OP HET SNIJVLAK VAN COGNITIE, WETENSCHAP EN FILOSOFIE: EEN WETENSCHAPSFILOSOFISCHE DUIDING VAN HET DENKEN OVER INTER-THEORETISCHE RELATIES IN DE TWINTIGSTE EEUW

1. INLEIDING

In geen enkel wijsgerig vakgebied, zo schrijft de Amerikaanse filosoof John Searle, staan de opvattingen van leken en professionals verder van elkaar af dan in de filosofie van de geest, ofwel *philosophy of mind*.[1] Waar in de overige disciplines, zo meent hij, de meningen van beroepsfilosofen vaak aansluiten op, of een uitwerking zijn van algemeen gedeelde intuïties, is het in de filosofie die zich bezig houdt met de relatie tussen lichaam en geest goed gebruik om voor buitenstaanders evidente uitgangspunten als twijfelachtig terzijde te schuiven, of zelfs geheel te ontkennen. Dit geldt niet in de laatste plaats voor de vraag naar het bestaan van de geest zelf. Hoewel het voor de leek vaststaat dat hij, naast zijn lichamelijke bestaan, ook nog zoiets als geestelijke vermogens bezit, is de waarheid van deze ogenschijnlijk plausibele opvatting binnen de filosofie van de geest inzet geworden van een langdurig en controversieel debat.

Om allerlei redenen is het oude, cartesiaanse dualisme (de geest opgevat als een onafhankelijke, van de lichamelijke werkelijkheid onderscheiden substantie[2]) onhoudbaar gebleken.[3] Uit het stof dat na de ineenstorting van het dualisme neerdwarrelde, is het materialisme of fysicalisme als dominante positie naar voren gekomen.[4] De geest, zo is de consensus in de filosofie van de geest, heeft geen zelfstandige ontologische status, los van de fysische werkelijkheid: al het 'spul' in de werkelijkheid is materieel.[5] Filosofisch gezien is deze positie natuurlijk weinig aantrekkelijk als ze niet gepaard gaat met een relaas over hoe het dan mogelijk is dat de intuïtie dat er *wel* iets immaterieels bestaat, zo ontzettend hardnekkig is. Met andere woorden, hoewel de geest als aparte substantie uit de metafysische blauwdruk van de werkelijkheid is verbannen, dient een goede fysicalist iets te zeggen over hoe mentale *toestanden* of *eigenschappen*, zoals het voelen van pijn, het hopen op mooi weer etc. dan *wel* samenhangen met de fysieke werkelijkheid, als zij geen toestanden of eigenschappen kunnen zijn van een onafhankelijke geest. De verschillende posities die in de loop van de twintigste eeuw in de filosofie van de geest geformuleerd zijn, kunnen worden beschouwd als evenzoveel antwoorden op deze vraag.

Er vallen hier zo veel posities te onderscheiden dat een volledig overzicht buiten het bestek van dit artikel ligt. Deze versnippering is vooral te danken aan de manier waarop de filosofische gemeenschap, indachtig de erfenis van het logisch positivisme, de ontologische

---

[1] J.R. SEARLE, *Mind. A Brief Introduction*. New York/Oxford, Oxford University Press, 2004, p. 8.

[2] Strikt genomen is het substantiedualisme in de filosofie van Descartes een afgeleid dualisme. Dit heeft te maken met de aan de scholastiek ontleende definitie van substantie die Descartes hanteert: een substantie is datgene wat op zichzelf kan bestaan, d.w.z. datgene wat voor zijn bestaan niet afhankelijk is van iets anders. Eigenlijk is het alleen God die zelfstandig bestaat. Lichaam en geest zijn dan substanties in de zin dat zij voor hun bestaan *van niets anders afhankelijk zijn dan van God.*

[3] Voor een overzicht van de verschillende gebreken van het dualisme kunt u in bijna iedere inleiding in de filosofie van de geest terecht; zeer helder en compleet is D. BRADDON-MITCHELL and F. JACKSON, *Philosophy of Mind and Cognition. An Introduction.* Oxford, Blackwell Publishers ltd, 1996.

[4] De term 'fysicalisme' werd aanvankelijk gebruikt om de zogenaamde identiteitstheorie aan te duiden (deze theorie zal late aan bod komen) en deze te onderscheiden van het eerdere behaviorisme, waarmee het de metafysische positie van het materialisme deelde. Het precieze verband tussen deze twee termen zal in het vervolg van dit artikel hopelijk duidelijk worden.

[5] Er zijn natuurlijk uitzonderingen. Voor een hedendaagse verdediging van het dualisme zie P. FORREST, 'Difficulties with Physicalism, and a Programme for Dualists', in: H. ROBINSON (Ed.), *Objections to Physicalism*, Oxford, Clarendon Press, 1996.

vraag naar de verhouding tussen lichaam en geest in de twintigste eeuw heeft geherformuleerd tot een wetenschapsfilosofische vraag naar de verhouding tussen psychologische en neurologische theorieën. In die zin is de term 'geherformuleerd' trouwens niet helemaal adequaat. Immers, niet alleen de vraagstelling is veranderd, maar ook het onderwerp.

Hoe dan ook, er zijn heel wat verschillende posities betrokken ten aanzien van de relatie tussen de psychologie en de neurowetenschappen. Hanteren wij echter een grove kam, dan kunnen we twee kampen onderscheiden:

1) Het reductionisme. Volgens deze stroming zullen psychologische theorieën worden gereduceerd of herleid tot neurologische theorieën. Zoals we zullen zien lopen de meningen over wat dit precies behelst sterk uiteen. Hiermee verbonden is de gedachte dat de toestanden waaraan de psychologie refereert op één of andere wijze kunnen worden begrepen als breintoestanden.
2) Het antireductionisme. Volgens deze positie is er sprake van twee verschillende wetenschappen, ieder met een eigen onderzoeksveld, die naast elkaar kunnen opereren zonder dat gevreesd hoeft te worden voor de zelfstandige status van de psychologie. De mentale toestanden waaraan de psychologie refereert bestaan, maar zijn op een dergelijk intieme manier verbonden met fysieke eigenschappen, dat het fysicalisme als zodanig niet hoeft te worden opgegeven.[6]

Hoewel deze vete tussen reductionisten en antireductionisten in het domein van de menswetenschappen wellicht het felst uitgevochten wordt, manifesteert ze zich ook op andere gebieden. Dezelfde houdingen (autonomie versus reductie) kunnen worden ingenomen met betrekking tot de relatie tussen bijvoorbeeld de biologie en de scheikunde, of de scheikunde en de natuurkunde. Veel van de posities die aan bod zullen komen zijn te veralgemeniseren tot modellen van inter-theoretische relaties als zodanig.

Dit artikel beoogt een historisch overzicht te geven van de tweestrijd tussen reductionisme en antireductionisme, zoals die zich in de tweede helft van de twintigste eeuw heeft ontwikkeld ten aanzien van de cognitieve wetenschappen. In het bijzonder zal mijn aandacht uitgaan naar de wisselwerking tussen wetenschapsfilosofische en metafysische opstellingen. Ik zal betogen dat zich hierin een patroon van *toenemende nuance* aftekent: onder druk van objecties en tegenvoorbeelden hebben zich steeds gematigder modellen ontwikkeld, zodat de hierboven omschreven posities van radicale autonomie en reductionisme zich als twee extremen laten duiden, waartussen een grote verscheidenheid aan hybride posities mogelijk is. Met deze toenemende nuance gaat een neiging gepaard zich uitsluitend te verlaten op de concrete resultaten van wetenschappelijk onderzoek, in plaats van filosofische speculatie.[7] Dit doet uiteindelijk de vraag rijzen in hoeverre de filosofie op dit gebied nog een eigenstandige methode kan worden toegedacht: als men niet verder komt dan het interpreteren van de allernieuwste wetenschappelijke bevindingen, is het dan niet eerder de autonomie van de *filosofie* waar we ons om moeten bekommeren, in plaats van de psychologie?

2. HET IDEAAL VAN DE EENHEID VAN WETENSCHAPPEN: KLASSIEK REDUCTIONISME EN DE IDENTITEITSTHEORIE

[6] Het gebruikelijke concept om deze intieme relatie te verduidelijken is 'superveniëntie'. In het kort komt het er op neer dat A supervenieërt op B als een entiteit alleen de A- eigenschap heeft *omdat* het de B- eigenschap heeft. Als de B- eigenschap eenmaal present is, dan krijgt het de A- eigenschap er 'gratis' bij.

[7] Een belangrijke uitzondering hierop vormt het werk van Jaegwon Kim, die op metafysische gronden argumenteert dat het antireductionisme slechts vol te houden is als we de causale effectiviteit van mentale toestanden opgeven (epifenomenalisme).

Het klassieke reductionisme wordt steevast in verband gebracht met het werk van Ernest Nagel. In zijn lijvige hoofdwerk *The Structure of Science*, lanceert deze zijn model voor inter-theoretische reductie.[8] De oorspronkelijke motivatie voor dit model was gelegen in het door de logisch positivisten gepropageerde ideaal van de *eenheid van wetenschappen*: grofweg het streven naar een situatie waarin alle wetenschappelijke kennis één coherent geheel vormt. Vanuit dit ambitieuze onderzoeksproject is de interesse voor theoriereductie natuurlijk zeer begrijpelijk. Een ander positivistisch kenmerk van Nagels wetenschapsfilosofie vormt het verklaringsmodel dat hij hanteert, het zogenaamde D-N of deductief-nomologisch model van verklaringen.[9] Volgens dit model zijn wetenschappelijke verklaringen *deductief geldige redeneringen*. Hetgeen verklaard moet worden, het zogenaamde *explanandum*, wordt gededuceerd uit een aantal premissen, waarbij tenminste één van die premissen een formulering van een wet moet zijn.[10] Wanneer een steen valt, kan die gebeurtenis verklaard worden door te laten zien dat ze een instantie is van een algemene zwaartekrachtwet.

Nu doet deze relatie van afleidbaarheid zich niet alleen tussen particuliere gebeurtenissen en algemene wetten voor, *maar ook tussen wetten of theorieën onderling*. Nagel was met name geïnteresseerd in gevallen uit de wetenschapsgeschiedenis waar een reeds geaccepteerde theorie geabsorbeerd wordt door een nieuwe, bredere theorie; een gebeurtenis die volgens hem kenmerkend is voor de wetenschappelijke vooruitgang in het algemeen.[11] Bedenken we daarbij dat hij het D-N model van verklaring aanhing, dan zal het niet verbazen dat hij afleidbaarheid (*derivability*) van de gereduceerde theorie T_r uit de reducerende theorie T_b als voorwaarde stelt voor reductie. Er was echter nog een tweede voorwaarde: terminologische verbindbaarheid (*connectability*). Willen wij een bewering afleiden uit een tweede, algemenere bewering, dan moet de betekenis van de termen in die beweringen overeenkomen (*meaning invariance*). In het geval van theoriereductie wil dit zeggen dat de wetten van T_r en T_b zich van gedeeltelijk overlappende concepten bedienen. Nagel sprak in zulke gevallen van homogene reductie. Zo kunnen Galileo's wet voor vrij vallende lichamen in de nabijheid van het aardoppervlak en Keplers wet van planetaire beweging reductief verklaard worden door de Newtoniaanse mechanica, omdat beide wetten zich van dezelfde termen bedienen als de mechanica (e.g. afstand, tijd, massa, versnelling). Dit specifieke voorbeeld laat zien hoe reductie werkelijk een vooruitgang in onze kennis teweeg kan brengen: twee soorten beweging die tot dan toe van toepassing geacht werden op twee fundamenteel verschillende gebieden (respectievelijk aardse versus hemelse beweging) worden in Newtons mechanica samengebracht als twee instanties van één en dezelfde, algemene wet.[12]

---

[8] E. NAGEL, *The Structure of Science: Problems in the Logic of Scientific Explanation*, London, Routledge & Kegan Paul, 1961.

[9] De *locus classicus* hier is C. HEMPEL and P. OPPENHEIM, 'Studies in the Logic of Explanation', *Philosophy of Science* 15/1948, pp. 135-175.

[10] De term explanandum verwijst hier zowel naar de gebeurtenis die verklaard moet worden, als naar de propositie die de gebeurtenis beschrijft.

[11] Reductie betreft hier dus uitsluitend een relatie tussen theorieën, die begrepen kunnen worden als verzamelingen van beweringen (axioma's, wetten, empirische hypothesen etc.). Hoewel het verleidelijk is om over de reductie van objecten, eigenschappen of toestanden te praten en ik zelf in dit artikel ook nogal eens voor die verleiding bezwijk, is het toch goed om te beseffen dat de relata in eerste instantie altijd theorieën zijn. Wanneer we dus spreken over de reductie van temperatuur naar gemiddelde kinetische energie, dan moet dat begrepen worden als een reductie van *claims* over temperatuur tot *claims* over gemiddelde kinetische energie. Wat de ontologische implicaties van een houding ten opzichte van inter-theoretische relaties zijn is natuurlijk een gewichtige zaak en zal in het vervolg uitvoerig aan bod komen, *maar blijft niettemin een apart vraagstuk*.

[12] Toch zijn in de literatuur bij dit voorbeeld wel bezwaren aangetekend. Galileo's wet zegt dat de versnelling van een lichaam dat zich, in de nabijheid van het aardoppervlak, in een vrije val bevindt constant is. In de Newtoniaanse mechanica is deze versnelling juist *niet* constant, maar varieert ze met de afstand van het vallende

Er dient zich echter een probleem aan: in veel gevallen (vaak juist de meest interessant) bedient T_b zich van concepten die in T_r helemaal niet voorkomen. Met andere woorden, het vocabulaire van de theorieën die worden geacht in de reductierelatie tot elkaar te staan komt niet overeen. Er is dan sprake van een zogenaamde *heterogene reductie*.

Een schoolvoorbeeld is de reductie van thermodynamica naar statistische mechanica in de negentiende eeuw. Thermodynamica gaat uit van concepten als hitte, temperatuur en entropie – concepten die in de statische mechanica plaats moeten maken voor noties als waarschijnlijkheid en moleculaire beweging. In deze heterogene gevallen moeten zich onder de premissen waaruit de gereduceerde theorie wordt afgeleid *correspondentieregels* of *brugwetten* bevinden, die de terminologische kloof tussen de beide theorieën overspannen door inter-theoretische identificaties te maken. In termen van het bovengenoemde voorbeeld: de wet van Boyle-Charles valt te deduceren uit de principes van de statische mechanica *en* allerlei hulphypothesen, aannames en postulaten aangaande moleculaire samenstelling van gassen, de beweging van moleculen en de relaties tussen de noties temperatuur en kinetische energie. Zo kon volgens Nagel met een beroep op brugwetten ook in gevallen van heterogene reductie aan de voorwaarde van terminologische verbindbaarheid worden voldaan.

Het bovenstaande model van theoriereductie wordt in de cognitiefilosofie sterk geassocieerd met wat te boek is komen te staan als de *identiteitstheorie*. In de jaren vijftig en zestig lanceerde een aantal filosofen de these dat mentale toestanden niets anders dan lichamelijke toestanden zijn (belangrijke namen hier zijn Ullin Place, Herbert Feigl en John Smart).[13] Net als bij Nagels brugwetten, zou deze identificatie door niveaus van beschrijving heen moeten geschieden. Het klassieke voorbeeld: het ervaren van pijn is niets anders dan het vuren van C-vezels.[14] Het zal niet verbazen dat de aanhangers van deze opvattingen de nageliaanse positie ten opzichte van inter-theoretische relaties tot wetenschapsfilosofische legitimatie van hun ontologische stellingname zouden maken. Net zoals het voortschrijden der wetenschappen het concept temperatuur tot gemiddelde kinetische energie heeft gereduceerd, zo was de gedachte, zullen in de toekomst de mentale toestanden die in de psychologie worden gepostuleerd (hoop, geloof, pijn, kortom elke mentale toestand waar we ons maar in kunnen bevinden) worden gereduceerd tot breintoestanden. Het verdient vermelding dat de identiteitstheorie hiermee vatbaar lijkt voor empirische weerlegging en zo het domein van de metafysica overstijgt: ze is tegelijk een ontologische claim en een voorspelling over de toekomst van de psychologie als zelfstandige discipline.

Enfin, voortgestuwd door de teloorgang van het behaviorisme, won de identiteitstheorie gedurende de jaren zestig steeds meer aan populariteit. Vooral de gedachte dat de theorie voor empirische weerlegging vatbaar is viel goed bij de metafysisch gedesillusioneerde filosofen van de twintigste eeuw. Wie de ondergang van het oude dualisme ernstig nam en het materialisme meer dan lippendienst wilde bewijzen, leek met de

---

lichaam tot het massacentrum van de aarde. Dit zou betekenen dat Galileo's wet niet consistent is met Newtoniaanse mechanica, en eerder door deze laatste werd *vervangen* dan gereduceerd. Hoewel Nagels model hiermee niet gelijk op de schop hoeft, brengt deze discussie wel een belangrijke aanname aan het licht, namelijk dat theorieën en wetten waarheidswaarden hebben (d.w.z. ze kunnen waar of onwaar zijn). Een instrumentalist zal deze aanname verwerpen: voor hem zijn theorieën en wetten niet waar of onwaar, maar meer of minder gebruiksvriendelijk. Nu hoeven we natuurlijk geen instrumentalist te zijn, maar we doen er met het oog op wat volgt goed aan te beseffen dat het klassieke reductionisme, juist omdat het deductie centraal stelt, moeite heeft met gevallen waarin de gereduceerde theorie (gedeeltelijk) onwaar is. Later zal ik op dit punt terugkomen.

[13] Deze theorie is ook bekend onder de namen *central state materialism*, *reductive materialism* en *mind/brain identity theory*.

[14] Het zij hier opgemerkt dat deze simpele voorstelling van zaken vanuit een neurofysiologisch oogpunt hopeloos achterhaald is. De C-vezels maken onderdeel uit van het perifere deel van het somatisch sensorisch systeem, en zijn als zodanig niet eens in het brein aanwezig! Desalniettemin, voor de juistheid van de identiteitstheorie is slechts van belang dat er *een* neurale toestand identificeerbaar is met het hebben van pijn. Ik zal daarom gemakshalve deze formulering aanhouden.

identiteitstheorie zijn filosofische ankerplaats te hebben gevonden. Deze tevredenheid zou niet lang duren.

3. HET ANTIREDUCTIONISTISCHE PROGRAMMA: DE OPKOMST VAN HET FUNCTIONALISME EN DE COMPUTERMETAFOOR

Reeds in de jaren zestig werd er aan de poten van het klassieke reductionisme gezaagd. De kritiek kwam van meerdere kanten. Ten eerste kon het reductionisme niet uit de voeten met gevallen waarin de te reduceren theorieën onwaar zijn. Als reductie werkelijk een bijzonder soort wetenschappelijke vooruitgang markeert, dan ligt het voor de hand dit proces te duiden in termen van waarheid: de nieuwe theorie is dan correcter of waarachtiger dan zijn achterhaalde voorgangers. Deze ogenschijnlijk plausibele intuïtie vormt echter een groot probleem voor wie gelooft dat deductie een voorwaarde voor theoriereductie is: uit een ware stelling kan men immers geen onware deduceren. Een tweede probleem is de *intransitiviteit van verklaring*. Hillary Putnam gaf een sprekend voorbeeld: stel we hebben een houten blad met twee gaten erin, één rond met een diameter van een centimeter en één vierkant waarvan de zijden ook een centimeter lang zijn, en een vierkanten pin die net iets minder dan een centimeter hoog is. Het explanandum is nu het feit dat de pin niet in het ronde maar wel in het vierkante gat past. Men zou dit feit kunnen verklaren door te verwijzen naar de hardheid van de tafel en de pin, terwijl die hardheid op zijn beurt kan worden verklaard in termen van de microstructuur van het hout waar beide van gemaakt zijn. Het is echter niet zo dat een analyse van de microstructuur van het hout ook verklaart waarom de pin niet door het ronde gat gaat. Daarvoor is het nodig te verwijzen naar de geometrische eigenschappen van respectievelijk de pin en de gaten. Hoewel deze geometrische eigenschappen vanuit natuurkundig oogpunt willekeurig genoemd kunnen worden, zijn ze relevant voor de verklaring. Er gaat dus iets verloren in de gang van macro-eigenschappen van objecten naar moleculair niveau.[15]

Dan was er nog een probleem met die andere voorwaarde die Nagel stelde: verbindbaarheid. Beïnvloed door het werk van Gödel en Quine begon een nieuwe generatie filosofen zich af te zetten tegen erfenis van het logisch-positivisme, met name het logisch-propositionele kennismodel en het ideaal van de eenheid der wetenschappen. Onder wetenschapsfilosofen als Paul Feyerabend en Thomas Kuhn begon een consensus te groeien dat wetenschappelijke vooruitgang geen lineair proces is van accumulatie en reductie, maar juist wordt gekenmerkt door revolutie, conceptuele incommensurabiliteit en revisie van ontologie. Wanneer een oude theorie gereduceerd wordt door een nieuwe, dan gaat dit gepaard met verandering van betekenis. Zo heeft het concept massa in Newtoniaanse natuurkunde een andere betekenis dan in de quantumfysica. In de wetenschappelijke praktijk lijkt er dan ook nauwelijks een geval van reductie te zijn dat voltoet aan Nagels strenge eisen.

Waren deze redenen al genoeg om het klassieke reductionisme aan het wankelen te brengen, de conceptuele doodsteek zou uitgerekend komen vanuit de filosofie van de geest. In 1967 publiceerde Putnam zijn 'Psychological Predicates' – het functionalisme was geboren.[16] In zijn artikel betoogde Putnam dat mentale toestanden functionele toestanden zijn, in de zin dat ze kunnen worden gedefinieerd in termen van hun vermogen om causale verbanden te leggen tussen stimulus en respons, eventueel via andere mentale toestanden. Stel, iemand heeft de perceptie dat het regent. Dan kan die perceptie een mentale toestand tot gevolg hebben, zeg de overtuiging dat het regent, terwijl die overtuiging op zijn beurt weer gedrag

[15] H. PUTNAM, 'Philosophy and our mental life', in: H. PUTNAM, *Philosophical Papers 2*, New York, Cambridge University Press, 1975.
[16] H. PUTNAM, 'Psychological predicates', in: W.H. CAPITAN and D.D. MERRILL (eds.) *Art, Mind, and Religion*, Pittsburgh, University of Pittsburgh Press, 1967.

veroorzaakt (het schuilen onder een afdakje). Het hebben van een overtuiging is volgens de functionalist niets anders dan het verkeren in een toestand die in dergelijke causale relaties tot andere toestanden staat.

Één van de meest in het oog springende voordelen van dit nieuwe, functionalistische paradigma is de zogenaamde *meervoudige realiseerbaarheid* van mentale toestanden. Om dit voordeel op waarde te schatten is het nodig even terug te gaan naar de oude identiteitstheorie.

Een belangrijk bezwaar tegen de identiteitstheorie was dat het leed aan *neuraal chauvinisme*: als het ervaren van pijn identiek is aan een of andere fysisch-chemische toestand van ons brein *N*, dan lijkt daaruit te volgen dat wezens die niet dezelfde neurale structuur bezitten als wij en wier brein derhalve niet in toestand *N* kan verkeren, geen pijn kunnen ervaren. Omgekeerd *moeten* all wezens die in *N* kunnen verkeren ook pijn kunnen ervaren. Om een lang verhaal kort te maken, de identiteitstheorie stelt een onredelijk strikte eis door te verlangen dat er voor elke mentale toestand een concrete neurale toestand is die zowel noodzakelijk als voldoende is.

Voor het functionalisme ligt dit heel anders. Immers, al wat belangrijk is om in een mentale toestand te verkeren is dat er een bepaalde functie wordt uitgeoefend. *Wat* deze functie uitoefent is in dit opzicht niet relevant. Omdat een functie op verschillende manieren kan worden gerealiseerd (door menselijke hersenen, kattenbreinen, siliconenchips etc.) hoeft het functionalisme zich niet ontologisch te committeren.

De these van meervoudige realiseerbaarheid was koren op de molen voor diegenen die de autonomie van de speciale wetenschappen bepleitten en het duurde dan ook niet lang voordat men Putnams inzicht te gelde zou maken in het debat over de status van de speciale wetenschappen. In zijn geruchtmakende artikel 'Special sciences, or the disunity of science as a working hypothesis' werkte Jerry Fodor de wetenschapsfilosofische consequenties van de meervoudige realiseerbaarheidsthese uit.[17] Predicaten zoals die in de psychologie en andere niet-fundamentele wetenschappen worden gebruikt, corresponderen slechts met oneindig disjunctieve reeksen van fysische predicaten en disjuncties zijn geen natuurlijke soorten die door een wet geïdentificeerd kunnen worden. Dit betekent dat er van de inter-theoretische identificaties waar Nagels brugwetten voor moesten zorgen, geen sprake kan zijn. Aangezien de brugwetten nodig zijn om aan de voorwaarde van terminologische verbindbaarheid te kunnen voldoen, kwam hiermee het klassieke reductionistische programma op losse schroeven te staan.

Maar er is nog een ander facet van het functionalisme dat hier de aandacht verdient: *flexibiliteit van analyse*. Door zich te concentreren op de functie die iets heeft voor het geheel biedt het functionalisme een grote mate van conceptuele vrijheid om complexe systemen te analyseren. Vanuit functioneel oogpunt beschouwd is bijvoorbeeld een verbrandingsmotor een systeem met als input energie en als output beweging. Deze algemene functie kan worden onderverdeeld in een aantal kleinere stappen of subfuncties, bijvoorbeeld de viertact of 'Otto-cyclus': injectie, expansie, verbranding en uitlaat. Op dit niveau van beschrijving hoeft niet duidelijk te worden gemaakt hoe deze subfuncties worden geïmplementeerd. Het onderdeel 'verbranding' bijvoorbeeld, kan extern worden gerealiseerd, zoals in een stoomlocomotief, of intern, zoals in een moderne auto. Op een bepaald niveau van abstractie zijn een stoomlocomotief en een moderne auto dus functioneel equivalent: het zijn allebei systemen die de Otto-cyclus vertonen. Functionele analyse gebeurt dus door het onderverdelen van de te verklaren functie in steeds verfijndere subfuncties: ze gaat gepaard met *decompositie*.

Het aardige is nu dat dezelfde flexibiliteit ter beschikking staat van de cognitiewetenschapper, die immers ook tot taak heeft bepaalde functies of capaciteiten te analyseren: in zijn geval betreft het typisch cognitieve functies. In de jaren zeventig en tachtig

[17] J.A. FODOR, 'Special sciences, or the disunity of science as a working hypothesis', *Synthese* 28/1974, pp. 77-115.

van de twintigste eeuw genoot deze onderzoeksstrategie grote populariteit onder cognitief psychologen, die functionele analyses maakten van cognitieve functies als patroonherkenning, taalverwerving, numerieke cognitie, motorische vaardigheden etc. Voor zover computers door onderzoekers in de artificiële intelligentie (AI) kunnen worden getraind om dezelfde functies uit te oefenen (denk aan de schaakvermogens van Deep Blue), zijn wij op een zeker niveau van abstractie zelfs functioneel equivalent aan een computer![18]

Dit laatste brengt me bij de zogeheten *computationele theorie van de geest*, kortweg computationalisme. Een belangrijke inspiratie hiervoor was het onderscheid tussen semantiek (de inhoud of betekenis van symbolen) en syntaxis (de manipulatie van symbolen volgens vastomlijnde regels) dat was opgekomen tijdens de ontwikkeling van de moderne symbolische logica aan het begin van de twintigste eeuw: men kwam tot het inzicht dat de geldigheid van een redenering slechts afhangt van de manier waarop de premissen en conclusie met elkaar samenhangen en niet van hun specifieke inhoud. Kort gezegd, voor het manipuleren van symbolen zijn alleen de syntactische eigenschappen van belang, niet de semantische. Met de opkomst van de eerste computers in de jaren veertig en vijftig kwam daar een krachtige metafoor voor de menselijke geest bij. Het samengaan van dit conglomeraat van ideeën zou heel lang de agenda van de cognitieve psychologie bepalen. Net zoals een computerprogramma bestudeerd kan worden zonder dat de onderzoeker of programmeur verstand hoeft te hebben van de hardware, zo kon de psycholoog de geest, nu opgevat als symboolverwerkende machine, bestuderen zonder zich iets van de hersenen aan te trekken. We hebben hier in feite te maken met radicale autonomie: een extreme positie die ook wel omschreven wordt als methodologisch dualisme. Het verhaal van Fodor en de computationalisten is een verhaal van twee wetenschappelijke disciplines die zich in isolement van elkaar ontwikkelen.

Een laatste woord over metafysica. De boedelscheiding waaraan ik zo even refereerde zou wellicht de indruk kunnen wekken dat het oude cartesiaanse dualisme via een omweg weer is binnengehaald. Dit zou echter een misvatting zijn. Al is het functionalisme in strikte zin ontologisch neutraal, in de praktijk wordt ze fysicalistisch geïnterpreteerd. Zeker, ze is gekant tegen de identiteitstheorie en het klassieke reductionisme dat ermee gepaard ging, maar dat wil niet zeggen dat de functionalisten een terugkeer naar het oude dualisme bepleiten. Hoewel het niet zo is dat iedere soort van mentale toestand identiek is met een zekere soort van breintoestand (*type physicalism* of soort-fysicalisme), wil dat niet zeggen dat zo'n identificatie niet mogelijk is op het niveau van concrete instanties van die soorten (*token physicalism* of teken-fysicalisme). Natuurlijk zijn onze psychologische toestanden identiek aan bepaalde neurale configuraties, maar het soort wetmatige identificatie waar de oude identiteitstheorie op aanstuurde is uitgesloten. Via het functionalisme komen we dus uit bij een positie die twee claims met zich verenigt: een metafysische (teken-fysicalisme) en een wetenschapsfilosofische (non-reductionisme). Deze positie, aangeduid als *non-reductionistisch fysicalisme* (NRF) is binnen de *philosophy of mind* is tot op de dag van vandaag dominant gebleven.

---

[18] Wel rijst hier de vraag of een capaciteit als 'kunnen schaken' nog *cognitief* te noemen is, wanneer ze wordt uitgeoefend door een computer. Ik moet bekennen dat mijn intuïties elkaar op dit punt enigszins tegenspreken. Laat ik volstaan met een verwijzing naar het werk van de Britse filosoof Andy Clark over zogenaamde *uitgebreide cognitie* (extended mind/cognition): het idee dat ook externe zaken deel kunnen uitmaken van het cognitieve proces. Zo zou een adressenboekje, wanneer geraadpleegd door een patiënt met Alzheimer, letterlijk de rol van een extern geheugen kunnen spelen: de patiënt raadpleegt het boekje net zoals ik mijn geheugen raadpleeg (A. CLARK and D.J. CHALMERS, 'The extended mind', *Analysis* 58/1998, pp. 10-23). Hoewel ik hier tal van problemen zie (Als, in plaats van een notitieboekje, iemand anders mij helpt om iets te herinneren, maakt die persoon als geheel dan onderdeel uit van mijn cognitieve proces?) moet ik Clark toegeven dat het uit filosofisch oogpunt arbitrair is om een proces niet langer cognitief te noemen, louter omdat het niet in iemands schedel plaatsvindt.

4. ELIMINATIVISME, CONNECTIONISME EN REDUCTIONISME NIEUWE STIJL

Het omarmen van NRF is evenwel niet de enige manier om op de teloorgang van het klassieke reductionisme te reageren. Reeds geanticipeerd door denkers als Wilfred Sellars, Paul Feyerabend en Richard Rorty, kwam in de jaren zeventig en tachtig het zogenaamde *eliminatief materialisme* op. Net als het functionalisme, beaamt het eliminativisme onomwonden de tekortkomingen van het nageliaanse reductionisme, maar trekt daar een radicaal andere les uit: als mentale toestanden pogingen tot reductie weerstaan, kunnen ze beter worden *afgeschaft*. Het idee is hier dat de neurowetenschappen de psychologie zullen vervangen, eerder dan ze te reduceren. In een artikel uit 1981, dat net als Putnam en Fodors programmatische artikelen over het functionalisme uit de decennia ervoor met recht tot de *capita selecta* van de cognitiefilosofie kan worden gerekend, betoogde Paul Churchland dat het moderne debat over de status van de psychologie en de mentale toestanden waar ze in haar verklaringen naar verwijst, nog steeds gevangen zit in wat hij *volkspsychologie* noemde.[19] Deze volkspsychologie betreft het geheel van onze doorgaans geaccepteerde, *common-sense* opvattingen over de geest. Het is als het ware een proto-wetenschappelijk raamwerk waarmee we het gedrag van anderen en onszelf kunnen verklaren en voorspellen.

Zo zullen de meesten van ons, wanneer we horen hoe de grote boze wolf het huisje van de drie biggetjes omver probeert te blazen, niet schromen dit gedrag te verklaren door naar een mentale toestand van het ondier te verwijzen, namelijk de begeerte om vanavond koteletten te eten. Maar dat is nog niet alles. We zijn gewoon deze mentale toestanden te interpreteren in termen van een 'denktaal' en, meer in het bijzonder, van *propositionele attitudes.* In het onderhavige voorbeeld zeggen we eigenlijk dat het subject (de wolf) zich op een bepaalde epistemische manier verhoudt tot een interne representatie of symbool waaraan een bepaalde betekenis toekomt ("ik heb trek in koteletten"). Met andere woorden, het vermeende mentale proces dat ten grondslag zou liggen aan het gedrag van de wolf bestaat volgens de volkspsycholoog uit het verwerken of lezen van die interne symbolen of representaties. Deze talige interpretatie van mentale processen deelt de volkspsychologie met de klassieke cognitieve psychologie, die de menselijke geest immers ziet als een weliswaar neuraal geïmplementeerde, maar toch onafhankelijke symboolverwerkende machine. De metafoor van de computer, die een formele taal bestaande uit enen en nullen manipuleert, versterkte deze gedachte alleen maar.

Het eliminativisme kwam hiertegen in opstand. De mentale toestanden en de interne representaties en symbolen, zo claimden zij, corresponderen nergens mee en dienen opgegeven te worden. De volkspsychologie is niets anders dan een gedegenereerd onderzoeksprogramma, waarin, ofschoon het nu al een geschiedenis van duizenden jaren heeft, nauwelijks enige vooruitgang te bespeuren valt. In alle andere wetenschappelijke disciplines, zo argumenteert Churchland, zijn onze volkstheorieën incorrect gebleken. Zo geloofden we ooit dat de zon door Helios op zijn kar langs de hemel getrokken werd, dat mislukte oogsten werden veroorzaakt door ontstemde goden en dat epileptische aanvallen werden opgewekt door boze geesten. In al deze gevallen zijn onze aanvankelijke, proto-wetenschappelijke opvattingen onjuist gebleken en hebben we met de opkomst van nieuwe, wetenschappelijke alternatieven niet alleen onze volkstheorieën vaarwel gezegd, maar ook het geloof in de entiteiten die door die theorieën werden gepostuleerd opgegeven. Hoe wonderbaarlijk zou het dan zijn als we het met betrekking tot de geest van het begin af aan bij het juiste eind hadden? Nee, besluit de eliminativist, de door de volkspsychologie

[19] P. CHURCHLAND, 'Eliminative materialism and the propositional attitudes', *Journal of Philosophy* 78/1981, pp. 67-90.

gepostuleerde mentale toestanden zijn niets anders dan luchtkastelen, verzinsels die door toedoen van de eenmaal volwassen geworden neurowetenschappen onherroepelijk op de vuilnisbelt van overbodige en wetenschappelijk achterhaalde ideeën zullen belanden, tezamen met heksen, phlogiston en het geocentrisch wereldbeeld.

Het spreekt voor zich dat het eliminativisme een aantal belangrijke tekortkomingen van het klassieke reductionisme vermijdt, juist omdat ze de eisen van afleidbaarheid en verbindbaarheid laat varen. Daarnaast won ze aan populariteit door een nieuw, veelbelovend onderzoeksprogramma dat rond de jaren zeventig opkwam in de cognitieve wetenschap en dat een alternatief bood voor de computationalistische modellen van de traditionele AI: het *connectionisme*. In tegenstelling tot de door de computermetafoor geïnspireerde, computationele modellen van cognitieve capaciteiten, die de geest begrijpen als een machine die interne symbolen manipuleert volgens vaste, voorgeprogrammeerde regels die uitgevoerd worden door hiërarchisch geordende onderdelen, verspreiden connectionistische modellen binnenkomende prikkels over een groot aantal gelijkwaardige knopen, die elkaar inhiberen of exciteren, zodat het netwerk ten slotte voor iedere input een unieke configuratie van activatie-graden vertoont.

Het valt te begrijpen dat dit nieuwe model voor menselijke cognitie de voorkeur genoot van eliminativisten, die al snel de voordelen van het connectionisme ten opzichte van de klassieke modellen zagen. Zo zijn neurale netwerken vele malen beter in het herkennen van patronen, vormen ze een getrouwere weergave van onze hersenen en zijn ze veel beter tegen beschadiging bestand dan computationele modellen met hun seriële architectuur. Toch is het enthousiasme voor neurale netwerken de laatste tijd weer bekoeld. De beloofde spectaculaire resultaten lieten langer op zich wachten dan gehoopt en hoewel ze neuraal aannemelijker zijn dan de oude computationele modellen, zijn ze nog steeds simpel en in hoge mate geïdealiseerd ten opzichte van het menselijk brein. Al deze technische details hoeven ons hier niet op te houden. Waar het om gaat is dat de controverse tussen klassieke en connectionistische modellen heeft laten zien dat er een alternatief is voor de talige, op symboolmanipulatie berustende interpretatie van cognitieve processen die zo dominant was in de klassieke cognitieve psychologie.

Terug naar het eliminativisme dat zich, net zoals zijn klassieke voorganger, lijkt te vergalopperen. Laat ik me beperken tot een tweetal problemen. Ten eerste werkt de wetenschap niet mee. Het lijkt simpelweg onwaar te zijn dat al onze volkstheorieën zijn geëlimineerd door de voortschrijdende wetenschap. Wetenschappelijke theorieën komen op, zijn aan verandering onderhevig, worden getest en weer verworpen, terwijl de ‘naïeve’ volkstheorieën bestendig zijn gebleken. Bovendien zijn ze uiterst succesvol: door mensen intenties, gedachten en voorkeuren toe te schrijven zijn we in de meeste gevallen goed in staat hun gedrag te voorspellen. Sterker, het hanteren van dit volkspsychologische idioom is wat het voor ons überhaupt mogelijk maakt ons in sociale situaties te bewegen. Vanuit dit perspectief kunnen de bezwaren van Churchland en andere eliminativisten eerder als aansporingen worden gezien om het karakter van onze volkstheorieën opnieuw te overdenken. Wellicht moeten we onze volkse opvattingen niet zien als concurrenten voor wetenschappelijke theorieën, maar eerder als leidraad of beginpunt voor verder wetenschappelijk onderzoek. Zij maken voorlopige indelingen, doen enkele praktische aanbevelingen en leveren de explananda aan.

Ten tweede, er is onenigheid over de vraag wanneer we nu precies gelegitimeerd zijn concepten te laten vallen. Één stroming, waartoe Churchland behoort, zegt dat de we niet langer over mentale toestanden moeten spreken wanneer we ontdekken dat ze met niets in de werkelijkheid corresponderen. Zoals we net gezien hebben is dit vanuit historisch oogpunt weinig plausibel. Een andere stroming houdt de mogelijkheid open dat concepten zoals die in de volkspsychologie gangbaar zijn wel degelijk naar iets in de werkelijkheid verwijzen, maar

claimt dat dit iets in de werkelijkheid niets anders is dan een breintoestand. Zodra wij deze identiteit vaststellen en we dus een wetenschappelijk alternatief hebben, kunnen we het verder stellen zonder het originele concept. Een probleem met deze laatste stroming is dat het niet duidelijk is in wat voor opzicht ze nog verschilt van de (teken) identiteitstheorie.

Zoals de laatste objectie al suggereert, is het eigenlijke probleem van het eliminativisme niet zozeer dat er nooit zoiets als theorie-eliminatie voorkomt in de wetenschap, maar eerder de manier waarop ze gepresenteerd wordt als *het* model voor wetenschappelijke vooruitgang. Het lijkt verstandiger om eliminatie als *een* mogelijke vorm van vooruitgang te zien en daarnaast ruimte te laten voor andere manieren waarop theorieën zich tot elkaar kunnen verhouden.

Dit laatste is precies wat er gebeurt in de jongste loot van het reductionisme: reductionisme-nieuwe-stijl (*new wave reductionism*) of RNS, waarvan John Bickle de voornaamste protagonist is.[20] RNS poogt een probleem op te lossen dat we in de bespreking van het klassieke reductionisme al zijn tegengekomen: hoe om te gaan met gevallen van reductie waarin de gereduceerde theorie T_r (gedeeltelijk) onwaar is. Bickle lost dit op door eerst een gecorrigeerde versie T_r* van de te reduceren theorie te construeren, in termen van de reducerende theorie T_b.[21] Met andere woorden, voordat de reductie plaats kan vinden moet er eerst gesleuteld worden aan de oude theorie, zodanig dat de herschreven versie zonder problemen gededuceerd kan worden uit de nieuwe theorie. Op deze manier worden de problemen geassocieerd met afleidbaarheid en verbindbaarheid van terminologie vermeden: RNS heeft geen brugwetten meer nodig, aangezien het vocabulaire van de twee theorieën per definitie de juiste overlap vertoont.

Hier blijft het echter niet bij. Bickle belicht een intrigerend facet van RNS dat het werkelijk boven zijn voorgangers doet uitstijgen. Als we toestaan dat de *mate waarin gesleuteld* moet worden aan T_r varieert, dan hebben we zo een manier in handen om individuele gevallen van reductie op een continuüm te ordenen. Aan de ene kant van dit continuüm staan gevallen waarin weinig tot geen veranderingen hoeven te worden aangebracht aan de te reduceren theorie. Dit zijn voorbeelden van 'gladde' reductie, waarin er een directe identiteit vastgesteld kan worden. Hier is sprake van retentie, van behoud van ontologie. Aan de andere kant van het spectrum staan gevallen waarin de te reduceren theorie grondig moest worden aangepast, zodanig zelfs dat er sprake is van eliminatie van ontologie: dit zijn voorbeelden van 'stroeve' reductie. Een concreet geval van theoriereductie zal zich ergens tussen deze twee polen bevinden.

RNS is dus een stuk verfijnder dan zijn voorgangers en lijkt een gezonde balans te vinden tussen het klassieke reductionisme enerzijds en het eliminatief materialisme van Churchland anderzijds. Het markeert een serieuze vooruitgang ten opzichte van zijn voorgangers, omdat het Bickle in staat stelt om uiteenlopende gevallen van theoriereductie onder te brengen in een overzichtelijk schema, dat juist in dit vermogen om voorbeelden uit de wetenschappelijke praktijk te duiden zijn rechtvaardiging vindt, in plaats van zich te beroepen op metafysische argumentatie. Dit gaat zelfs zover dat Bickle actief probeert om zich van metafysica te distantiëren. Tegenwoordig spreekt hij dan ook liever over *metawetenschap*:

> One feature of my new wave metascience deserves explicit discussion. It eschews all *traditional* concern with ontology and "metaphysics" […] The job of new wave metascience is simply to illuminate concepts like reduction as these imbue actual scientific practice. To what end? *Not* to achieve some better way of addressing reformulated "external" questions about the existence and nature of "theory-independent ontology." Rather, because a reasonable explanatory goal is to

[20] J. BICKLE, *Psychoneural Reduction, The New Wave*, Cambridge Massachusetts: MIT Press, 1998.
[21] Suggesties in deze trant werden eerder gedaan door Kenneth Schaffner en Clifford Hooker, hoewel Schaffner voorstelde om T_r* te construeren uit T_r in plaats van T_b. Bickles model sluit dus aan bij dat van Hooker.

> understand practices "internal" to important current scientific endeavors and the scope of their potential application and development. The tasks of this book are part of a *metascience* of contemporary psychology and neurobiology, not part of some "ontology of mind."[22]

5. Naar een wetenschappelijk geïnformeerde middenpositie: verklaringspluralisme, heuristische identiteiten en de mechanistische beweging

Toch lijkt ook RNS niet zaligmakend te zijn. Bickles continuüm ruimt beoogt weliswaar plaats te bieden aan uiteenlopende gevallen van theoriereductie, toch behandelt het al die gevallen als *opeenvolgend in tijd*: oude theorieën worden gecorrigeerd of vervangen door nieuwe theorieën. Het is echter ook mogelijk dat twee theorieën *tegelijkertijd* bestaan en onderling invloed op elkaar uitoefenen. Er is dan sprake, niet van diachronische opvolging, maar van synchronische co-evolutie. Zo hebben de klassieke mendeliaanse genetica en de biochemische genetica decennia lang naast elkaar bestaan. Het betreft hier eigenlijk twee onderzoeksprogramma's die op naburige niveaus van analyse elkaar wederzijds beïnvloeden. Zo droeg de klassieke genetica indertijd de explananda voor de biochemische genetica aan en hielp zij haar vraagstellingen te preciseren, terwijl biochemici hypothesen voorstelden over welke structuren nu eigenlijk verantwoordelijk waren voor de door Mendel beschreven patronen, zoals behoud van eigenschappen doorheen één of meerdere generaties.

Dit laatste voorbeeld illustreert een belangrijke *modus operandi* van wetenschappers die onderzoek doen op naburige niveaus van beschrijving: het opwerpen van hypothetische of heuristische identiteiten. Een tentatieve brug wordt opgeworpen tussen twee domeinen. Het betreft hier echter geen ontologische stellingname: de voorgestelde identiteit wordt net zo lang volgehouden als zij vruchten afwerpt en het onderzoek verder helpt. Blijkt zij echter geen resultaten te boeken of zelfs vooruitgang in de weg te staan, dan wordt ze net zo makkelijk weer verlaten.[23]

Zo komen we uit op een zeer tolerante positie ten aanzien van inter-theoretische relaties. Er bestaan vele niveaus van beschrijving, elk met hun eigen methoden, concepten en verklaringstechnieken. Wat op één niveau werkt hoeft niet per se op een ander niveau ook te werken.[24] Hoewel dit verklaringspluralisme antireductionistisch genoemd mag worden, markeert zij evenwel een belangrijke stap voorwaarts in het autonomiekamp. Net zoals RNS een nuancering inhield van het klassieke reductionisme, is het verklaringspluralisme een stuk gematigder dan het rigide methodologisch dualisme van Fodor en de computationalisten. Zeker, twee theorieën kunnen naast elkaar bestaan, maar dat wil niet zeggen dat zij zich ook in isolatie van elkaar ontwikkelen. Door een plaats te geven aan interdisciplinaire kruisbestuiving is het voor de verklaringspluralisten mogelijk een belangrijke bron van wetenschappelijke vooruitgang in hun model te accommoderen.

Zoals we bij de bespreking van het reductionisme ook al zagen, vertoont het antireductionisme eveneens een tendens tot nuancering. Met het verstrijken van de tijd hebben de twee kampen de neiging zich, onder druk van voorbeelden uit de wetenschappelijke praktijk, naar elkaar toe te bewegen. In dit verband kan de mechanistische beweging niet onbesproken blijven, omdat zij feitelijk een programmatische invulling geeft aan dit oecumenische project.

---

[22] J. Bickle, *Philosophy and Neuroscience. A Ruthlessly Reductive Account*, Dordrecht, Kluwer, 2003, pp. 31-32.

[23] R.N. McCauley and W. Bechtel, 'Explanatory pluralism and the heuristic identity theory', *Theory and Psychology* 11/2001, pp. 736-760

[24] R.N. McCauley, 'Explanatory pluralism and the coevolution of theories in science', in: R.N. McCauley (ed.), *The Churchlands and their critics*, Oxford, Blackwell, 1996, pp. 17-47.

Herinneren we ons de flexibiliteit van analyse van het functionalisme: de mogelijkheid om systemen te analyseren in termen van steeds verfijnder subfuncties, zonder dat het noodzakelijk is te specificeren hoe een gegeven functie ontologisch geïmplementeerd is. Hoewel deze onderzoeksstrategie, zoals reeds vermeld, in de jaren zeventig en tachtig van de twintigste eeuw de cognitieve psychologie domineerde, is ze de laatste tijd aan erosie onderhevig. Wat aanvankelijk als één van de grootste pluspunten van het functionalisme werd gepresenteerd, de grote mate van abstractie, wordt tegenwoordig juist als zwakte beschouwd. Het gaat er in de cognitiewetenschappen immers om te verklaren hoe cognitieve functies *door mensen* worden gerealiseerd, niet door computers, zombies of bewoners van Putnams hypothetische tweelingaarde. Willen we bijvoorbeeld de verminderde motorische vaardigheden van een patiënt met hersenletsel herstellen, dan doen we er goed aan als we op de hoogte zijn van details over de neurale implementatie van die vaardigheden. Een puur functionele analyse die de vaardigheid onderverdeeld in allerlei hypothetische deelvaardigheden blijkt hier weinig effectief.

Een belangrijke stroming in het hedendaagse denken over neurobiologie heeft de taak op zich genomen deze lacune te verhelpen en de functionele modellen een concretere invulling te geven: de mechanistische beweging. Hoewel begrippen als ‘mechanisme’ en ‘mechanistische verklaring’ al veel langer geleden hun weg naar het filosofisch discour over de menswetenschappen hadden gevonden, nam de beweging pas echt een vlucht rond de millenniumwisseling met een beroemd artikel dat nu al tot de klassieken gerekend mag worden: ‘Thinking about Mechanisms’, geschreven door Peter Machamer, Lindley Darden en Carl Craver.[25] Wat wetenschappers in werkelijkheid doen als ze een cognitieve functie verklaren, zo betogen de auteurs, is dat ze het voor die functie verantwoordelijke mechanisme beschrijven. Een mechanisme bestaat uit *delen* en hun *activiteiten* die zo georganiseerd zijn dat ze samen het explanandum (de cognitieve functie) realiseren.

Zo kunnen we bijvoorbeeld een mechanistische verklaring geven van osmoregulatie, ofwel het vermogen van het lichaam om vocht tussen de cellen vast te houden (het ECV of extracellulaire volume). Cellen worden door het ECV van zuurstof en voedingstoffen voorzien en voeren hun afvalstoffen erdoor af. Wanneer we zout voedsel tot ons hebben genomen of getranspireerd hebben zonder het vocht te hebben aangevuld, heeft dit een stijging van de osmolariteit tot gevolg (de moleculaire samenstelling van het bloed verandert). Dit wordt opgemerkt door osmoreceptoren in de hypothalamus, die daardoor vasopressine of antidiuretisch hormoon (ADH) aanmaakt. Dit hormoon reguleert de variatie van waterafscheiding via urine en prikkelt het dorstcentrum, zodat we gaan drinken. Al deze subfuncties, onderdelen en hun organisatie werken dus samen om de capaciteit osmoregulatie te realiseren.

Twee aspecten zijn hier van belang. Ten eerste: de verklaring die hier gegeven wordt refereert aan entiteiten en processen op verschillende niveaus beschrijving, van moleculaire veranderingen in de samenstelling van het bloed tot wijzigingen in het gedrag. Nadat het explanandum eenmaal vaststaat, is er geen principiële keuze voor één niveau van beschrijving meer te maken. De verklaring gaat op en neer, langs vele niveaus van beschrijving om die entiteiten en activiteiten er uit te pikken die van belang zijn. Niveaus kunnen hier dus alleen maar lokaal, dat wil zeggen met betrekking tot het gegeven mechanisme, gedefinieerd worden.[26]

---

[25] P.K. MACHAMER, L. DARDEN and C.F. CRAVER, ‘Thinking About Mechanisms’, *Philosophy of Science* 67/2000: pp. 1-25.

[26] Hier gaan enkele noties van het begrip ‘niveau’ vervelend door elkaar lopen. Enerzijds zijn we, vanuit de traditie van *theorie*reductie die een gang van specialere wetenschappen naar fundamentele natuurkunde voorspelt, geneigd over niveaus na te denken in termen van wetenschappelijke manieren van beschrijving of disciplines: ruwweg het rijtje sociologie, psychologie, neurofysiologie etc. Anderzijds hanteren we in het

Ten tweede: een mechanistische verklaring behelst, net als de abstracte functionele verklaring, onder andere de decompositie van een functie in subfuncties. Een functie (een balans in het extracellulair volume te behouden) wordt opgesplitst in een aantal deeltaken (het detecteren van schommelingen in de osmosewaarden, de aanvoer van nieuwe vloeistoffen etc.). Echter, ze gaat verder dan een puur functionele analyse: ze identificeert voor elke subfunctie ook door welke entiteit ze wordt uitgevoerd (osmoreceptoren en ADH respectievelijk). Het moge duidelijk zijn dat bij het behandelen of voorkomen van verstoringen van de vochthuishouding (zoals bij diabetes) dergelijke kennis onmisbaar is.

Enfin, een mechanistische verklaring doet dus meer dan de verschillende onderdelen van een proces ontleden, ze doet ook een poging de uitvoerende instanties te identificeren. Hoe correcter deze identificatie verloopt, hoe groter de verklarende kracht van het model, zo is de gedachte. We krijgen zo een continuüm dat zich lijkt te spiegelen aan Bickles spectrum van theoriereductie: van *mogelijke*, via *plausibele* tot *feitelijke* modellen.[27] Aan de ene kant van het spectrum staan modellen die mogelijk aangeven hoe het mechanisme werkt. Dit zijn speculatieve, abstracte weergaven van hoe het mechanisme dat verantwoordelijk is voor een cognitieve capaciteit er uit zou *kunnen* zien. Daarna komen plausibele modellen. Hoewel nog steeds hypothetisch, zijn deze modellen opgesteld met medeneming van additionele informatie. Zo kan het bijvoorbeeld zijn dat op grond van nieuw, experimenteel bewijs bepaalde entiteiten kunnen worden uitgesloten als uitvoerders van een gegeven deeltaak. Aan de andere zijde van het spectrum vinden we dan feitelijke modellen. Dit zijn ideale, complete beschrijvingen van het mechanisme zoals dat in de werkelijkheid voorkomt. Hoewel de abstracte modellen zo hun nut kunnen hebben, bestaat wetenschappelijke vooruitgang volgens de mechanisten daarin, dat we een aanvankelijk speculatief model geleidelijk aan steeds verfijnder en waarheidsgetrouwer maken, met de complete weergave van het feitelijke mechanisme als regulatief ideaal.

6. SLOTWOORD OVER DE FILOSOFISCHE METHODE

Tot zover de ontwikkelingen die het denken over inter-theoretische relaties in de twintigste eeuw heeft doorgemaakt. Zoals RNS een belangrijke nuancering inhoudt ten opzichte van het klassieke reductionisme, zo is ook het beeld dat de verklaringspluralisten ons voorhouden veel gematigder dan het radicale methodologisch dualisme van Fodor. De twee stromingen die aanvankelijk zo tegenover elkaar stonden zijn dus als het ware naar elkaar toegegroeid. Het is niet zo dat de psychologie constant in zijn autonomie bedreigd wordt door de neurowetenschappen, maar evenmin is het correct om te zeggen dat de twee disciplines zich in volkomen isolatie van elkaar ontwikkelen. Prominente denkers van beide kampen hebben ingezien dat hun aanvankelijke modellen faalden *juist omdat het algemene modellen waren*:

---

opstellen van mechanistische verklaringen een mereologisch begrip van niveaus, dat wil zeggen, we begrijpen niveaus in termen van de grootte van de entiteiten die we beschrijven: een geheel wordt verklaard door naar het niveau van de delen af te dalen. Zoals Bechtel terecht opmerkt zijn deze twee noties niet met elkaar te verenigen (W. BECHTEL, 'Reducing psychology while maintaining its autonomy via mechanistic explanations,' in: M. SCHOUTEN and H. LOOREN DE JONG (eds.), *The Matter of the Mind: Philosophical Essays on Psychology, Neuroscience, and Reduction*, Oxford, Blackwell Publishing, 2007, pp. 172-198). Natuurkunde, als wetenschappelijke discipline, bestudeert entiteiten variërend van subatomische deeltjes tot complete sterrenstelsels, terwijl bijvoorbeeld de biologie zich zowel buigt over virussen als over ecosystemen. De beschrijvingsnotie en de mereologische zin van niveaus zijn dus niet met elkaar te verenigen. De problemen hier zijn uiterst complex en een meer dan oppervlakkige behandeling ervan zou een apart artikel in beslag nemen. Voor een heldere en uitvoerige uiteenzetting over deze problematiek en een poging ze te overwinnen, zie hoofdstuk vijf van C.F. CRAVER, *Explaining the Brain*, Oxford, Clarendon Press, 2007, pp. 163-195.

[27] C.F. CRAVER, 'When mechanistic models explain', *Synthese* 153/2006, pp. 355-376.

het bleek zeer moeilijk iets betekenisvols te kunnen zeggen dat voor alle gevallen van inter-theoretische relaties opgaat. Wellicht is de wetenschappelijke praktijk simpelweg te grillig en te dynamisch om gevangen te kunnen worden in termen van één statisch model. Gegeven deze stand van zaken lijkt pluralisme de aangewezen weg. In concrete gevallen kan er sprake zijn van reductie, eliminatie of wederzijdse beïnvloeding van theorieën, maar deze gevallen laten zich niet veralgemeniseren.

Dit doet echter wel de vraag rijzen welke rol de filosoof in dit alles nog kan spelen. Als elke poging iets algemeens over inter-theoretische relaties te zeggen ofwel door tegenvoorbeelden op de klippen loopt, ofwel verzandt in trivialiteiten, dan is het moeilijk om in te zien wat een specifiek filosofische analyse hier nog kan toevoegen. In dit verband is het veelzeggend dat het proces van matiging en nuancering waarvan in dit artikel gewag werd gemaakt vooral gedreven wordt door een toenemende aandacht van filosofische kant voor concrete voorbeelden uit de wetenschappelijke praktijk. Dit gaat zelfs zo ver dat de wetenschapsfilosofie, traditioneel begrepen als de conceptuele reconstructie van die praktijk, een slechte naam heeft gekregen. Filosofen lijken tegenwoordig over elkaar heen te buitelen in hun poging om zich toch vooral te distantiëren van hun eigen vakgebied. Bickles metawetenschap is hiervan wellicht het meest drieste voorbeeld. Waar Quine nog zei: "Philosophy of science is philosophy enough," luidt het motto van de huidige generatie cognitietheoretici eerder: "Enough philosophy already!" Durven we in deze omstandigheden nog te vragen naar de toekomst van de *good old philosophy of mind*?

Toch is deze voorstelling van zaken te somber. Zeker, een herbezinning op de filosofische methode lijkt, gezien de teloorgang van de oude monistische modellen, meer dan gerechtvaardigd, maar dat wil geenszins zeggen dat een conceptuele reconstructie van het wetenschappelijke bedrijf geen waardevolle inzichten kan opleveren, ook met betrekking tot de cognitiewetenschappen. Laat ik tot besluit van dit artikel een tweetal overwegingen noemen.

Ten eerste vormen de cognitiewetenschappen bij uitstek een domein dat, reeds vanaf het beginstadium in de jaren veertig en vijftig, gekenmerkt wordt door een vruchtbare interactie met de filosofie. Zo vond het computationele paradigma van de traditionele cognitieve psychologie zijn inspiratie in de propositionele duiding van het menselijk kenvermogen zoals die ons, vanaf het nominalisme in de middeleeuwen, via Hobbes en de logisch positivisten werd overgeleverd. Deze wisselwerking heeft belangrijke bijdragen geleverd aan de ontwikkeling van de psycholinguïstiek en de AI. De evolutionaire psychologie ontving op zijn beurt een belangrijke stimulans van de modulaire theorie van de geest zoals die door Fodor en anderen in de jaren tachtig en negentig werd uitgewerkt. Bovendien lijkt deze wisselwerking zich in de nabije toekomst voort te zetten: in het spoor van Feyerabend, Kuhn en Andy Clark (zie noot 12) tekent zich de laatste jaren een heel nieuw onderzoeksveld af dat zich richt op zogenaamde *embedded/embodied cognition:* een onderzoeksprogramma dat ons cognitief apparaat uit het cartesiaanse isolement wil halen en uitdrukkelijk als onderdeel van reeds een aanwezige, externe omgeving probeert te begrijpen. Hierbij kan men zelfs de naam Heidegger horen vallen.[28]

Het dilemma ten slotte, dat de wijsbegeerte ofwel zijn hand overspeelt door de wetenschap vanuit een leunstoel op een procrustesbed van dwingende, uitputtende modellen te leggen, ofwel verwatert tot een onsamenhangende, *case-by-case* beschrijving van een eindeloze hoeveelheid concrete voorbeelden uit de wetenschappelijke praktijk, valt of staat met de ambities van de filosoof in kwestie. Het formuleren van een reeks noodzakelijke voorwaarden die samen voldoende zijn zou na Wittgenstein ook in de filosofische deelgebieden die zich bezig houden met wetenschap en cognitie niet langer als ideaal moeten

[28]Zie bijvoorbeeld M. WHEELER, *Reconstructing the Cognitive World: The Next Step*, Cambridge, Massachusetts, MIT Press, 2007.

gelden. Wellicht is het mogelijk om door conceptuele bezinning op het wetenschappelijke handwerk tot positieve, niet-triviale inzichten te komen omtrent de verschillende onderwerpen die in dit artikel aan bod gekomen zijn, zonder daarbij de illusie te wekken dat deze inzichten op alle gevallen in gelijke mate van toepassing zijn. In dat geval is het niet de methode, maar veeleer de ambitie die onderwerp van herbezinning zou moeten zijn.

SUMMARY: *At the Intersection of Cognition, Science and Philosophy. A Philosophical Interpretation of Twentieth Century Reflection on Intertheoretic Relations.*

This article provides a critical survey of the debate on intertheoretic relations as it has manifested itself in the twentieth century. Particular emphasis is put on the cognitive sciences, as the discussion has arguably reached its most refined stage in that area. I begin by distinguishing two opposing sides, reductionism and antireductionism, and proceed by tracking the changes these positions underwent in the twentieth century. It appears that these changes consist to a significant degree of smoothing out the rough edges of both, so that the original positions can be understood as crude extremes. The monadic, all-or-nothing-accounts of intertheoretic relations were traded in for more tolerant and nuanced approaches. In the most recent developments, reductionism and anti-reductionism seem almost to converge, as theorists from both sides no longer think of their models as having an exhaustive scope.

I argue that this tendency away toward nuance is chiefly inspired by an increasing focus on actual scientific practice. This raises a question about the philosophical method. If philosophy recedes to a case-by-case description of particular examples taken from scientific practice, it is hard to see what the philosopher of science can still contribute to our understanding of intertheoretic relations.

SLEUTELWOORDEN

Reductionisme, inter-theoretische relaties, cognitie, psychologie, neurologie, fysicalisme, filosofie van de geest

KEY WORDS

Reductionism, intertheoretic relations, cognition, psychology, neurology, physicalism, philosophy of mind

PERSONALIA

Raoul Gervais
Geb. 12-05-1981
Doctoraatsstudent/promovendus wijsbegeerte
Belfortstraat 34
9000 Gent
België
Tel. 0473915589
Raoul.Gervais@UGent.be