

DISCUSSIE

TOCH GEEN REPRESENTATIES? EEN REACTIE OP J. SLEUTELS' ZONNEBRANDONTOLOGIE

F. Buekens*

De problemen die Sleutels aansnijdt in zijn replek (*ANTW* 88, 1996, nr. 2, 148-152) op mijn recensie van zijn dissertatie raken het hart van de hedendaagse filosofie van de geest en de methodologie van de cognitiewetenschappen. Ik zal me in deze korte reactie beperken tot enkele verduidelijkingen bij de objecties die ik tegen zijn relationele theorie van mentale inhoud formuleer. Hoewel: bij het lezen en herlezen van zijn boek kom ik steeds weer tot de vaststelling dat zijn relationisme best in orde is. Alleen, er is net iets te véél aan, namelijk mentale representaties. Die stelling zal ik in wat volgt verdedigen.

1. Hoe levensbelangrijk zijn representaties?

'Mentale symbolen zijn volgens Buekens nodeloze ballast, ontologische verdubbeling van de dingen. Dit punt is natuurlijk van levensbelang voor iedere cognitiewetenschap en naturalistische epistemologie, wier kernbegrip immers de mentale representaties is', stelt Sleutels. Dit is het vertrekpunt van de discussie.

Eerst een externe kritiek op het 'levensbelang' van mentale symbolen en representaties voor de cognitiewetenschappen. Het is helemaal niet zo dat het verwerpen van het bestaan van mentale representaties de ondergang der cognitiewetenschappen zou impliceren. Er worden op dit ogenblik talloze interessante modellen voor menselijke cognitie (inclusief de simulatie van cognitie) ontwikkeld die juist niet vertrekken van de assumptie dat een goed model van menselijke kennis berust op representaties en dat de cognitiewetenschappen staan of vallen met de aanname van representaties. Er zijn de theorieën over de belichaming van kennis (Haugeland 1995), over transactionele kennis (Toth 1995, met talloze verdere verwijzingen), gesitueerde kennis (Brooks 1991, met als veelzeggende titel 'Intelligence without Representation'), er is situatietheorie (Perry 1987 als voorbeeld), en op filosofisch terrein is er het werk van McDowell en Davidson die beide genuanceerde vormen van externalisme verdedigen. Wat deze theorieën gemeen hebben is hun stelling dat het een dogma is van de cognitiewetenschappen dat ze moeten vertrekken van representaties om precies dát te doen wat Sleutels wil: kennis *naturaliseren*, d.i. het opvatten als een deel van de werkelijkheid waartoe ook tafels en stoelen, electrons en zwarte gaten behoren. Het zou dus best wel eens kunnen zijn dat cognitieve wetenschappen lijden aan wat Bickhard (1993) het

* Faculteit der Wijsbegeerte van de Katholieke Universiteit Brabant.

‘substantie-syndroom’ van premature wetenschappen heeft genoemd. Bijna alle wetenschappen hebben een vroege, historische fase moeten doormaken waarin het onderzoeksobject dat geacht werd centraal te staan, een *substantie* was. Substantie-ontologieën liepen meestal faliekant af: we accepteren geen phlogiston-theorieën over vuur, calorie-theorieën over warmte, fluïdum-theorieën over magnetische verschijnselen of élan-vital-theorieën in de biologie. Al deze theorieën zijn verworpen. Voor de semantiek deed Quine overigens hetzelfde met betekenissen. Het lijkt me absoluut mogelijk dat de cognitieve psychologie ooit hetzelfde zal doen met mentale representaties. Natuurlijk bestaan er representaties: er zijn schilderijen, teksten, beelden, landkaarten, morse-codes, enz. Wat ik beweer is dat mentale representaties niet bestaan: het soort dingen dat we in een theorie over overtuigingen, kennis, verlangens, hoop en vrees nodig hebben. Het spreekt vanzelf dat ik het (net als Sleutels) in dit debat enkel heb over mentale representaties.

2. *Representaties als ‘proxies’?*

Sleutels gaat er dus van uit dat een wetenschappelijke, cognitieve psychologie slechts mogelijk is voor zover ze van de assumptie vertrekt dat kennis ‘is realized by means of internal representations: for a subject to be cognitively related to the world is for him to be tokening mental representations that go proxy for external states’ (Sleutels 1994 p. 206). Het centrale begrip is hier natuurlijk ‘proxy’. Hoe relationeel Sleutels mentale toestanden ook wil opvatten, je zou vervallen in wat hij (terecht) een omkering van het cartesianisme indien je het subject zou opvatten als een lege huls, een machine ‘op afstand bestuurd door zijn omgeving.’ (Sleutels, 2) Als dit het alternatief is voor het opgeven van representaties, dan kan ik mij daar natuurlijk ook niet in vinden. Maar omgekeerd: als er al relaties zijn met externe toestanden, moeten er dan ook nog eens ‘proxies’ zijn voor externe toestanden?

Twee kwesties lijken me hier belangrijk. Ik heb nooit willen (en zou ook nooit kunnen) beweren dat zich in ons hoofd niets zou bevinden. In ons hoofd bevindt zich een brein, en interactie met onze omgeving moet noodzakelijk neurale sporen achterlaten in die hersenen. Van zodra we datgene wat zich in onze hersenen zit, semantisch willen karakteriseren, hebben we relationele concepten nodig. We kunnen dan gebruik maken van onze gewone noties: wat X weet, verlangt of hoopt dat het geval is. Loopt er iets mis met onze hersenen, dan loopt er iets mis met die mentale huishouding. Je kan dus evengoed zeggen dat mentale toestanden identiek zijn met hersentoestanden, net zoals Xantippe identiek is met de echtgenote van Socrates, of een bepaalde toestand T van de huid identiek is met zonnebrand. De eerste karakterisering is niet-relationeel, de tweede is dat wel. Mentale toestanden (d.i. toestanden beschreven met behulp van mentale eigenschappen) zijn relationele toestanden.

Door de nadruk te leggen op mentale toestanden als *kennistoestanden* vermijdt Sleutels overigens terecht de neiging om in Cartesiaans solipsisme te vervallen. Je kan over kennistoestanden niet anders denken dan in relationele termen. Voor overtuigingen kan dat om allerlei redenen minder voor de hand liggen. Hierover beneden meer.

De neiging kan nu ontstaan om de relationele beschrijving die een inhoudelijke of semantische beschrijving is van een bepaalde gebeurtenis (of een configuratie van interne gebeurtenissen: het is best mogelijk dat mijn overtuiging dat ik nu voor mij PC zit over mijn hersenen verspreid zit) te willen vervangen door een be-

schrijving van diezelfde gebeurtenis die *niet* relationeel is, maar toch ook van semantische aard is. Dat laatste kan niet. Welnu, Sleutels' conceptie van representaties lijkt me te oscilleren tussen beide posities. Wanneer hij het heeft over representaties als dingen die noodzakelijk in relationele termen moeten beschreven worden, dan heeft hij het eigenlijk over wat ik liever iemands overtuigingen, verlangens, vermoedens of wat dan ook zou willen noemen. In dat geval is de discussie afgelopen, want spreken we verschillende talen (Sleutels verkiest 'representaties', waar ik 'mentale toestanden' verkies).

Maar als hij het heeft over 'mental representations that go proxy for external states' lijkt hij me de tweede optie te verkiezen. Als men dan ook nog blijft spreken over 'mentale symbolen' die 'beslist niet kunnen worden gemist' (Sleutels 1996, p.150) wordt de keuze voor die tweede optie alleen nog maar versterkt. Maar wat kan de functie zijn van interne symbolen? Waarom moeten de interne neurale sporen die interactie met onze omgeving achterlaat ook beschreven worden als *symbolen*? De functie van mentale symbolen (en de omzetting van externe prikkels in interne symbolen: het transductieprobleem) is dat hun inhoud ook moet toegankelijk zijn onafhankelijk van datgene waar symbolen voor staan. Is dat trouwens niet de pointe van een symbool: te staan voor iets wat afwezig kan zijn? En is de cognitief-psychologische notie van een mentale representatie niet precies hetzelfde? Lees er Tim Crane's laatste boek *The Mechanical Mind* (Crane 1995) maar op na: het staat vol van dit soort retoriek.

In de volgende passage lijkt Sleutels me de nagel op de kop te tikken: 'elke kenrelatie (moet) tevens een relatie tot een specifiek gestructureerd subject zijn. Zo'n structuur (die ook niet-relationeel gespecificeerd kan worden, maar dan niet als representatie geldt) is een noodzakelijke voorwaarde voor kennis; *voldoende* voorwaarde wordt hij pas wanneer hij wordt begrepen *als* representaties, d.w.z. in relationele termen.' (p. 4) Ik denk dat dit absoluut waar is, maar ook waar blijft als je 'representatie' overal in dit citaat vervangt door 'mentale toestand'. Het probleem is nu dat, als deze substitutie correct is (en er is geen enkele reden om daaraan te twijfelen), Sleutels eigenlijk niets anders zegt (en niets anders hoeft te zeggen) dan dat mentale toestanden relationeel moeten worden beschreven. Daar waren we mee begonnen en waren we het over eens. Representaties lijken ondertussen wél uit de boot te vallen. Ik blijf erbij dat Sleutels' relationisme er beter bij gebaat is, representaties uit de boot te laten vallen. Ik vind het overigens merkwaardig dat in dit citaat toch nog sprake is van de kenrelatie als een relatie tot een representatie. Dan verkies ik liever wat Sleutels enkele regels ervoor schrijft: representaties (of mentale toestanden, als je mijn optie volgt) zijn datgene waarin we ons verhouden wanneer we ons kennend verhouden tot iets.

De moeilijkheid is dat men vaak twee verschillende noties van relationaliteit door elkaar gebruikt: representatie-opgevat-als-mentaal-symbool maakt gebruik van relaties van mentale symbolen tot een virtuele wereld die, wanneer die mentale symbolen kennistoestanden karakteriseren, de actuele wereld is – de wereld die toelaat de interne representaties te kwalificeren als kennisdragers. De representationalist kan de wereld waartoe de representatie zich verhoudt, wijzigen in functie van de attitude die de representatie moet kenmerken: een verlangen verhoudt zich tot een virtuele wereld die, wanneer het verlangen gerealiseerd is, samenvalt met de actuele wereld. De relationaliteit waar Sleutels het in het tweede citaat over heeft (representaties als voldoende voorwaarde voor kennis) is relationaliteit die

enkel in actualistische termen gekarakteriseerd kan worden. Als we zeggen dat zonnebrand een relationeel concept is, houdt dit niet in dat er een mogelijke wereld is waarin die toestand van de huid door de zon is veroorzaakt. Nee: hij moet in de *actuele wereld* door de zon veroorzaakt zijn, zoniet is het geen zonnebrand maar een toestand van de huid die lijkt op zonnebrand. Natuurlijk kan je contrafactisch redeneren over de relatie tussen de zon en je huid; maar als je zegt dat je zonnebrand hebt, moet je *de facto* ook verbrand zijn door de zon.

Sleutels maakt een allusie op beide noties van relationaliteit – relationaliteit gebaseerd op actuele, feitelijke causale verbindingen en de lossere (tegenfeitelijke) ‘causal powers’. Uiteraard heb ik het over actuele causale verbindingen, niet over tegenfeitelijke causal powers. De reden hiervoor ligt in de lijn van Sleutels’ pleidooi voor kennis als centrale cognitieve categorie (zie ook Buekens 1994 voor een verdere uitwerking van dit causaal actualisme). Uit het feit dat iemand weet dat *p*, kun je iets afleiden over de wereld, (nl. dat *p* het geval is). En omgekeerd: wanneer *p* het geval is, zal een redelijk persoon, indien zij zich de moeite getroost, zichzelf in een toestand kunnen brengen waarin zij weet dat *p*. Dat doet ze door het juiste causale contact met externe objecten en gebeurtenissen te organiseren. Weten dat *p* wordt beter niet opgevat als een specifiek geval van het bezitten van informatie, nl. informatie die gaat over de actuele wereld. (In dat geval geldt dat ‘geloven dat *p*’ zijn inhoud zou ontleen aan een mogelijke, niet noodzakelijk geactualiseerde wereld.) Het alternatief is dat weten dat *p* opgevat worden als een mentale configuratie waarin behalve een mentale toestand ook de omgeving van die persoon is betrokken. Verandert die omgeving, dan verandert niet de waarheidswaarde of het epistemisch statuut van die toestand: de toestand zelf verdwijnt. Ik geef overigens grif toe dat dit nadere uitleg behoeft (zie andermaal Buekens 1994).

Ik besluit met een bedenking die ik elders hoop uit te werken. Een van de redenen waarom Sleutels gelijk heeft als hij stelt dat kennis centraal moet staan in een naturalistische, niet-reductionistische theorie over cognitie, is dat ‘weten dat *p*’ als cognitieve attitude van factieve aard is. (het heeft dit gemeen met ‘zich herinneren dat *p*’, ‘vergeten dat *p*’, etc.) Deze mentale toestanden zijn intrinsiek relationeel: uit wat je zegt over een persoon (‘*X* weet dat hij zich in Argentinië bevindt’) volgt tevens informatie over wat het geval is (de complementzin is immers waar). En ook omgekeerd: wanneer iemand zich in Argentinië bevindt, zal deze persoon, indien hij rationeel is, vroeg of laat erachter komen dat hij zich in Argentinië bevindt (*ceteris paribus*). Redelijke wezens laten toe dat je kan redeneren van wat iemand denkt naar wat het geval is, en van wat het geval is naar wat iemand hoort te denken. Dit, lijkt me, is de kern van de relationaliteit van mentale toestanden. Wanneer we daarentegen hebben met niet-factieve constructies als ‘*X* gelooft dat *p*’, dan lijken we van die relationaliteit meteen iets magisch te moeten maken: ‘*X* gelooft dat *p*’ wordt immers (onder druk van een bepaalde epistemologie?) niet alleen opgevat als een cognitieve attitude, maar ook als een epistemische attitude: *X* gelooft iets dat waar of onwaar kan zijn. Het kan zijn inhoud niet ontleen aan wat het geval is als *X* gelooft dat *p* (wat gespecificeerd wordt in de waarheidsvoorwaarde voor ‘*p*’), maar aan iets wat aanwezig is ongeacht de waarheid van ‘*p*’. Ik denk dat veel discussies in de ‘philosophy of mind’, en met name een aantal posities die neigen naar methodologisch solipsisme, zich te sterk laten inspireren door epistemologische overwegingen: er moet nog sprake kunnen zijn van (smalle?) inhoud, zelfs indien men de externe context (of, zo men wil, de

causale geschiedenis van een mentale toestand buiten het lichaam) weglaat. Want wie garandeert ons dat onze overtuigingen werkelijk waar zijn en niet gaan over de wijze waarop de wereld (of beter: een wereld) ons voorkomt? McDowell heeft het in deze context overigens vaak over de ‘desastreuze’ gevolgen van het klassieke ‘argument from illusion’ in de epistemologie en de filosofie van de geest.

Literatuur

- Bickhard, M., ‘Representational Content in Humans and Machines’, in *Journal for Experimental and Theoretical Artificial Intelligence* 5 (1993), 285-333
- Brooks, R., ‘Intelligence without Representation’, *Artificial Intelligence* 47 (1991), 139-160
- Buekens, F., ‘Externalism, Content, and Causal Histories’, *Dialectica* 48 (1994), 267-287
- Crane, T., *The Mechanical Mind. A Philosophical Introduction to Minds, Machines and Mental Representation* London 1995
- Haugeland, J., ‘Mind Embodied and Embedded’, in L. Haaparanta e.a. (eds), *Mind and Cognition. Philosophical Perspectives on Cognitive Science and Artificial Intelligence*, Helsinki 1995 (Acta Philosophica Fennica: vol. 58)
- Perry, J., ‘Thought Without Representation’, *Proc. Arist. Soc.* 1987, 136-151
- Sleutels, J., *The Problem of Content in Neural Epistemics*, Nijmegen 1994
- Toth, P., bespreking van K. Ford en P. Hayes, eds., *Reasoning Agents in a Dynamic World: the Frame Problem*, in *Artificial Intelligence* 73 (1995), 323-369