde school van denken in) de *wetenschappelijke biologie*? Weliswaar werd door bepaalde biologen in de negentiende eeuw inderdaad een metafysisch principe van vitaliteit gepostuleerd ter verklaring van levensverschijnselen, maar dat is een buitenissigheid waarvan de volksbiologie verschoond blijft. Al spreken wij in het dagelijks leven over 'het leven dat in de amaryllis zit' of dat 'uit opa is geweken', daarmee vergrijpen wij ons nog niet aan een dualistische metafysica. In feite doen wij niet meer dan een oppervlakkig onderscheid aanbrengen tussen twee soorten van verschijnselen, namelijk levensverschijnselen en andere verschijnselen. De onderliggende verklaring van deze verschillen wordt gedelegeerd aan filosofen en wetenschappers. Analoog hieraan, zo zouden wij kunnen redeneren, wordt in de volkpsychologie een onderscheid gemaakt tussen cognitieve en andere verschijnselen. Wat er precies in ons hoofd zit dat als verklaring voor het verschil tussen deze verschijnselen zou kunnen worden geciteerd, wordt door de volkpsychologie als zodanig open gelaten. De leek, de praktizerend volkpsycholoog, delegeert het antwoord op deze kwestie aan filosofen en wetenschappers, in wier vaardige handen de onderliggende oorzaak inderdaad de meest zonderlinge metafysische vormen aanneemt, uiteenlopend van een onstoffelijke ziel tot een complex zenuwstelsel.

Het spreekt voor zich dat hiermee het laatste woord over het eliminatief materialisme nog niet is gesproken. Elk van de hier aangerode kwesties is het onderwerp van verhitte controverse. Veel meer onderzoek zal nodig zijn, zowel binnen de cognitiewetenschappen als binnen de filosofie van de geest, om uit te maken of de sombere verwachtingen die het eliminatief materialisme ten aanzien van de traditionele psychologie koestert gerechtvaardigd zijn of niet. Laat ik hier echter althans één lijn van onderzoek noemen die heden ten dage grote belangstelling geniet van de zijde van zowel filosofen als cognitiewetenschappers, en die bijzonder nauw lijkt aan te sluiten bij het eliminatief materialisme — inderdaad: het connectionisme. Op het eerste gezicht lijkt deze nieuwe stroming een ontkenning te zijn van vrijwel alles wat de 'klassieke' cognitiewetenschap dierbaar is. Connectionistische modellen bedienen zich niet van expliciete, gearticuleerde regels en procedures voor het verwerken van expliciete, gearticuleerde symbolen. Het geheugen, bijvoorbeeld, wordt niet langer voorgesteld als een soort geordende opslagplaats van zinnen, met een centrale verwerkingseenheid die met een boekje in de hand de zinnen kan terugvinden, ontleden, associëren en verder verwerken. Het geheugen wordt nu eerder voorgesteld als een hologram, waarbij elke bit informatie verspreid ligt over een groot aantal punten; de klassieke afzonderlijke 'geheugeninhouden' zijn nu verdeeld over een groot netwerk van onderling doorverbonden verwerkingseenheden, die ieder betrokken zijn bij de representatie van meerdere van die 'geheugeninhouden'.²³

Als dit connectionistische beeld van het mentale correct is, lijkt de volkpsychologie ons inderdaad een bedrieglijk beeld van de geest voor te spiegelen. Het *lijkt* weliswaar alsof onze geest zich door welbepaalde regels laat leiden en alsof er binnen in ons hoofd mentale zinnen worden verwerkt, maar in werkelijkheid treffen wij daar alleen complexe stromen van activiteit

²³ Voor diverse filosofische aspecten van het connectionisme, in het bijzonder de aard en status van representaties in het connectionisme en de verhouding tussen connectionisme en klassiek computationeel cognitivisme, zie de bijdrage van Sleutels & Geurts elders in deze bundel. Voor meer informatie, zie voorts Rumelhart & McClelland (1986), Churchland (1986), pp. 458vv, Churchland (1988), Pinker & Prince (1988) en Smolensky (1988), alsmede de artikelen van Phaf & Murre en Levelt elders in deze bundel.

in uitgestrekte netwerken van verwerkingseenheden aan. De waargenomen regelmaat is een zogenaamd 'emergente eigenschap' (*emergent property*) van de onderliggende statistische processen in de netwerken van verwerkingseenheden; elk van deze eenheden werkt volgens statistische principes die niet zinvol kunnen worden weergegeven in termen van een cognitieve regel, maar desniettemin vertoont het netwerk als geheel gedrag dat globaal bezien geleid schijnt te worden door cognitieve regels. De afzonderlijke verwerkingseenheden kan men zich overigens wellicht het beste voorstellen als een soort geïdealiseerde zenuwcellen. Een aanzienlijk deel van de aantrekkingskracht van het connectionisme lijkt trouwens te kunnen worden toegeschreven aan het betrekkelijk gemak waarmee connectionistische modellen kunnen worden geïnterpreteerd als een model van de menselijke hersens.

Hoe dicht het connectionisme en het eliminatief materialisme op bepaalde punten ook bij elkaar mogen liggen (bijvoorbeeld ten aanzien van het belang dat beide hechten aan neurobiologische modellen de evenredige geringschatting van volkpsychologie), er zijn nog andere interpretaties of ontwikkelingen van het connectionisme denkbaar. Zo onderscheiden Pinker en Prince drie mogelijke relaties tussen connectionisme en traditionele ('symbol processing') psychologie: eliminatie, implementatie en revisie. Elders in deze bundel zal nader op deze onderscheiden mogelijkheden worden ingegaan.²⁴

Een daarvan is het zojuist besproken eliminatief connectionisme, waarin connectionistische modellen de traditionele cognitieve modellen verdringen. Daarnaast zou men echter kunnen verdedigen dat connectionistische modellen in feite ergens moeten worden ingeschaald *tussen* de neurofysiologie en de cognitieve psychologie, zodanig dat het connectionisme beschrijft hoe de functies en procedures van de psychologie nader zijn geïmplementeerd in netwerken van elementaire verwerkingseenheden, terwijl de neurofysiologie op haar beurt beschrijft hoe deze netwerken zijn geïmplementeerd in de concrete *wetware* van het zenuwstelsel. De cognitieve psychologie beschrijft als het ware *wat* er wordt berekend; het connectionistisch model beschrijft *hoe* deze berekeningen precies worden gemaakt; en de neurobiologie beschrijft hoe het rekenapparaat zelf gebouwd is. Deze werkverdeling, die wij *implementatieel connectionisme* zouden kunnen noemen, kan als een directe nazaat van het boven besproken functionalisme worden beschouwd. Zij zou ruimte laten voor elk van de drie benaderingen van het menselijk kenapparaat. Of deze functionalistische coëxistentie ook een even grote (methodologische en metafysische) onafhankelijkheid van elk der drie niveaus met zich mee zou brengen als oorspronkelijk door het functionalisme werd geclaimd voor neurofysiologie en psychologie, staat echter nog te bezien. Het lijkt er eerder op dat met de komst van een intermediair verklaringen- en beschrijvingsniveau, dat van het connectionisme, de conceptuele afstand tussen biologische *wetware* en cognitieve *software* aanmerkelijk wordt geslecht. Een implementatiele interpretatie van het connectionisme zou dan ook wellicht goede diensten kunnen bewijzen als conceptuele brug waarlangs toenadering en samenwerking tussen neurowetenschap en cognitieve psychologie mogelijk is. Deze speculatie (als zij louter speculatie is²⁵) zou kunnen worden gestaafd door de verdere ontwikkeling van connectionistische modellen — wederom een voorbeeld van een filosofisch belangwekkende belofte die vooralsnog in de boezem der empirie besloten ligt.

Een derde en laatste mogelijke interpretatie van het connectionisme die door Pinker en Prince wordt genoemd is het 'revisionistisch' of 'gemengd' connectionisme, een begrip waarachter eigenlijk een breed scala van mogelijke middenwegen tussen de vorige twee uitersten schuilgaat. De cognitiewetenschap zou kunnen vasthouden aan het idee dat mentale processen inderdaad symboolverwerking met behulp van procedures zijn, met dien verstande dat deze

²⁴ Pinker & Prince (1988), pp. 75-78. Zie de bijdrage van Sleutels & Geurts elders in deze bundel.

²⁵ Een soortgelijk theoretisch model met drie onderscheiden niveaus van beschrijving en verklaring, voorgesteld door wijlen David Marr, kan reeds bogen op aanzienlijk empirisch succes, juist omdat het ruimte schept voor een intensieve uitwisseling van begrippen en *constraints* tussen deze niveaus. Zie Marr (1982), of voor een beknopt samenvattend overzicht bijvoorbeeld Kitcher (1988).

procedures en symbolen wellicht heel anders van aard zijn dan onze spontane volkspychologie ons zou willen doen geloven. De regels die ons inwendig taalapparaat volgt, bijvoorbeeld, zouden helemaal niet hoeven te lijken op de vertrouwde regels van de grammatica, en de symbolen die door het apparaat worden verwerkt zouden ons zeker niet meteen aan letters, woorden of zinnen hoeven te doen denken. In plaats daarvan zouden de symbolen abstracties kunnen blijken te zijn over een schijnbare wirwar van verwerkingseenheden, en zouden de procedures in feite kunnen blijken te bestaan in de niet minder verwarrende wijze waarop deze verwerkingseenheden met elkaar tot netwerken zijn doorverbonden. Deze interpretatie van het connectionisme, die overigens de vorige bepaald niet hoeft uit te sluiten, ligt zeer dicht bij de boven gegeven schets van de volkspychologie als een betrekkelijk oppervlakkige en beschrijvende inventarisatie van wat er überhaupt verklaard moet worden ten aanzien van het mentale, die zelf in het midden laat hoe de onderliggende verklaring er precies zal uitzien.

6 Modulariteit. Een verdeelde geest?

Uit bovenstaand overzicht van enkele van de voornaamste controverses omtrent aard en werkelijkheid van mentale processen zal bovenal één ding duidelijk zijn geworden: dat de geesten vooralsnog verdeeld zijn. Het wordt echter nog krasser: er is zelfs een theorie volgens welke *de geest zelf* verdeeld is, de zogenaamde 'modulaire' opvatting van het mentale. Een eleganter besluit van dit overzicht is nauwelijks voorstelbaar. De discussie over de modulariteit van het kenapparaat staat als het ware dwars op de boven besproken thema's in de filosofie van de cognitiewetenschappen (straks zal ik suggereren dat deze orthogonale ligging zelfs vrij letterlijk mag worden genomen). Inzet van de discussie is de vraag of het kenapparaat moet worden opgevat als een naadloos monolithisch geheel, dan wel als een min of meer losse formatie van zelfstandige kleinere eenheden, elk specialist op zijn eigen gebied. Ofschoon de eerste *unitaristische* theorie de oudste rechten heeft wint de concurrerende *modulaire* opvatting de laatste tijd steeds meer aan terrein. Zij steekt op tal van plaatsen en in tal van vormommingen de kop op. Zo ligt zij in de informatica ten grondslag aan het idee van gestructureerd, modulaair programmeren. In de psychologie duikt zij op als de theorie dat de cognitieve architectuur zich bedient van een aantal afzonderlijke subsystemen met specifieke rekencapaciteiten (onder meer in David Marrs theorie van gezichtswaarneming). In de psycholinguïstiek ligt zij ten grondslag aan Chomsky's idee van een gespecialiseerd 'taalorgaan' en van de 'onderdelen van de grammatica'. De meeste bekendheid heeft de modulariteitstheorie echter gekregen in de filosofie van de cognitiewetenschappen, waar zij nader is uitgewerkt in het werk van onder anderen Zenon Pylyshyn en Jerry Fodor.²⁶

Vooral Fodors idee van 'informatieel ingekapselde modules' heeft wortel geschoten. Hij presenteert zijn modulariteitsthese tegen de achtergrond van de geschiedenis van de faculteitenpsychologie, d.w.z. de theorie dat de geest bestaat uit een aantal onderscheiden cognitieve vermogens of faculteiten, onder meer voor geheugen, wil, waarneming en zo meer. Er zijn ruwweg twee soorten van faculteitenpsychologie, een horizontale en een verticale variant. Volgens de horizontale variant omvat de geest een aantal vermogens die als het ware bestemd zijn voor gemeenschappelijk gebruik door de diverse mentale processen; de specifieke aard van elk proces wordt dan bepaald door de eigen melange van faculteiten waarvan het gebruik maakt. Het kernpunt van de theorie is dat het steeds dezelfde faculteit zou zijn die door de uiteenlopende processen op de diverse gebieden van cognitie wordt gebruikt. Er zou dus maar één geheugen zijn, één waarnemingsvermogen en één oordeelsvermogen, waarvan zowel bij het drinken van een Chateau de Lamarque 1976 als bij het vervangen van een kapotte gloeilamp gebruik moet worden gemaakt. Deze 'horizontale' visie lijkt betrekkelijk dicht bij ons volkspychologisch beeld van mentale processen te blijven, die wij ons immers ook spontaan en intuïtief voorstel-

²⁶ Zie respectievelijk Marr (1982), met name pp. 8-38; Chomsky (1980); Pylyshyn (1984), met name pp. 13vv; Fodor (1983).

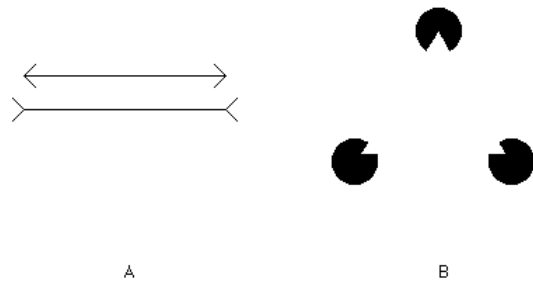
len als een soort collectieve inspanning van een bundel van steeds dezelfde vermogens. Dat deze intuïtieve plausibiliteit geen garantie is voor de waarheid van de horizontale faculteittenpsychologie, is een van de lessen die het eliminatief materialisme ons heeft ingepeperd.

De verticale variant van de faculteittenpsychologie (waarvan de oorsprong door Fodor bij de frenologie of knobbelkunde van Gall wordt gezocht) ontkent het bestaan van dergelijke 'horizontale' vermogens. Zij ontkent dat er maar één geheugen en één waarnemingsvermogen zou zijn. In plaats daarvan voorziet zij in een groot aantal afzonderlijke faculteitten, ruwweg één voor ieder domein of iedere soort van cognitieve processen. Deze afzonderlijke faculteitten staan 'verticaal' naast elkaar, als het ware als evenzovele zelfstandige kolommen op een rij; zij hoeven geen rekenfaciliteiten te delen met andere vermogens of processen. Het muzikaal vermogen, bijvoorbeeld, zal zijn eigen specifieke muzikaal geheugen hebben, dat geheel onafhankelijk is van de geheugenvoorzieningen van andere faculteitten, bijvoorbeeld het rekenvermogen.

Tegen deze achtergrond kan Fodors eigen theorie worden beschreven als een gemengd horizontale en verticale psychologie. Fodor maakt een onderscheid tussen twee grote groepen van mentale processen, namelijk de 'centrale' processen en de processen in de diverse 'perifere' systemen. De centrale processen bedienen zich van een aantal gemeenschappelijke, horizontale vermogens; de perifere systemen fungeren daarentegen als evenzovele verticale faculteitten of, zoals Fodor ze noemt, 'modules', elk met zijn eigen onafhankelijke rekenfaciliteiten. Tot deze modules worden onder meer bepaalde delen van de sensorische inputsystemen gerekend (met name die voor visuele waarneming en voor spraakherkenning), alsmede bepaalde delen van de outputsystemen (met name die voor motoriek en taalproductie).

Fodor bespreekt een aantal eigenschappen die hij kenmerkend acht voor zijn hypothetische modulaire bouwstenen. De vier belangrijkste kenmerken zullen wij hier in het kort bespreken.²⁷ In de eerste plaats is een module 'informatieel ingekapseld', waarmee wordt bedoeld dat zij enkel toegang heeft tot informatie die direct binnen de module zelf beschikbaar is, en niet tot enigerlei 'achtergrondinformatie' die van elders moet komen. Tot de bekendste aanwijzingen voor het bestaan van zo'n informatiele inkapseling behoren diverse optische zinsbegoochelingen, onder meer de illusie van Müller-Lyer (figuur 2a) en de illusie van Kalisza, ook wel schertsend genoemd de 'Omo'- of 'Witter dan wit'-illusie (figuur 2b). Ook al hebben wij nagemeten dat de twee lijnen in figuur 2a even lang zijn, desondanks zien wij de een als langer dan de ander; de door meting verkregen achtergrondinformatie is kennelijk onbereikbaar voor de (veronderstelde) module die de visuele beelden verwerkt. Hetzelfde geldt, *mutatis mutandis*, voor figuur 2b. Zelfs nadat wij hebben vastgesteld dat de afbeelding objectief gezien slechts uit twee kleuren bestaat, zwart en wit, blijven zich de 'subjectieve contouren' van een witter-dan-witte driehoek aftekenen tegen de witte 'achtergrond'.

In de tweede plaats werkt een module volkomen onwillekeurig, d.w.z. zij treedt automatisch in werking zodra een bepaalde input wordt aangeboden. Wij kunnen de module ook niet stoppen of haar ervan weerhouden in werking te treden. Hier kan wederom worden gedacht aan de zoëven genoemde voorbeelden, maar ook aan het bekende verschijnsel dat wij een uitspraak in onze moedertaal onmiddellijk en onwillekeurig horen als een zin, en niet als een losse verzameling van betekenisloze klanken.



Afbeelding 2
Twee optische illusies

²⁷ Voor een meer gedetailleerde bespreking zij verwezen naar de desbetreffende andere bijdragen in deze bundel.

In de derde plaats zouden modulaire systemen gekenmerkt worden door specifieke domeinen van toepassing; elke module verwerkt alleen stimuli van een bepaalde soort die kenmerkend is voor die module. Ook voor deze eigenschap bestaan duidelijke experimentele aanwijzingen, waarbij wij onder meer kunnen denken aan bepaalde delen van het visuele systeem waarvan is aangetoond dat zij enkel reageren op retinale beelden van lijnen die in een bepaalde richting bewegen, of aan bepaalde computationele systemen voor spraakanalyse die alleen reageren op akoestische signalen die (al dan niet terecht) voor taaluitingen worden gehouden.

Een vierde en laatste belangrijke eigenschap van modulaire systemen zou zijn dat zij als het ware volledig zijn gemechaniseerd, d.w.z. dat zij bestaan uit specifieke, locale neurale structuren. Verwacht mag worden dat modules op grond van dit kenmerk een overeenkomstig specifiek patroon van mankementen zullen vertonen, hetgeen ook inderdaad bevestigd lijkt te worden door neuropsychologisch onderzoek.²⁸ Locale hersenletsels (waarbij de kans groot is dat afzonderlijke modules worden beschadigd) blijken geassocieerd te zijn met geheel of gedeeltelijk verlies van specifieke cognitieve functies, vooral in het geval van de zintuigsystemen en taalmechanismen (agnosie, afasie, enzovoort). De door de theorie gepostuleerde centrale processen lijken daarentegen niet gelocaliseerd te zijn; zij zijn veeleer 'equipotentieel' verdeeld over uiteenlopende hersenstructuren. Grofgezegd lijkt er niet een bepaald deel van de hersens gereserveerd te zijn voor de afleidingsregel van *modus ponens*.

Deze vier basiskenmerken van modules brengen nog tal van andere eigenschappen met zich mee. Een daarvan is bijvoorbeeld dat modules bijzonder snel zouden kunnen werken. Omdat de module onwillekeurig in werking treedt gaat er geen tijd verloren aan beraadslagingen en aarzelingen vooraf. Omdat zij alleen op specifieke domeinen opereren kunnen modules bovendien inspelen op bepaalde contingente eigenschappen van dat domein, zonder zich hoeven te bekommeren om de vraag of die eigenschappen ook gelden in andere domeinen. Nog meer tijd wordt gewonnen door de informatiele inkapseling van de modules, omdat dit betekent dat er minder informatie hoeft (en kan) worden geraadpleegd en dat alle benodigde informatie onmiddellijk binnen handbereik is. En ten slotte zijn de modules *hardwired*, zodat de verwerkingssnelheid aanmerkelijk wordt opgevoerd.

Al deze kenmerken gelden volgens Fodor niet voor de zogenaamde centrale systemen en processen. Aangezien in de centrale processen op de een of andere manier de informatie van de verspreide perifere modules bij elkaar moet worden gebracht, kunnen zij onmogelijk domeinspecifiek zijn. Evenmin kunnen zij informatieel ingekapseld zijn, want zij brengen de uiteenlopende brokken informatie van de diverse modules op de een of andere manier met elkaar in verband. In principe kan iedere bit informatie die vanuit de ene module binnenkomt beïnvloeden hoe de informatie van de andere bronnen moet worden verwerkt. Dit holistisch karakter van centrale processen maakt het erg moeilijk om ze precies te analyseren; alles hangt er met alles samen, zodat wij nog niet eens kunnen bepalen waar onze analyse zou moeten beginnen. In feite worden wij geconfronteerd met wat in de AI het 'raamprobleem' heet (*frame problem*): hoe kunnen wij paal en perk stellen (er als het ware een raamwerk omheen plaatsen) aan de kennis die gereviseerd moet worden in het licht van nieuwe informatie? Fodor spreekt hier van het 'Quineaans' of 'isotropisch' karakter van centrale processen. Elk onderdeel van onze kennis is van mogelijk belang voor de beoordeling en verwerking van elk ander onderdeel. Maar als er niet een soort van Archimedisches punt is waarop ons onderzoek kan aangrijpen, hoe kunnen wij centrale processen dan ooit begrijpen?

De modulariteitsthese mag zich verheugen in een bijzonder levendige belangstelling, zowel van filosofische als van empirisch-wetenschappelijke zijde (een feit waarvan de onderhavige bundel getuigenis aflegt). Nu wij in het voorgaande al zó vaak hebben kunnen noteren hoe nauw de wetenschappelijke en filosofische aspecten van de cognitiewetenschap met elkaar samenhangen, zal deze 'vermenging van motieven' (zo men daar überhaupt nog van wil spreken) niemand meer verbazen. In feite schuilt achter de modulariteitstheorie een compleet research-

²⁸ Zie bijvoorbeeld de schat aan klinisch materiaal verzameld in Luria (1973).

programma, een onderzoeksvorstel dat niet alleen op zijn empirische merites beoordeeld kan worden, maar ook op zijn methodologische en filosofische implicaties. De theorie heeft niet alleen empirisch toetsbare consequenties, maar houdt ons evenzeer een welbepaalde metafysica van het mentale voor en brengt een eigen methode van onderzoek met zich mee. Bijgevolg kan worden verwacht dat er drie lijnen van kritiek mogelijk zijn op de modulariteitsthese — empirisch, filosofisch, en methodologisch. Wij zullen hier van elk type kritiek een voorbeeld laten zien.²⁹

(I.) Een van de terreinen waarop belangrijk *empirisch* onderzoek wordt gedaan naar de modulaire structuur van ons cognitief systeem, in dit geval van ons taalapparaat, betreft de herkenning van woorden in een tekst of taaluiting (*lexical processing*). Een van de twistpunten op dit gebied is de vraag of en in hoeverre de verwerking van afzonderlijke woorden wordt beïnvloed door de (linguïstische en andere) context waarin deze woorden voorkomen. Twee typen van theorieën staan lijnrecht tegenover elkaar, de zogenaamd 'modulaire' en de 'interactieve' theorieën. Volgens de modulaire theorieën is lexicale verwerking ondergebracht in een autonoom modulaair subsysteem van het taalapparaat. Deze visie weet zich gesteund door het feit dat woordherkenning in normaal taalgebruik automatisch, onwillekeurig en zeer snel gebeurt, en (zoals experimenteel kan worden aangetoond) kennelijk zonder daarbij te interfereren met andere cognitieve processen. Als woordherkenning zich inderdaad afspeelt in een afzonderlijke module, informatieel afgeschermd van andere delen van ons cognitief systeem, kan tevens worden voorspeld dat de werking van deze lexicale module niet wordt beïnvloed door de context. Om precies te zijn kunnen wij voorspellen (i) dat dezelfde stimuli in verschillende context zullen worden herkend als dezelfde woorden, en (ii) dat de snelheid waarmee woordherkenning plaatsvindt niet wordt beïnvloed door de context.

Deze voorspellingen worden rigoreus ontkend door de interactieve theorieën. Volgens deze laatste wordt de verwerking van zintuiglijke informatie van meet af aan beïnvloed door contextuele informatie. Zintuiglijke processen zouden gedurende heel het proces van woordherkenning worden gestuurd door contextuele informatie 'van bovenaf', zodanig dat de kandidaatwoorden waaruit het zintuiglijk proces een keus moet maken tevens voortdurend blootstaan aan een selectie op grond van hun geschiktheid of waarschijnlijkheid binnen de aanwezige context. Dezelfde woorden zullen zodoende sneller herkend worden in rijke contexten dan in verarmde contexten. Op grond van deze context-gestuurde selectie van kandidaatwoorden kan bovendien worden verwacht dat dezelfde stimuli in verschillende contexten soms zullen worden geïnterpreteerd als verschillende woorden.

Deze lijnrecht tegenover elkaar staande voorspellingen raken aan de kern van de modulariteitsthese. Tot dusver zijn er nog geen doorslaggevende empirische resultaten pro of contra voor handen; het lot van de modulariteit van *lexical processing* ligt in de hand van nader experimenteel onderzoek.³⁰

(II.) Een *filosofische* tegenwerping tegen de modulariteitstheorie is de suggestie dat alle cognitieve processen, inclusief de vermeend 'modulaire' processen, in werkelijkheid door en door 'Quineaans' en 'isotropisch' zijn, d.w.z. dat in feite alle processen centrale processen zijn. Bijgevolg zouden modules niet bestaan. Volgens Fodors 'First Law of the Nonexistence of Cognitive Science' zou dit het einde van de cognitiewetenschap betekenen.³¹ De werking van modules kunnen wij nog begrijpen, omdat modules een aantal min of meer duidelijke beperkingen kennen (localisering, inkapseling, automatisering). Maar als er geen modules bestaan is de

²⁹ Voor verdere informatie, zie onder meer de recente bundels van Gopnik en Gopnik (1986), en Garfield (1987).

³⁰ Voor meer informatie over dit onderwerp zij verwezen naar de bijdrage van Seidenberg en Tanenhaus in Gopnik & Gopnik (1986), pp. 135vv. Een middenweg tussen extreem-modulaire en extreem-interactieve modellen van woordherkenning wordt voorgesteld door onder anderen Tanenhaus, Denn & Carlson, in Garfield (1987), pp. 83vv.

³¹ Fodor (1983), p. 107.

kans groot dat de werking van de geest voor eeuwig een raadsel zal blijven, gezien het bovengenoemde holistisch karakter van centrale processen.

Het bestaan van modules wordt ontkend door onder anderen de psycholoog John Anderson, die daaraan echter niet Fodors pessimistische conclusie ten aanzien van de mogelijkheid van een cognitiewetenschap wenst te verbinden. Als uitgesproken voorstander van een 'unitaristische' cognitieve architectuur probeert hij te beargumenteren dat de vermogens van de geest eerder horizontaal georganiseerd zijn dan verticaal. Anderson wijst onder meer op de betrekkelijk korte ontwikkelingsgeschiedenis die de meeste cognitieve functies achter de rug hebben, een evolutie die in feite veel te kort lijkt te zijn om gespecialiseerde mentale 'organen' à la Fodor te kunnen hebben voortgebracht. Het is volgens hem waarschijnlijker dat de diverse mentale processen bijzondere toepassingen zijn van een en dezelfde, evolutionair ontwikkelde, universele cognitieve architectuur. Als er al gespecialiseerde mentale organen zouden zijn, zo betoogt hij verder, dan zouden zij in de diverse cognitieve processen dermate verweven zijn dat het bijzonder moeilijk, zo niet onmogelijk, zou zijn om hun individuele bijdragen in kaart te brengen.

Bij Andersons theorie moeten evenwel twee kanttekeningen worden gemaakt. In de eerste plaats betreffen zijn argumenten vooral processen die door Fodor wellicht tot de centrale processen zouden worden gerekend (bijvoorbeeld logica, schaken, beeldhouwen en het schrijven van computerprogramma's). Zelfs als Anderson gelijk heeft voor wat deze centrale processen betreft, zou Fodors modulariteitstheorie bijgevolg van toepassing kunnen blijven op meer perifere processen. In de tweede plaats is het niet ondenkbaar dat Andersons analyse zich op een ander niveau van beschrijving en verklaring afspeelt dan die van Fodor. Er zijn tal van niveaus waarop het triviaal waar is dat alle mentale processen zich bedienen van één universele onderliggende architectuur, zonder dat dit van invloed is op of van belang is voor de stelling dat (een deel van) de cognitieve architectuur modulair georganiseerd is. Dat geldt bijvoorbeeld voor het niveau van de biochemie van onze zenuwcellen. Elk van de door Fodor voorgestelde modules bedient zich, net als trouwens de centrale processen, van dezelfde onderliggende biochemische werkingsprincipes; in die zin is het triviaal waar dat elk onderdeel van ons kenapparaat gebouwd is uit een universele architectuur. Er zijn aanwijzingen dat Andersons analyse zich inderdaad op een ander niveau beweegt dan die van Fodor. Anderson probeert bijvoorbeeld te beargumenteren dat neuropsychologische gegevens over de localisatie van cognitieve functies irrelevant zijn voor de modulariteit van de geest. Hij trekt in dat verband een vergelijking tussen cognitieve functies en computerprogramma's. Net zoals twee programma's die zich op twee verschillende plaatsen in het computergeheugen bevinden dezelfde bouwprincipes kunnen hebben (zij zijn bijvoorbeeld geschreven in dezelfde taal, of maken gebruik van eendere procedures), zo kunnen volgens Anderson twee cognitieve functies op verschillende plaatsen in de hersenen gelocaliseerd zijn maar niettemin dezelfde architectuur bezitten.³² Hier blijkt duidelijk dat Andersons opmerkingen over een 'gemeenschappelijke architectuur' zich afspelen op een ander niveau dan Fodors opmerkingen over 'informatieel ingekapselde modules'; het feit dat twee programma's in dezelfde taal zijn geschreven zegt niets over de vraag of deze programma's ook toegang hebben tot elkanders input en output.

(III.) Tot slot een snelle blik op een *methodologische* lijn van kritiek. Het is bepaald niet ondenkbaar dat de vermeende modulariteit van een cognitief systeem in feite niets meer is dan een onbedoeld bijproduct van de onderzoeksstrategie die wordt gevolgd. Door de bank genomen zijn de cognitieve modellen die in de cognitiewetenschap worden ontwikkeld bijzonder specialistisch van aard. Het gaat meestal om modellen van betrekkelijk kleine, specifieke en welgedefinieerde onderdelen van een hypothetisch groter geheel. Dat groter geheel wordt betrekkelijk vaag gelaten en alleen in termen van zijn input/output-verbindingen met het subsysteem aangeduid. Het subsysteem komt er op deze manier weliswaar *uit te zien* als een module in Fodors zin van het woord, maar hoeft dat in werkelijkheid helemaal niet te zijn. De inkapseling en verzelfstandiging van het subsysteem zou ook een afspiegeling kunnen zijn van de specialisatie

³² Zie Anderson (1983), p. 8.

en verzelfstandiging van het onderzoek. Vooral in het geval van de psycholinguïstiek lijkt dit methodologisch gevaar levensgroot aanwezig te zijn. Linguïstische theorieën zijn vaak nog ingewikkelder dan de taal zelf en vergen overeenkomstig specialistische *know-how*. De neiging is groot om de specialistische kennis van de (psycho-)linguïst als het ware te projecteren in een afzonderlijk, in taal gespecialiseerd mentaal orgaan. Vanzelfsprekend biedt zo'n projectie geen enkele garantie voor de juistheid van de modulariteitstheorie.³³

7 Slotsom. Een conceptuele vectorruimte

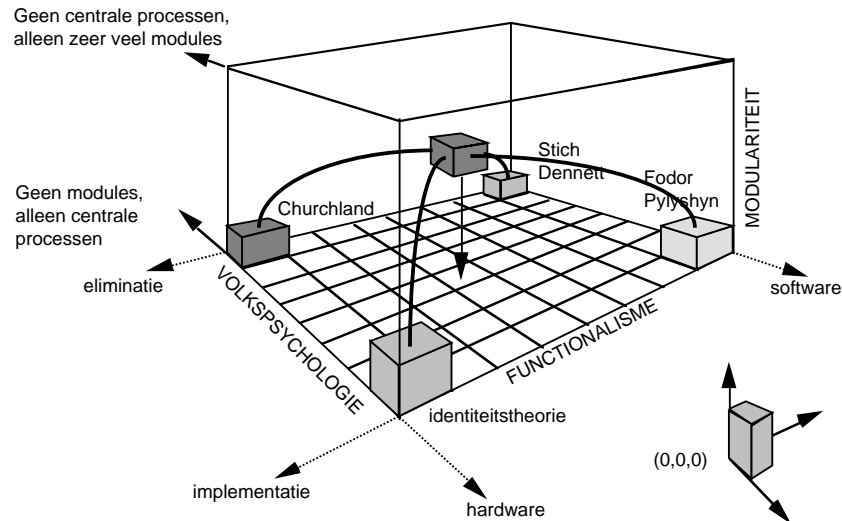
Hiermee zijn wij aan het eind gekomen van deze *sightseeing tour* van de grondslagen van de theorie van het kenapparaat. Het is overigens een vrij willekeurig einde, in de zin dat er nog tal van andere, niet minder interessante vraagstukken uit de filosofie van de cognitiewetenschappen aan de orde gesteld hadden kunnen worden. Zo hebben wij het zogenaamd 'methodologisch solipsisme' en de causale theorie van representatie links laten liggen, hebben wij het vraagstuk van de qualia nauwelijks aangeroerd, en is de controverse tussen de syntactische en de representatieve theorie van de geest alleen terloops ter sprake gebracht.³⁴ De onderwerpen die wél aan bod zijn gekomen geven echter een goede algemene indruk van het bijzonder levendige en boeiende strijdtonel dat de hedendaagse cognitiewetenschap en haar filosofie te bieden hebben.

Een van de constanten bij al deze controverses bleek de verstrengeling van filosofische en wetenschappelijke argumenten en posities te zijn. Filosofische theorieën ten aanzien van het mentale spelen zich niet af in een cognitief vacuüm, maar staan bloot aan de invloed van tal van wetenschappelijke en alledaagse denkbeelden en resultaten. Wetenschappelijk, empirisch succes van een bepaalde theorie van cognitie kan de met deze theorie geassocieerde filosofie een enorme duw in de rug geven. En anderzijds brengt het innemen van bepaalde filosofische posities uiteraard welbepaalde mogelijkheden en grenzen met zich mee ten aanzien van de ontwikkeling van wetenschappelijke theorieën. De filosofische positie kan als inspiratiebron dienen of met behulp van filosofische argumenten een zekere conceptuele beschutting bieden aan geestverwante lijnen van wetenschappelijk onderzoek. Op deze vormen van wisselwerking tussen filosofie en wetenschap zijn tal van varianten mogelijk, waarvan wij hierboven diverse voorbeelden zijn tegengekomen.

Vooral de modulariteitstheorie lijkt zich op dit moment te bevinden op het snijpunt van de grote wetenschappelijke en filosofische controverses in de cognitiewetenschap. Zij staat als het ware dwars op de overige discussies over de ontologie van het mentale. Dit kan zelfs vrij letterlijk worden opgevat, wanneer wij enkele van de voornaamste controverses weergeven in een 'conceptuele vectorruimte'. Tot de belangrijkste parameters bij de standpuntbepaling in de filosofie van de cognitiewetenschappen behoren op dit moment de status van de volkspychologie, het functionalisme en de modulariteit van cognitieve systemen. Wanneer deze parameters worden voorgesteld als onafhankelijke assen in een geometrische ruimte, kunnen de verschillende theoretische keuzemogelijkheden worden weergegeven in een driedimensionaal 'conceptueel universum' of een 'conceptuele vectorruimte', zoals afgebeeld in figuur 3. Concrete, specifieke theorieën, gekenmerkt door een welbepaalde verhouding tot de volkspychologie, een welbepaald niveau van functionele analyse en een welbepaalde graad van modulariteit, worden weergegeven door een *vector* in deze ruimte. Breder opgezette onderzoeksprogramma's, waarin ruimte wordt gelaten voor uiteenlopende specifieke theorieën, d.w.z. waarin de keuze

³³ Zie onder meer de bijdrage van Marslen-Wilson en Tyler in Garfield (1987), pp. 37vv. Vanuit een iets ander perspectief merkt ook McCauley (1987) op dat er onder psycholinguïsten een zekere neiging lijkt te bestaan om hun vakgebied te isoleren van ander psychologisch onderzoek en het aldus te behoeden voor usurpatie door de algemene psychologie.

³⁴ Vooral het methodologisch solipsisme en de naturalistische, causale theorieën van representatie verheugen zich op dit moment in een groeiende belangstelling. Zie ook de bijdrage van Sleutels & Geurts elders in deze bundel, Fodor (1987), Garfield (1988), alsmede Sleutels (1989).



Figuur 3

Een deel van de conceptuele vectorruimte van de filosofie van de cognitiewetenschappen

voor een bepaalde invulling van de parameters nog in meerdere of mindere mate wordt opengelaten, kunnen worden weergegeven als *lichamen* in de conceptuele ruimte.³⁵

In de vier uithoeken van het basisvlak in figuur 3 zijn enkele van de in dit artikel behandelde posities weergegeven: het eliminatief materialisme van Paul Churchland, het reductief materialisme van de (soort-)identiteitstheorie, het implementationeel functionalisme van Jerry Fodor en Zenon Pylyshyn (waarbij het aspect van modulariteit nog even buiten beschouwing is gelaten), en het eliminatief functionalisme van Daniel Dennett en Stephen Stich. Zoals hierboven op diverse plaatsen werd opgemerkt, is er tussen deze extremen ruimte voor tal van tussenposities. Wat betreft de verhouding tussen cognitiewetenschap en volkspychologie, is er tussen eliminatie en implementatie plaats voor (uiteenlopende gradaties van) revisie, d.w.z. gedeeltelijke overname, gedeeltelijke verfijning en gedeeltelijke eliminatie van volkspychologische ideeën. Wat betreft het niveau van functionele analyse dat relevant is voor de verklaring van cognitieve verschijnselen, is er tussen het allerlaagste niveau van de neurofysiologie (of hardware) en het allerhoogste niveau van abstracte cognitieve functies (of software) ruimte voor tal van tussenliggende niveaus, zoals beschreven in §§ 2-4.

Bij het innemen van tussenposities tussen de genoemde vier extremen doet zich echter het probleem voor dat er geen reden is om aan te nemen dat niet het *hele domein* van het mentale *over de volle linie* ofwel eliminatief ofwel implementationeel, en ofwel als hardware ofwel als software moet worden behandeld. Gezien vanuit het tweedimensionale perspectief van het basisvlak van figuur 3 is er geen *principiële, rationele* mogelijkheid om bijvoorbeeld sommige delen of aspecten van de cognitieve organisatie als hardware en andere als software te analyseren; elke keuze die hier gemaakt wordt lijkt willekeurig en *ad hoc* te zijn. Deze situatie verandert wanneer wij een derde parameter bij deze overwegingen betrekken, namelijk die van de modulariteit van het mentale. De invoering van een nieuwe as in de vectorruimte betekent een nieuwe dimensie van keuzemogelijkheden, en in het bijzonder een nieuwe manier om tussenliggende posities in het basisvlak rationeel te verantwoorden. Uitgaande van Fodors onderscheid tussen modulaire en centrale systemen, hoeven cognitieve verschijnselen niet langer *uniform* te worden

³⁵ Voor zover er naast de hier besproken drie parameters nog andere zijn, kan de vectorruimte van figuur 3 worden beschouwd als een driedimensionale *doorsnede* van een conceptuele ruimte van meer dimensies.

begrepen, maar kan een onderscheid worden gemaakt tussen twee grote groepen van verschijnselen met eigensoortige verklaringen. Zo is het alleszins voorstelbaar dat de diverse partijen in genoemde controverses zich zullen kunnen vinden in een theorie of programma waarin de door hen voorgestane eigenschappen op een bepaalde manier verdeeld zijn over modules en centrale systemen. Laten wij hier besluiten met het noemen van een drietal voor de hand liggende mogelijkheden:

- Materialisten en functionalisten zouden elkaar wellicht ergens halverwege tegemoet kunnen komen in een opvatting van het mentale waarin de werking van de neuraal gelocaliseerde modules vatbaar is voor een reductieve verklaring in neurofysiologische termen, terwijl de centrale processen en de globale, hiërarchische organisatie van de modules alleen een meer abstracte functionele analyse toelaten.

- Eliminativisten en implementationisten zouden elkaar wellicht kunnen vinden in een theorie waarin de werking van neuraal gelocaliseerde, *hardwired* modules volledig neurobiologisch kan worden verklaard, d.w.z. zonder beroep op de volkpsychologie, terwijl ingevolge Fodors 'First Law' de centrale processen zich opwerpen als ware bolwerken van volkpsychologisch verzet tegen elke vorm van eliminatieve analyse.

- Wij hebben gezien dat door sommige auteurs in twijfel wordt getrokken of het functionalistisch postulaat van een kenapparaat een afdoende verklaring van mentale processen kan bieden. Misschien kan het door de modulariteitstheorie gemaakte onderscheid tussen centrale en perifere processen de geesten ook op dit punt verenigen. De modulaire perifere processen met hun informationele inkapseling lijken zich immers bij uitstek te lenen voor een functionalistische en computationalistische analyse; de bezwaren tegen het computationalisme lijken vooral gericht op de centrale processen waar intentionaliteit, isotropie en allerlei andere moeilijke zaken een functionalistische analyse in de weg staan.

Uiteraard zijn dit hoogst speculatieve suggesties. Maar de gedachte dat de theorie van de verdeelde geest de geesten zou kunnen verenigen is toch eigenlijk té mooi om er niet voor te zwichten.

Bibliografie

- ANDERSON, JOHN, *The architecture of cognition*. Cambridge, MA: Harvard University Press 1983.
- BLOCK, NED, Troubles with functionalism. *Minnesota Studies in the Philosophy of Science* 9, 1978, 261-325.
- CHOMSKY, NOAM, *Rules and representations*. New York: Columbia University Press 1980.
- CHURCHLAND, PAUL M., *Scientific realism and the plasticity of mind*. Cambridge: Cambridge University Press 1979.
- CHURCHLAND, PAUL M., Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78, 1981, 67-91.
- CHURCHLAND, PAUL M., *Matter and consciousness. A contemporary introduction to the philosophy of mind*. Cambridge, MA: MIT Press 1984¹, 1988².
- CHURCHLAND, PATRICIA SMITH, *Neurophilosophy. Toward a unified science of the mind/brain*. Cambridge, MA: MIT Press 1986.
- DENNETT, DANIEL C., *Brainstorms. Philosophical essays on mind and psychology*. Montgomery, VT: Bradford Books 1978.
- DENNETT, DANIEL C., *The intentional stance*. Cambridge, MA: MIT Press 1987.
- FODOR, JERRY A., *Representations. Philosophical essays on the foundations of cognitive science*. Brighton, Sussex: Harvester Press 1981.
- FODOR, JERRY A., *The modularity of mind. An Essay on Faculty Psychology*. Cambridge, MA: MIT Press 1983.
- FODOR, JERRY A., *Psychosemantics. The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press 1987.
- GARDNER, HOWARD, *The mind's new science. A history of the cognitive revolution*. New York: Basic Books 1985.

- GARFIELD, JAY (ed.), *Modularity in knowledge representation and natural-language processing*. Cambridge, MA: MIT Press 1987.
- GARFIELD, JAY, *Belief in psychology. A study in the ontology of mind*. Cambridge, MA: MIT Press 1988.
- GOLDMAN, ALVIN, *Epistemology and cognition*. Cambridge, MA: Harvard University Press 1986.
- GOPNIK, IRWIN & MYRNA GOPNIK (eds.), *From models to modules. Studies in cognitive science from the McGill Workshops*. Norwood, NJ: Ablex Publishing Co 1986.
- HOFSTADTER, DOUGLAS & DANIEL DENNETT (eds.), *The mind's I. Fantasies and reflections on self and soul*. New York: Basic Books 1981.
- KITCHER, PATRICIA, Marr's computational theory of vision. *Philosophy of Science* 55, 1988, 1-24.
- LINDSAY, PETER & DONALD NORMAN, *Human information processing. An introduction to psychology*. New York: Academic Press 1977².
- LURIA, ALEKSANDR, *The working brain. An introduction to neuropsychology*. Harmondsworth, Middlesex: Penguin books 1973.
- MARR, DAVID, *Vision. A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman & Co 1982.
- MCCAULEY, ROBERT N., The not so happy story of the marriage of linguistics and psychology, or: Why linguistics has discouraged psychology's recent advances. *Synthese* 72, 1987, 341-353.
- PINKER, STEVEN & ALAN PRINCE, On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28, 1988, 73-193.
- PUTNAM, HILARY, *Mind, language and reality. Philosophical papers, volume 1*. Cambridge: Cambridge University Press 1975.
- PYLYSHYN, ZENON, *Computation and cognition. Toward a foundation for cognitive science*. Cambridge, MA: MIT Press 1984.
- QUINE, W.V.O., *Word and object*. Cambridge, MA: MIT Press 1960.
- RUMELHART, DAVID E., JAY L. MCCLELLAND, et al., *Parallel distributed processing. Explorations in the micro-structure of cognition*. Cambridge, MA: MIT Press 1986.
- SEARLE, JOHN, Minds, brains, and programs. *Behavioral and Brain Sciences* 3, 1980, 417-457.
- SEARLE, JOHN, *Intentionality. An essay in the philosophy of mind*. Cambridge: Cambridge University Press 1983.
- SLEUTELS, J.J.M., Eliminatief materialisme en de autonomie van de bottom-up benadering. *Algemeen Nederlands Tijdschrift voor Wijsbegeerte* 80, 1988, 41-62.
- SLEUTELS, J.J.M., Natuurlijke teleologie. Het probleem van misrepresentatie in de fysicalistische philosophy of mind. *Nijmegen Studies in the Philosophy of Nature and Its Sciences* 10, 1989 (in druk).
- STICH, STEPHEN, *From folk psychology to cognitive science. The case against belief*. Cambridge, MA: MIT Press 1983.