

REDUCTIE OF ELIMINATIE VAN QUALIA?

Superzwakke, zwakke, sterke en supersterke qualia
bij Paul Churchland

door

J.J.M. Sleutels

1. Diskwalificatie van qualia?

Het door Paul Churchland verdedigde idee van een *vervanging* van onze vertrouwde psychologie door een volwassener *neurowetenschap* is behept met een zekere dubbelzinnigheid: wordt de oude ontologie van mentale entiteiten en processen *geëlimineerd* door de nieuwe theorie, of is er veeleer sprake van *reductie*? Of-schoon Churchland vooral bekend is geworden als een voorstander van de eerste interpretatie, aangeduid als 'eliminatief materialisme', zijn beide tendenties in zijn werk aantoonbaar. Lezers staan voor de vraag hoe deze aspecten zich tot elkaar verhouden.¹

In dit artikel wil ik mij beperken tot een analyse van Churchlands behandeling van de zogeheten *qualia*, de onmiddellijk gegeven kwalitatieve inhoud van het fenomenaal bewustzijn. Qualia worden veelal beschouwd als moeilijk inpasbaar in het fysicalistisch mensbeeld. Kleuren, geuren, smaken en andere bewoners van ons bewustzijn² bezitten een subjectieve, onmiddellijk ervaren *hoedanigheid*; hun 'van binnenuit aangevoelde' identiteit zou zich onttrekken aan elke uitwendige derdepersoonsbenadering.³ Dennett heeft in tal van publicaties laten zien dat de identiteit

¹ Cf. Sleutels 1988.

² Boeiende qualia beschrijft de Amerikaanse dichteres Diane Ackerman in *A natural history of the senses* (Ackerman 1990).

³ Voor enkele bekende, op qualia gebaseerde objecties tegen fysicalistische theorieën van het mentale, zie o.m. Kripke 1972, Nagel 1974, Block 1978, Searle 1980, Jackson 1982, Robinson 1982. Een bonte verzameling gedachtenexperimenten pro en contra het fysicalisme is Dennett & Hofstadter 1981. Als de 'moeder van alle argumenten' kan het fameuze argument van Kripke worden beschouwd. Een korte, *nuwe* versie ervan iuudt als volgt. Kwalitatieve bewustzijnstoestanden, bijv. pijn, worden geregeerd door identiteitscriteria die volledig verschillen van de criteria in termen waarvan neurale toestanden worden gespecificeerd. Daarom kan men zich gemakkelijk voorstellen dat een bepaald type neurale toestand N zich voordoet zonder dat N gepaard gaat met een bepaald type bewustzijnstoestand B. In filosofenjargon: als N en B al identiek zijn, dan is dit hooguit een *contingente* identiteit. Als zodanig is deze identiteit wetenschappelijk niet interessant, aangezien wetenschappelijke identiteiten, bijv. die van water en H₂O, steeds *noodzakelijk* zijn: het is onvoorstelbaar dat iets H₂O is maar geen water. Op de basisstructuur van Kripke's argument zijn vele gedachtenexperimenten geënt, waarin telkens de fenomenale identiteit van bewustzijnstoestanden constant wordt gehouden, terwijl de interne (neurale, functionalistische) of externe (ecologische, historische, sociale, linguïstische) omstandigheden variëren. Voor het regi-

Jan Sleutels (1960) studeerde filosofie en rechten aan de KU Nijmegen, waar hij assistent in opleiding was bij de vakgroep Wetenschapsfilosofie. Sedert 1989 is hij universitair docent bij de vakgroep Kennisleer en Metafysica van de RU Leiden. Adres: Faculteit der Wijsbegeerte, RU Leiden, Postbus 9515, 2300 RA Leiden

van qualia ronduit vaag is; aan onze intuïties omtrent de inhoud van qualia liggen elkaar tegensprekende criteria ten grondslag, die in kundig gekozen gedachtenexperimenten zichtbaar kunnen worden gemaakt.⁴ Ingevolge Quines fameuze adagium 'no entity without identity' zouden qualia wellicht beter kunnen worden geschrapt uit de ontologie van het mentale. Anderzijds valt moeilijk te ontkennen dat gewaarwordingen bestaan: zijn wij ons niet op elk ogenblik van hun bestaan bewust? Een dilemma dient zich aan: de fysicist zal geneigd zijn qualia te elimineren omdat zij niet passen bij zijn theorie, terwijl zijn tegenstander zal concluderen dat het fysicalisme niet deugt omdat het geen raad weet met qualia.

Churchlands naturalistische filosofie biedt een op het eerste gezicht aannemelijke verklaring van kwalitatieve mentale inhouden. Centraal in zijn theorie staat het idee dat ons zenuwstelsel fungeert als een meetinstrument, een detector van de eigenschappen van distale objecten.⁵ Zijn oplossing voor het probleem van qualia is op het eerste gezicht de eenvoud zelve: de detectietoestanden *zijn zelfde* introspecteerbare qualia van het fenomenaal bewustzijn. Waarneming is de theoretische exploitatie van de informatie die deze introspecteerbare detectietoestanden bevatten.

Perception consists in the conceptual exploitation of the natural information contained in our *sensations or sensory states*.⁶

Gewaarwordingen ('sensations') worden door Churchland gelijkgesteld aan detectietoestanden ('sensory states'). Achter deze ogenschijnlijk onschuldige stap gaat echter een *Hinterland* van filosofische problemen schuil. Tussen beide bestaan traditioneel belangrijke verschillen. Gewaarwordingen worden primair gekenmerkt door een eigen *epistemische* of *semantische* identiteit. Ook Churchland zelf blijft spreken over het specifieke, kwalitatieve karakter van sensaties: hun inhoud is subjectief introspecteerbaar door de drager ervan. Waarnemingstoestanden, daarentegen, bezitten primair een *causale* (en *meer* in het bijzonder: *neurofysiologische*) identiteit. Zij worden gespecificeerd in termen van hun causale rol in het perceptieproces; als zodanig kunnen zij natuurlijk worden geïdentificeerd met specifieke toestanden van het perceptieapparaat. In het jargon dat in hedendaagse discussies over 'mental content' wordt gehanteerd: spreekt Churchland over gewaarwordingen, dan gebruikt hij een semantische taxonomie van qualia; spreekt hij over detectietoestanden, dan gebruikt hij een syntactische (of formele) taxonomie.⁷ De vraag is nu of Churchlands *identificatie* van deze taxonomieën niet veel-

me van filosofisch surrealisme waartoe deze gedachtenexperimenten aanleiding hebben gegeven, zie ook noot 24 en de verwijzingen aldaar.

⁴ Zie bijv. Dennett 1990 en 1991.

⁵ Deze causale perceptietheorie kampt met [al van randproblemen, waaronder dat van 'misrepresentatie'. Dergelijke problemen zijn evenwel niet uniek voor Churchlands causale theorie; m.n. ook de causale theorieën van zijn critici, waaronder die van Fodor, staan voor hetzelfde probleem. Elders heb ik beargumenteerd dat een naturalistische oplossing op basis van Churchlands materialisme superieur is aan de oplossing die o.a. Fodor zelf biedt voor het probleem van misrepresentatie (Sleutels 1991).

* Churchland 1979, p. 7 (curs. JS).

⁷ Voor een verhelderende bespreking van het afwijkende gebruik van de begrippen 'semantisch', 'syntactisch' en 'formeel' in dit verband, zie Devitt 1990.

eer een *verwarring* is.

In Churchlands benadering van qualia schuilt nog een tweede, verwante dubbelzinnigheid. De ene keer lijkt hij naar een *eliminatieve* positie te neigen, dan weer dringt hij aan op een *reductie* van qualia. Vooral in zijn eerdere werk zweemt zijn theorie naar eliminativisme. Van gewaarwordingen ('sensations'), zo heet het daar, is de intrinsieke kwalitatieve identiteit volmaakt irrelevant; wat telt zijn enkel de systematische relaties die zij onderhouden met hun oorzaak onder normale omstandigheden en het oordeel waartoe zij een subject normaliter disponeren. *Qualia* kunnen gewaarwordingen gewoon worden gemist.

The intrinsic qualitative identity of one's sensations is irrelevant to what properties one can or does perceive the world as displaying. (...) Sensations are just *causa!* middle-men in the process of perception, and one kind will serve as well as another so long as it enjoys the right causal connections. (...) In principle they might even be *dispensed* with, so far as the business of learning and theorizing about the world is concerned. As long as there remain systematic causal connections between kinds of states of affairs and kinds of singular judgements, the evaluation of theories can continue to take place.⁸

Een voor de hand liggende suggestie zou zijn om Churchlands dubbezzinnige positie als volgt te begrijpen: *gewaarwordingen* kunnen worden geëlimineerd, terwijl de causaal relevante *detectietoestanden* worden gereduceerd. Toch is dit duidelijk niet Churchlands bedoeling, getuige de volgende karakteristieke passages waarin precies de mogelijkheid van een reductie van kwalitatieve gewaarwordingen centraal staat.

"The *objective* qualia (redness, warmth, etc.) should never have been 'kicked inwards to the minds of observers' in the first place. They should be confronted squarely, and they should be reduced where they stand: *outside* the human observer. (...) If objective phenomenal properties are so treated, then subjective qualia can be confronted with parallel *fortrightness*, and can be reduced where they stand: inside the human observer".⁹

"The 'ineffable' pink of one's current visual sensation may be richly and precisely expressible as a 95Hz/80Hz/80Hz 'chord' in the relevant triune cortical system. The 'unconveyable' taste sensation produced by the fabled Australian health tonic Vegamite might be quite poignantly conveyed as a 85/80/90/15 'chord' in one's four-channeled gustatory system (a dark corner of taste-space that is best avoided). And the 'indescribable' olfactory sensation produced by a newly opened rose might be quite accurately described as a 95/35/10/80/60/55 'chord' in some six-dimensional system within one's olfactory bulb".¹⁰

⁸ Churchland 1979, p. 15.

⁹ Churchland 1985 (1989, p. 57).

¹⁰ Churchland 1986 (1989, p. 106).

In dit artikel wil ik een poging wagen de oorzaken van deze ambiguïteit in Churchlands opvatting over qualia te traceren en zijn theorie zo mogelijk te zuiveren van dubbelzinnigheden. Na een schets van de basisbeginselen van Churchlands epistemologie introduceer ik een onderscheid tussen vier soorten qualia: supersterke, superzwakke, zwakke en sterke qualia, corresponderend met resp. het klassieke empirisme, eliminativisme en de beginselen van zwakke en sterke plasticiteit van perceptie. In de daaropvolgende paragrafen ga ik de status van zwakke en sterke qualia in Churchlands werk na. Centraal daarbij staat de vraag of deze qualia in werkelijkheid supersterke qualia zijn, zoals critici beweren.

2. Sterke, zwakke, supersterke en superzwakke qualia.

Churchlands opvattingen over qualia hangen nauw samen met een aantal kernpunten in zijn filosofie. Voor wij het eigenlijke qualia-probleem onderhanden nemen, geven wij eerst een korte schets van de drie basisbeginselen die in dit verband het meest relevant zijn: de netwerktheorie van betekenis, het empirisch realisme en de plasticiteit van perceptie.¹¹

Volgens Churchlands netwerktheorie van betekenis is de betekenis van elke uitspraak een functie van het geheel van samenhangende overtuigingen waarin zij functioneert. Dit geldt niet alleen voor theoretische uitspraken maar ook voor waarnemingsuitspraken: hierin onderscheidt deze *radicale* netwerktheorie zich van een zwakkere variant die vasthoudt aan de relatieve semantische en epistemische autonomie van waarnemingsuitspraken, zoals verdedigd door o.a. Quine.¹²

Anders dan de relatieve variant lijkt de radicale netwerktheorie onverenigbaar met een tweede basisbeginsel van Churchlands filosofie, zijn empirisch realisme. Deze term gebruik ik hier voor Churchlands combinatie van wetenschappelijk realisme en empirisme. De kern ervan luidt dat waarnemingsuitspraken direct verslag doen van de waarneembare eigenschappen van distale objecten." Dit beginsel verzet zich tegen een vorm van empirisme waarin mentale representaties (bijv. gewaarwordingen) fungeren als het onmiddellijk gekende, het *primum notum*. Empirisch realisme is op het eerste gezicht onverenigbaar met de netwerktheorie van betekenis: is de betekenis van waarnemingsuitspraken een functie van de theorie die de waarnemer hanteert, dan wordt zij niet bepaald door het object *zelf*, maar door onze theorie *over* het object. Om de afstand tussen deze posities te overbruggen dient een boud, derde beginsel, dat van de plasticiteit van perceptie.

Volgens Churchland veroorzaken de eigenschappen van distale objecten bepaalde neurale representaties die hier 'detectietoestanden' worden genoemd. Deze worden op hun beurt conceptueel geëxploiteerd, d.w.z. verwerkt in andere representa-

¹¹ De hier gegeven presentatie volgt enkele hoofdlijnen uit Philippe's genealogische reconstructie van de sterke plasticiteit van perceptie (1990, pp. 129-146). Philippe geeft een zeer erudiete maar vaak verwarrend complexe schets van Churchlands positie.

¹² Cf. Quine 1990, pp. 2vv: "some semences (...) are directly and firmly associated with our stimulations. (...) I call them observation sentences. (...) Observation sentences are the link between language, scientific or not, and the real world that language is all about."

¹³ Dit 'empirisch realisme' is een verbijzondering van Churchlands wetenschappelijk realisme in het algemeen: (theoretische zowel als observationele) uitspraken doen verslag van de eigenschappen van distale entiteiten. Wij beperken ons hier tot de empirische stellingname vanwege het gekozen onderwerp, qualia.

ties, om uiteindelijk te resulteren in gedrag, onder meer in waarnemingsuitspraken. Het beginsel van de plasticiteit van perceptie stelt nu dat wij de werkelijkheid anders gaan waarnemen wanneer wij een nieuwe theorie verwerven. Zo gesteld is dit beginsel echter dubbelzinnig, zoals Philipse terecht opmerkt.¹⁴ Want wat verandert er in de zich wijzigende waarneming: onze waarnemingsuitspraken, of ook onze *gewaarwordingen*? Een zwakke vorm van plasticiteit laat de identiteit van onze waarnemingen of gewaarwordingen intact terwijl de geassocieerde waarnemingsuitspraken (d.w.z. onze spontane, non-inferentiële uitspraken over distale factoren) veranderen als functie van onze theorieën. Een sterke interpretatie van het plasticiteitsbeginsel laat de waarnemingen of gewaarwordingen zelf meeveranderen met de waarnemingsuitspraken en theorieën; wij nemen dan de wereld *direct, kwalitatiefanders* waar in termen van een nieuw conceptueel apparaat.

Beide lezingen van plasticiteit komen in Churchlands werk voor. In de loop der jaren heeft zijn positie zich geradicaliseerd. In *Scientific realism and the plasticity of mind* (1979) ging hij uit van de starheid van onze gewaarwordingen als onderscheiden van de plasticiteit van onze waarnemingsoordelen ('conceptual' of 'judgmental plasticity', zwakke plasticiteit), terwijl in latere publicaties wordt verdedigd dat ook onze *gewaarwordingen* plastisch zijn ('sensational plasticity', sterke plasticiteit).¹⁵

Om Churchlands positie op dit punt beter te kunnen plaatsen is het nuttig haar te vergelijken met een klassieke perceptietheorie als die van Locke, zoals bekriftiseerd door o.a. Kant en Sellars.¹⁶ Volgens Lockes middellijk empirisme is niet het distale object maar het quäle (de gewaarwording of zintuiglijke indruk) onmiddellijk gegeven in de waarneming. De waarnemer is zich primair bewust van zijn eigen inwendige toestand, bijv. van een 'rood'-indruk. *Middels* deze kwalitatieve indrukken (re-)construeert hij vervolgens een theorie over de eigenschappen van distale objecten. Voor het gemak van verwijzing noem ik de qualia in Lockes theorie 'supersterk':

- (1) *Supersterke qualia* =_{def} fenomenale bewustzijnstoestanden die de betekenis van waarnemingsuitspraken bepalen.

Supersterke qualia zijn epistemisch actieve 'gegevens*' in de door Seliars bekritiseerde zin van het woord.¹⁷ Volgens de 'Myth of the given' wordt ons waarne-

¹⁴ Philipse 1990, m.n. pp. 132-133.

¹⁵ Zoals gesteld in Churchlands commentaar op Fodor (1988): "My own 1979 position (...) simply assumes the generally constant character of our sensory responses to the environment. The plasticity that excited me there was confined to the conceptual frameworks within which we make our judgmental responses to the passing contents of our sensory manifold. (...) To be sure, sensational plasticity would constitute an *additional* argument for the plasticity of perception. (...) And I (...) am now willing to defend it *vigorously*" (Churchland 1988b (1989, pp. 277-278)).

¹⁶ Het doet hier niet ter zake of het navolgende een correcte exegese van Lockes positie is. Voor een lezing van Locke als 'direct' realist, zie o.m. Yolton 1987 en Mandelbaum 1964.

¹⁷ Cf. Sellars' 'Empiricism and the philosophy of mind'. 1963, pp. 127-196. In dezelfde geest observeerde Kant n.a.v. Lockes empirisme, "man glaubte, die Sinne lieferten uns nicht allein Eindrücke, sondern setzten solche auch so gar zusammen" (1781, A 120, Anm.).

mingsoordeel 'X is A' zowel veroorzaakt als gerechtvaardigd door een 'A'-indruk. Supersterke qualia bezitten een eigen semantische identiteit, een eigen 'A'-heid; de vorming van oordelen is enkel een kwestie van de passende begrippen zoeken.

Sellars distantieerde zich scherp van Lockes empirisme, dat volgens hem op een 'naturalistic fallacy' zou berusten. Locke verwarde de oorzaak van het oordeel met zijn epistemische grond. Kennis is een zaak van *uitspraken*, terwijl gewaarwordingen *zelf* nu eenmaal geen uitspraken zijn. Gewaarwordingen hebben geen eigen (prelinguïstische) 'conceptuele inhoud' en kunnen dus niet epistemisch actief zijn.

Wordt het bestaan van supersterke qualia ontkend, dan doet zich het probleem voor wat gewaarwordingen *überhaupt* nog met waarneming te maken hebben. Als het quale niet bepaalt *wat* wordt waargenomen kan het ook niet bepalen *dat* wordt waargenomen; voor waarneming is het niet constitutief. Ontkenning van epistemisch actieve qualia leidt aldus tot een positie waarin qualia *superzwak* worden:

- (2) *Superzwakke qualia* =_{def} fenomenale bewustzijnstoestanden die *niet* (mede-) constitutief zijn voor waarneming.

Churchland onderschrijft Sellars' kritiek op supersterke qualia, getuige zijn opmerking dat "sensations themselves are not yet truth-valuable or semantically-contentful states": zij behoren tot "the wrong logical space: it is only an observation judgment, or belief, or report that can be logically consistent or inconsistent with any theory".¹⁸ Toch betekent dit niet zonder meer dat zijn qualia *dus* superzwak zijn. Zoals boven al werd opgemerkt is Churchlands positie dubbelzinnig: soms plaatst hij de redundantie van qualia voorop ("they might even be dispensed with"), maar op andere plaatsen overheerst het idee van reductie. Bovendien gebruikt Churchland twee verschillende versies van het beginsel van plasticiteit van perceptie, beide van invloed op zijn positie t.a.v. qualia. Laten wij ter aanduiding van Churchlands qualia twee nieuwe begrippen introduceren: *zwakke* en *sterke* qualia, geregeerd door de beginselen van zwakke resp. sterke plasticiteit.

- (3) *Zwakke qualia* =_{def} fenomenale bewustzijnstoestanden waarvan de kwalitatieve identiteit onafhankelijk is van de conceptuele exploitatie.
(4) *Sterke qualia* =_{def} fenomenale bewustzijnstoestanden waarvan de kwalitatieve identiteit een functie is van de conceptuele exploitatie.

Supersterke qualia zijn duidelijk onverenigbaar met de absolute netwerktheorie van betekenis. De plasticiteitsstelling, daarentegen, wordt door Churchland ingeroepen om zijn netwerktheorie te verzoenen met het empirisch realisme. Deze poging wordt echter niet door iedereen als geslaagd ervaren. In enkele recente publicaties is door critici beargumenteerd dat Churchlands uitgangspunten onverenigbaar zijn. In het bijzonder zou het plasticiteitsbeginsel er onvoldoende in slagen

¹⁸ Churchlands dupliek aan Fodor (1988) n.a.v. diens commentaar op de plasticiteit van perceptie (Churchland 1988b (1989, pp. 267-268 en 277)). Cf. Sellars' oorspronkelijke formulering, dat "in characterizing an episode or state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in *the logical space of reasons*, of justifying and being able to justify what one says" (Sellars 1963, p. 169, curs. JS).

zich los te maken van het klassieke empirisme à la Locke: zowel zwakke qualia als sterke qualia zouden in werkelijkheid verkapte supersterke qualia zijn. In de volgende paragrafen onderwerpen wij deze kritiek aan een nauwkeurig onderzoek, eerst de zwakke qualia, daarna de sterke.

3. *Wal zeggen zwakke qualia?*

Volgens Churchland zijn zelfs waarnemingstermen als 'rood' en 'warm' geen namen voor gewaarwordingen, maar verwijzen zij naar de objectieve, distale eigenschappen waardoor de desbetreffende gewaarwordingen worden teweeggebracht. Hij beargumenteert dit aan de hand van een voorbeeld waarin verschillende causale routes leiden naar 'dezelfde waarneming'.¹⁹ Warmte kan met de tastzin worden waargenomen maar ook met behulp van infrarood-detectoren, zoals bij sommige diersoorten (waaronder de ratelslang) het geval is. Men kan zich gemakkelijk een ras van mensachtige wezens voorstellen, zo redeneert Churchland, wier ogen functioneren als detector van infrarode straling. Gesteld dat deze wezens sterk op ons lijken, dan kunnen wij ons verder voorstellen dat zij een warm voorwerp zien zoals wij een wit voorwerp zien, terwijl de waarneming van een koud voorwerp voor hen lijkt op onze waarneming van iets zwarts. Hoe moeten wij nu de waarnemingsoordelen en introspectierapporten van dergelijke wezens verstaan? Ais beschrijvingen van kleuren en kleurgewaarwordingen, of als beschrijvingen van temperatuur en warmtesensatie? Churchland pleit voor de tweede interpretatie.

On viewing a very hot object they have what we would describe as a sensation of an incandescent *while* object, and on viewing a very cold object they have what we would describe as a sensation of a *black* object, and so on. They, of course, describe these sensations quite differently - as sensations of heat, of coldness, and so on.²⁰

Objectief gezien (qua distale oorzaak) zijn de gevoelde en geziene warmte identiek, maar subjectief gezien (qua waarneming of gewaarwording) verschillen zij. Intersubjectief gezien (qua taalgedrag) zijn volgens Churchland de oordelen waartoe de verschillende waarnemingen aanleiding geven wederom identiek, nl. op grond van de voorgestelde vertaling.

Bovenstaand argument introduceert het beginsel van *zwakke plasticiteit*. Zoals in de vorige paragraaf werd geëist, slaat het een brug tussen Churchlands empirisch realisme en zijn netwerksemantiek. De betekenis van waarnemingsuitspraken (bijv. 'Dit is warm', zoals gezegd door een infrarood-waarnemer) wordt weliswaar direct bepaald door de waargenomen distale eigenschap, maar blijkt tevens een functie van de achterliggende theorie, in dit geval die over de aard van de waargenomen eigenschap, een theorie die wij met de infraroodwaarnemers delen. Het belangrijke punt is hier dat zwakke qualia volmaakt onbelangrijk zijn in het kenproces: zij bepalen noch wat wordt waargenomen, noch hoe het wordt beschreven.

In een recente kritiek op Churchlands positie wordt door Philipse een tegen-

¹⁹ Churchland 1979, pp. 7-14; Philipse 1990, pp. 139-142.

²⁰ Churchland 1979, p. 9.

gestelde conclusie bepleit. Hij wijst erop dat in bovenstaand argument het *zien* en het *voelen* van een warm voorwerp kwalitatief verschillende waarnemingen opleveren, ofschoon dit kwalitatieve verschil niet tot uitdrukking komt in een verschil in betekenis van de bijbehorende waarnemingsoordelen. Het infrarood-oordeel dat gepaard gaat met een 'wit-achtige' gewaarwording (bij waarneming van een warm voorwerp) vertalen wij als een oordeel over warmte, niet over kleur. Welnu, "if the way the world appears to the hominoids were not determined by their sensations", aldus Philipse, "it would not make sense to consider the possibility of a sensation-guided translation, albeit as a possibility which is eliminated (...). In other words, it is essential to Churchland's argument that his biological fiction presupposes the very variety of the representative theory of perception which is rejected once the argument has done its service".²¹ Blijkens de context is de 'variety of the representative theory of perception' die Churchland zou vooronderstellen het Lockeans empirisme.

Op het eerste gezicht doet deze objectie bizar aan. Als je aantoont dat de kwalitatieve inhoud van gewaarwordingen *irrelevant* is voor je perceptietheorie, zoals Churchland doet, volgt toch zeker niet dat je perceptietheorie (derhalve!) dergelijke inhouden *essentieel vooronderstelt*, zoals de kritiek concludeert? Merk in dit verband op dat Churchlands argument, ontologisch gezien, zelfs een epifenomenalisme t.a.v. zwakke qualia impliceert, of althans niet uitsluit. De subjectief waargenomen kwalitatieve inhoud is causaal volstrekt inert. Zelfs *zonder* qualia zouden de systematische connecties tussen omgeving, interne verwerking en (taal)gedrag *ex hypothesi* volledig intact blijven. Dergelijke inerte qualia kunnen uiteraard geen epistemologische dienst doen; zij bepalen noch wat wordt waargenomen, noch wat wij erover zeggen. Zij zijn kentheoretisch overbodig.²²

Toch kan de objectie niet zo gemakkelijk worden afgedaan. De inertie van zwakke qualia is dermate compleet dat zij *superzwak* lijken te zijn. Dit is vermoedelijk het achterliggende motief van Philipse's enthymem: Churchlands zwakke qualia moeten *ofwelsuperzwak* zijn *ofwelsupersterk*. Zijn zij *superzwak*, dan zijn zij niet constitutief voor *waarneming*. In dat geval mist Churchlands theorie een belangrijk element van waarneming, als onderscheiden van de pure neurofysiologie van hersenprocessen. Zijn zwakke qualia daarentegen *supersterk*, dan is er geen sprake van plasticiteit. Het is dan immers de inhoud van gewaarwordingen die bepaalt wat wij waarnemen. Churchlands notie van 'zwakke qualia' is een hybride: zij leent van *superzwakke* qualia hun inertie, van *supersterke* qualia hun verband met waarneming. Dit zou de reden kunnen zijn waarom Churchland, zoals eerder opgemerkt, in zijn perceptieeler steevast de weifelende uitdrukking '*gewaarwordingen of waarnemingstoestanden*'* gebruikt: hij wil tegelijk de kool en de geit sparen.²³

²¹ Philipse 1990, p. 141.

²² Churchland 1979, p. 15: "In principle they might even be *dispensed* with, so far as the business of learning and theorizing about the world is concerned. As long as there remain systematic causal connections between kinds of states of affairs and kinds of singular judgements, the evaluation of theories can continue to take place".

²³ Cf. Philipse 1990, pp. 157-158: "As sensations do not determine the way the world appears to us, they have no other function than sensory or neural states have, and may either be '*dispensed*'

In haar nieuwe gedaante is de objectie intuïtief plausibel. Misschien is zij zelfs een van de redenen waarom Churchland zich in zijn latere werk steeds reductionistisch is gaan opstellen t.a.v. qualia en is overgegaan op sterke qualia. Het probleem is echter dat de objectie erg moeilijk *precies* te maken is. Onduidelijk is met name de wijze waarop qualia *constitutief* zijn voor perceptuele processen. De objectie vooronderstelt dat alleen *supersterke* qualia waarneming kunnen constitueren. Deze premisse heeft echter ongerijmde gevolgen, zoals blijkt wanneer wij haar toepassen op Churchlands argument.²⁴ Wat de infrarood-waarnemers in wezen doen, zo luidt de nieuwe tekst, is hun intrinsieke 'wit'-inhouden (die zij met ons delen) *interpreteren* *ah* tekens van warme distale objecten. De onmiddellijke waarneming van een 'wit'-qualie brengt sommige wezens ertoe te concluderen dat er zich een wit voorwerp voor hen bevindt, terwijl andere wezens op grond van dezelfde waarneming concluderen dat er een *warm* voorwerp aanwezig is. Het supersterke qualie geeft in beide gevallen dezelfde waarneming, die echter verschillend wordt geïnterpreteerd. Het element van waarnemingsconstitutie schuilt kennelijk in het introspectief bewustzijn van de 'naakte', kwalitatieve 'gegevens' voorafgaand aan het oordeel.

Tenminste drie fundamentele bezwaren kunnen tegen deze opvatting worden aangevoerd. In de eerste plaats vooronderstelt zij een homunculus die de qualia waarneemt, interpreteert en omzet in gedrag; daarmee is zij als verklaring van waarneming circulair. Bovendien is de 'gegevens' inhoud van de qualia ofwel arbitrair (waarom zouden wij hem 'wit' noemen en niet 'warm?'), ofwel oneindig disjunctief ('warm of wit of ...'). Dit holt de notie van 'inhoud' zodanig uit dat supersterke qualia niet meer epistemisch relevant zijn. Hoe kunnen zij dan dienen ter identificatie van 'echte' waarneming? Ten derde strookt bovenstaande theorie niet met de fenomenologie van de waarneming: aan mijn (exterspectief) bewustzijn van witte voorwerpen gaat geen (introspectief) bewustzijn van naakte qualia vooraf. Eerder het omgekeerde is het geval: reflecterend over mijn uitwendige

with' (zwakke plasticiteit, *JS*) or be identified with these states (sterke plasticiteit, *JS*). But in spite of this repression of the representative theory of perception by the strong plasticity thesis (evenals de zwakke, *JS*), the latter owes its appearance of intelligibility completely to the former, which survives, as it were, in the dubious disjunctive clause '*sensations or sensory states*'. Het dilemma waarvoor de eliminatieve perceptietheorie staat is een bijzondere vorm van de algemene aporie waarin de eliminatief materialist zich bevindt: zonder 'volkpsychologie' geen wetenschap van *cognitie*, maar met volkpsychologie geen *wetenschap* (cf. Sleutels 1988 en 1989. pp. 271-278).

²⁴ Merk overigens op dat Churchlands vergelijking tussen normale 'wit'-qualia en infrarood 'wit'-qualia niet meer is dan een retorisch hulpmiddel om duidelijk te maken hoe het bij benadering zou zijn om infrarood-ogen te hebben, *ceteris paribus*. In deze *cetera* zit enig venijn: hoe exotischer de assumpties, des te wankeler onze intuïties. Gesteld dat onze netvliezen worden vervangen door andere met infrarood-gevoelige cellen, terwijl de rest van het visueel systeem intact wordt gelaten - hoe zou de wereld er dan voor ons uitzien? Het wordt aanmerkelijk moeilijker om deze vraag te beantwoorden wanneer verder wordt aangenomen dat de operatie ook andere aanpassingen met zich meebrengt, waaronder een verbinding tussen onze nieuwe ogen en de oude neurale vectorruimte voor *lastgebonden warmteperceptie*. Zien wij dan nog hetzelfde als eerst, of *voelen* wij het veeleer? Ik weet niet hoe het bij de lezer is, maar mijn intuïties laten mij hier gewoon in de steek. Over het regime van surrealisme in de moderne filosofie van de psychologie, zie ook Kathleen Wilkes 1988. pp. 1-48; cf. Sleutels 1989. pp. 265-270.

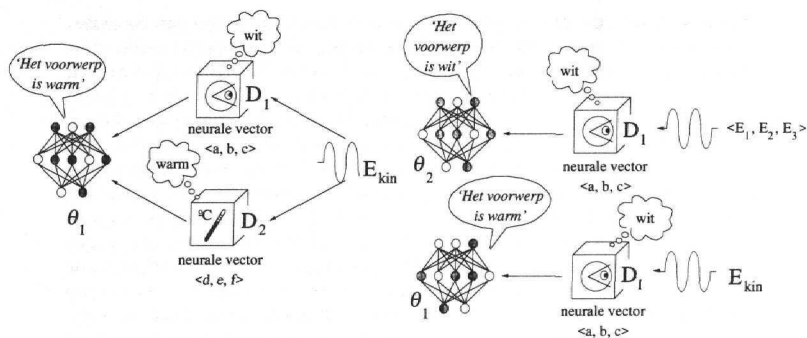
waarnemingen kan ik mij bewust worden van de inwendige toestanden die er mee gepaard gaan. Al deze bezwaren maken de stelling dat supersterke qualia *constitutief* zijn voor waarneming dubieus. Churchlands zwakke qualia dienen zich aan als een levensvatbaar alternatief: zij doen hetzelfde werk voor minder geld.

Vergeleken met zijn mededinger is Churchlands theorie beduidend sterker. Zwakke plasticiteit biedt een principiële verklaring van ons kwalitatief bewustzijn waar supersterke qualia tot ongerijmdheden leiden. Ten eerste kan Churchlands perceptietheorie verklaren hoe kwalitatief identieke gewaarwordingen kunnen leiden tot verschillende waarnemingen, en hoe kwalitatief verschillende gewaarwordingen kunnen leiden tot dezelfde waarneming. Beschouw het volgende voorbeeld. Stel dat wij onze normale ogen operatief laten vervangen door infraroodgevoelige ogen van hetzelfde type als door Churchland beschreven, terwijl de rest van onze fysiologie intact blijft. Na oefening zijn wij in staat om de temperatuur van voorwerpen zowel te zien als te voelen. Ofschoon het *dezelfde* kinetische energie is die in beide gevallen wordt geregistreerd, zijn wij intuïtief geneigd om te zeggen dat deze registraties kwalitatief *verschillende* effecten met zich meebrengen. Het beginsel van zwakke plasticiteit respecteert deze intuïtie en heeft er zelfs een plausible causale verklaring voor. Twee verschillende detectoren worden gekoppeld aan hetzelfde afleesmechanisme: kwalitatieve verschillen kunnen systematisch worden verklaard als verschillen in toestanden van de detectoren (zie fig. 1a). Een soortgelijke verklaring dient zich aan voor gevallen waarin identieke zwakke qualia leiden tot verschillende waarnemingen. In Churchlands oorspronkelijke voorbeeld is dezelfde detector gekoppeld aan verschillende invoersystemen (normale resp. infrarode netvliezen) en bovendien aan verschillende afleesystemen; de toestand van de detector bepaalt het quale, terwijl het systeem waarin de detector ligt ingebed de inhoud van de waarneming en de betekenis van de waarnemingsuitspraak bepaalt (zie fig. 1b).²⁵

De hier voorgestelde identificatie van zwakke qualia en detectietoestanden doet afstand van de 'eigen' betekenis van qualia. Daarmee verdwijnt ook het probleem van willekeur of disjunctie dat zich voordeed bij de bepaling van de inhoud van supersterke qualia. Zwakke qualia hebben geen 'inhoudelijke' specificatie nodig omdat deze voor de semantiek van waarneming irrelevant is. Hun inhoud kan ostensief gedefinieerd worden: de kwalitatieve inhoud van een zwak rood-quale is wat je voelt als je in detectietoestand $\langle x_1, x_2, \dots, x_n \rangle$ bent. Het is interessant om op te merken dat deze causale theorie verrassend genoeg ook een vindicatie is van enkele traditionele filosofische eigenschappen van qualia: subjectiviteit, onuitsprekelijkheid ('ineffability'), mogelijkheid van inversie en intersubjectieve empathie.²⁶ Zwakke qualia zijn radicaal subjectief en gebonden aan een *point of view*

²⁵ Cf. Churchlands verdediging van het functionalisme tegen qualia-objecties (1988a, pp. 38-42). Hij stelt de volgende oplossing voor: "identify the qualitative nature of your sensations-of-red with that physical feature (of the brain states that instantiates it) to which your mechanisms of introspective discrimination are in fact responding when you judge that you have a sensation-of-red. If materialism is true, then there must be some internal physical feature or other to which your discrimination of sensations-of-red is keyed: *that* is the quale of your sensations of red" (Churchland 1988a, p. 40).

²⁶ Voor een kritische bespreking van enkele van deze klassieke eigenschappen van qualia, zie o.m. Dennett 1990, Cf Churchland 1998a, pp. 38-42.



Figuur 1: reductie van zwakke qualia.

Toestanden van een detector D worden voorgesteld als een vector $\langle x_1, x_2, \dots, x_n \rangle$ in een n -dimensionale neurale phase space. Deze vectoren worden door een neurale netwerk 6. dat als afleesmechanisme functioneert, omgezet in representaties van waargenomen distale eigenschappen. Links (1a) veroorzaken verschillende detectietoestanden dezelfde waarneming van warmte (kinetische energie, E_{kin}), rechts (1b) leiden identieke detectietoestanden tot waarneming van verschillende distale eigenschappen (kinetische energie resp. een bepaalde samenstelling van elektromagnetische straling, weergegeven als triplet $\langle E_1, E_2, E_3 \rangle$).

omdat zij toestanden zijn van fysiologisch unieke detectoren in een individuele waarnemer. Toegang tot andermans qualia is aldus uitgesloten, al mag ik op grond van overwegingen van analogie aannemen dat andermans qualia op de mijne lijken. Deze overwegingen blijken een neurale basis te bezitten in de door waarnemers gedeelde fysiologie: naarmate andermans neurale vectorruimten meer verschillen van de mijne, wordt het moeilijker mij voor te stellen hoe diens qualia aanvoelen. De mogelijkheid van inter- en intrapersoonlijke omkering van qualia wordt verklaard als een omkering van een neurale vector; gedragseffecten blijven achterwege wanneer de koppeling aan het afleesmechanisme tegelijkertijd meeveert. En ten slotte zijn qualia onuitsprekelijk omdat zij geen eigen conceptuele inhoud bezitten: zij kunnen op onbepaald veel manieren conceptueel worden 'geëxploiteerd'.

4. De 'labeling fallacy'

Churchlands zwakke qualia, reductionistisch opgevat, breken het dilemma van superzwakke en supersterke qualia. Zij blijken tal van aspecten van ons kwalitatief bewustzijn inzichtelijk te kunnen maken. Dit zien wij vooral in Churchlands latere werk, waar hij zich, na de epifenomenalistische aarzeling van *Scientific realism and the plasticity of mind*, een krachtig voorstander van de reductie van qualia toont.²⁷ In deze paragraaf bekijken wij een concreet voorbeeld van zo'n reductie: kleurwaarneming. Wij zullen zien dat Churchlands ideeën over qualia worden ondersteund door recente connectionistische theorieën over mentale representatie.

²⁷ Zie o.m. Churchland 1985, 1986 en 1988a, pp. 38vv., 73vv., 146vv.

Aan de hand van dit voorbeeld gaan wij vervolgens in op de kritiek die onlangs door Fodor en Lepore is geleverd op Churchlands toepassing van de connectionistische notie van *vector coding*.²⁸

Een toonbeeld van de reductie van quaiia zoals Churchland die op het oog heeft speelt zich af op het gebied van kleurwaarneming.²⁹ Vastgesteld is dat de drie soorten kegeltjes in ons netvlies een selectieve gevoeligheid vertonen voor licht op bepaalde frequenties: 450 nm, 525 nm en 555 nm, corresponderend met resp. blauw, groen en rood licht. Eveneens staat vast dat de kleur van een voorwerp afhankelijk is van de samenstelling van de uitgestraalde of gereflecteerde elektromagnetische straling.³⁰ Deze gegevens maken het aannemelijk dat de waargenomen kleur van een oppervlak ergens in het brein wordt gerepresenteerd in termen van de activatie van elk der drie soorten kegeltjes. Deze activatietoestanden kunnen worden weergegeven als een geordend drietal $\langle a, b, c \rangle$, waarbij a , b en c een functie zijn van de gedetecteerde hoeveelheid licht in de drie kritische frequenties. Grafisch weergegeven ontstaat zo een 3-D vectorruimte of *state space*, waarin elk drietal wordt gepresenteerd door een vector (*vector coding*, zie figuur 2a). Connectionistische theorieën geven een causale verklaring van de wijze waarop dergelijke neurale representaties in netwerken worden gerealiseerd en computationeel worden getransformeerd tot andere representaties en tot gedrag. *Vector coding* beperkt zich overigens niet tot perceptie. Algemeen gesproken kunnen neurale representaties in het connectionisme worden weergegeven als een n -dimensionale vector of een partitie in een n -dimensionale ruimte; alle mentale inhouds, zowel percepten als concepten, worden op deze wijze opgevat (zie figuur 2b).³¹

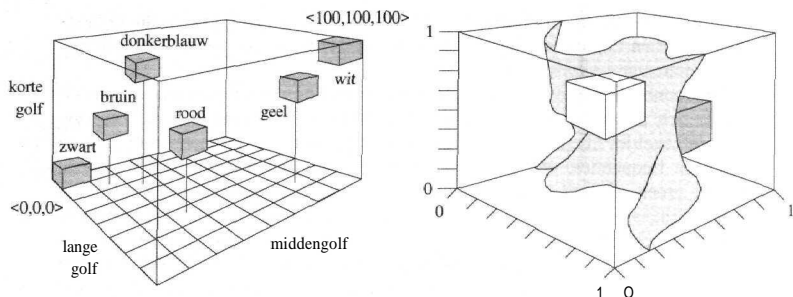
De 'qualia-kubus' in figuur 2a, of althans het netwerk waarvan de kubus een wiskundige beschrijving is, is precies de hypothetische detector die in de vorige paragraaf werd geïntroduceerd, bijv. detector D_1 uit figuur 1b. Chromatische qualia, kleurgewaarwordingen, zijn letterlijk toestanden van deze detector, d.w.z. partities op de chromatische vectorruimte. In figuur 2a is een aantal van deze partities ingetekend, met een zwart-gewaarwording linksonder en een wit-gewaarwording rechtsboven. Het model past precies bij onze uiteenzetting over zwakke quaiia in de vorige paragraaf. Daarnaast beschikt het over diverse andere deugden. Zo geeft het een goede verklaring voor onze subjectieve inschatting van gelijkenis tussen kleuren (hoe groter de afstand in de qualia-kubus, des te minder lijken twee kleuren op elkaar), en verklaart het de drie basisvormen van kleurenblindheid (als één

²⁸ Fodor & Lepore 1992, pp. 187-207.

²⁹ Zie o.m. Churchland 1986 (1989, pp. 102-105) en 1988a, p. 148. Cf. Land 1977, Campbell 1982. Voor een overzicht van inmiddels klassiek werk op het gebied van reductief onderzoek naar waarneming van o.m. kleur, kromming, diepte, textuur, spraak, zang, melodien, evenwicht en kracht, zie Richards 1988.

³⁰ Diane Ackerman maakt de volgende treffende observatie: "When light hits a red car on the streetcorner, only the red rays are reflected into our eyes, and we say 'red'. The other rays are absorbed by the car's paintjob. When light hits a blue mailbox, the blue is reflected, and we say 'blue'. The color we see is always the one being reflected, the one that doesn't stay put and is absorbed. We see the rejected color, and say 'an apple is red'. But in truth an apple is everything but red" (Ackerman 1990, p. 252).

³¹ Voor een gedegen inleiding in neurocomputatie, zie o.m. Richards 1988 en Churchland & Sejnowski 1992. Een eenvoudiger kennismaking bieden Churchland 1988b en de bijdragen van Phaf, Murre, Sleutels en Geurts in Brown, Hagoort & Meijering 1989, pp. 164-201 en 220-261.



Figuur 2: twee voorbeelden van neurale vectorcodering.

(a) Vector-representatie van het waarneembare kleurenspectrum.

Vectoren en volumes in de afgebeelde state space staan voor reflectantiewaarden (0-100%) op drie kritische frequenties in het zichtbare deel van het spectrum, corresponderend met de absorptiemaxima van pigmenten in de kegeltjes: korte golf (450 nm, blauw), middengolf (525nm, groen) en lange golf (555 nm, rood). (Vrijnaar Churchland 1989, p. 104.³²)

(b) Concepten als partities van een vectorruimte.

Het semantisch verschil tussen twee concepten, bijv. tussen 'boom' en 'struik', is een functie van de afstand tussen de corresponderende vectoren 'm' een relevante neurale vectorruimte. Links afgebeeld is een vectorruimte die door een 'gordijn' in tweeën wordt gedeeld. Alle activatiepatronen waarvan de vector vóór van het gordijn belandi zijn representaties van struiken; vectoren die erachter belanden staan voor bomen. De concepten hebben een prototypische semantische kern, welke neuraal tot uitdrukking komt als een volume in de vectorruimte. (Vrij naar Churchland 1989, pp. 169 en 203.)

van de drie koppelingen tussen oog en detector ontbreekt valt een kleurdimensie weg en klappt de kubus ineen tot een plat vlak).

Op bovenstaande poging tot reductie van quaiia is onlangs fundamentele kritiek geleverd door Jerry Fodor en Ernest Lepore.³³ Ofschoon hun kritiek betrekking heeft op Churchlands *algemene* theorie van mentale representatie, gebruiken zij precies het voorbeeld van chromatische perceptie. De essentie van de kritiek laat zich gemakkelijk uitbreiden tot andere vormen van mentale representatie, m.n. ook de *state space*-benadering van begrippen, theorieën en andere supra-zintuiglijke constructen.³⁴ Wij beperken ons hier tot de problematiek van het quäle in Churchlands filosofie.

Fodor en Lepore confronteren Churchland met een inmiddels bekend dilemma. De neurale vector $\langle a, b, o \rangle$ waartoe kleurqualia gereduceerd zouden worden is in

³² Churchlands theorie is gebaseerd op de kleurenkubus in de Retinex-theorie van Edwin Land (zie bijv. Land 1977). Voor een kritische, anti-reductionistische bespreking van de filosofische implicaties van Lands theorie, zie Campbell 1982.

³³ Fodor & Lepore 1992, pp. 187vv.

³⁴ Overigens is de keuze van het voorbeeld bepaald niet toevallig: Fodor en Lepore betogen dat Churchlands theorie een vorm van reductief neo-empirisme is - "a recidivist empiricism is in fact its boitom line" (a.w., p. 192).

wezen een fysiologische toestand — wat hier een *detectietoestand* is genoemd. De identiteit van een vector of van een partitie, bijv. de 'rood'-partitie in de qualia-kubus, is volledig afhankelijk van de dimensies van de vectorruimte. Nu doet zich de vraag voor waar deze dimensies zelf voor staan. Als zij worden gespecificeerd in zuiver psychofysische termen (bijv. in termen van de activiteit van cellen in het netvlies), dan worden de qualia *superzwak*: de theorie gaat dan niet over waarneming van kleuren maar over de fysiologie van het oog. Worden de dimensies daarentegen zelf gespecificeerd in termen van kleuren (blauw, groen en rood), dan kan de theorie weliswaar verklaren waarom een bepaalde partitie een 'bruin'-quale is (een bepaalde samenstelling van blauw, groen en rood), maar wordt de verklaring circulair. Voorondersteld is dan immers de kwalitatieve inhoud van de neurale vectoren $\langle 100, 0, 0 \rangle$, $\langle 0, 100, 0 \rangle$ en $\langle 0, 0, 100 \rangle$.

What Churchland has is a dilemma: it may be that he isn't intending to require that the dimensions of his state space correspond to properties of the contents of the mental states (objects, events) they taxonomize. In that case, he isn't doing semantics at all. He's doing, as it might be, psychophysics (. . .). If, on the other hand, Churchland is taking the talk about neural representation seriously, his move to state spaces leaves all the old problems about content identity still to be solved.³⁵

Churchland wil volgens Fodor en Lepore het midden houden tussen eliminatie en reductie. Als eliminatief materialist zou Churchland vrede moeten hebben met superzwakke qualia. "An eliminativist doesn't need a notion of semantic identity. An eliminativist doesn't want to reconstruct semantic discourse; he wants to change the topic".³⁶ Maar dit is niet wat Churchland wil. Hij houdt vast aan de notie van neurale *representaties* en is dus verplicht om een theorie over de *betekenis* hiervan te geven. Daarmee komt hij terecht bij supersterke qualia, "for a semantics taxonomizes mental states by their *contents*, not by their *causes*".³⁷

Churchland maakt volgens Fodor en Lepore een typisch connectionistische denkfout: hij verwacht de *labels* van de dimensies in zijn vectorruimte met de structuur van de *state space* zelf. Als wij de dimensies *merken* met namen voor kleuren is het natuurlijk geen wonder dat de vectoren en partities kunnen worden begrepen als functies van deze kleuren! Als naam voor deze redeneerfout stel ik hier de term *labeling fallacy* voor. Volgens de bekende kritiek van Fodor en Pylyshyn bestaat de *labeling fallacy* in het connectionisme hierin dat knopen of verzamelingen van knopen in een netwerk worden *gemerkt* met bepaalde begrippen, bijv. 'boterham' of 'kaas'.³⁸ Worden beide (verzamelingen van) knopen tegelijk actief, dan zou dit volgens connectionisten de neurale representatie zijn van een boterham met kaas. Wat in deze optiek over het hoofd wordt gezien, aldus de kritiek, is dat de labels *door de onderzoeker* zijn aangebracht en door hem ook worden geïnterpreteerd, terwijl het netwerk *zelf* ze niet kan lezen. De knopen

³⁵ Fodor & Lepore 1992, pp. 204-205.

³⁶ T.a.p.

³⁷ A.w., pp. 200-201 (curs. JS).

³⁸ Fodor & Pylyshyn 1988.

hebben geen eigen, intrinsieke inhoud, en kunnen derhalve geen mentale representaties zijn.³⁹

De *labeling fallacy* is vanzelfsprekend een drogredenering. Als je een mantelpak verwart met de couturier wiens naam op het etiket staat word je opgesloten. Het is echter nog maar de vraag of deze diagnose ook opgaat in de speciale context waarmee wij hier te maken hebben, nl. die van mentale representaties. Laten wij ter vergelijking een niet-connectionistisch alternatief nemen. Fodor heeft zich in de loop der jaren sterk gemaakt voor een theorie van mentale inhouden waarin de semantiek van onze representaties volledig wordt *weerspiegeld* in hun non-semantic (syntactische en causale) structuur. Zo hoopt hij te kunnen verklaren hoe semantische eigenschappen specifieke causale gevolgen kunnen hebben. Gesteld in termen van de *labeling fallacy* zoekt Fodor een zodanige verdeling van de labels over de functionele toestanden van onze machinerie dat wij exact de goede toestanden blijken te doorlopen: de 'kaas'-toestand en de 'boterham'-toestand veroorzaken precies de toestand die gemerkt is als 'boterham-met-kaas'. Dit onderscheidt zich natuurlijk in niets van de boven geschetste semantiek van neurale vectoren, behalve dan dat Fodor van *zijn* taxonomie meent dat zij de *intrinsieke betekenis* van de vectoren blootlegt: label en vector zijn intrinsiek verbonden. Ik kan mij niet aan de indruk onttrekken dat Fodor de couturier *zelf* in het mantelpak wil naaien.⁴⁰

Aan bovenstaande redenering ligt ten grondslag dat mentale inhouden een intrinsieke betekenis moeten hebben om cognitief relevant te zijn. Wat van deze premisse ook waar moge zijn in het geval van andere mentale representaties, voor zwakke qualia is zij te sterk. In de vorige paragraaf zagen wij dat Churchlands zwakke qualia causale inertie kunnen combineren met cognitieve relevantie. Het volstaat dat kwalitatieve verschillen en overeenkomsten tussen bewustzijnstoestanden verklaard kunnen worden in termen van causale detectietoestanden. Een intrinsieke epistemische inhoud is hiervoor niet nodig. Toegepast op de vermeende *labeling fallacy* betekent dit dat zwakke qualia een *blanco* label dragen.⁴¹ Fodor en Lepore lijken te vrezen dat Churchlands identificatie van neurale toestanden en zwakke qualia *willekeurig* is: je zou de *phase space* partities kunnen noemen hoe je wilt, want de *eigenlijke* betekenis staat niet op de assen vermeld. Eerder zagen wij echter dat voor zwakke qualia een ostensieve definitie volstaat, aangevuld met plausibele overwegingen over de functionele organisatie van onze zintuigen; meer 'inhoud' hebben zwakke qualia niet nodig. Churchlands theorie verklaart tal van eigenschappen van qualia zonder te vervallen in de absurditeiten van 'naakte', 'uitgeklede' bewustzijnsgegevens, wier 'intrinsieke inhoud' door subjecten op on-

³⁹ "Strictly speaking, the labels play no role at all in determining the operation of a Connectionist machine; in particular, the operation of the machine is unaffected by the syntactic and semantic relations that hold among the expressions that are used as labels. To put this another way, the node labels in a Connectionist machine are not part of the causal structure of the machine" (Fodor & Pylyshyn 1988. p. 17).

⁴⁰ Voor een uitgebreid commentaar op Fodors argumentatie, zie Sleutels en Geurts 1989.

⁴¹ Vanuit een ander gezichtspunt kun je ook zeggen dat zwakke qualia niet zelf een label *dragen*, maar dat zij juist het label van een detectietoestand *zijn* - een label dat *niet* eerst door de waarnemer 'gelezen' wordt om vervolgens tot waarneming te leiden!

bestemde wijze wordt 'geïnterpreteerd' - het ongerijmde idee, kortom, van supersterke qualia.

5. Sterke qualia en kattequalia

In de inleiding werd opgemerkt dat de verschuiving van eliminatie naar reductie niet de enige tendentie is in Churchlands positie t.a.v. qualia. Na de oorspronkelijke, zwakke versie is hij overgegaan op een vorm van plasticiteit die veel verder gaat: het beginsel van de *sterke plasticiteit van perceptie*. Volgens deze opvatting is de gewaarwording zelf een functie van de theoretische exploitatie ervan. Nieuwe theorieën brengen een kwalitatief nieuw waarnemingsbewustzijn roet zich mee, d.w.z. nieuwe sterke qualia. Onder invloed van theoretische ontwikkelingen in de natuurwetenschappen, in het bijzonder ook op het gebied van de neurofysiologie, zou een dramatische expansie van ons bewustzijn kunnen plaatsvinden, zowel van ons bewustzijn-van-de-wereld als van ons introspectief zelfbewustzijn. In deze paragraaf zullen wij aan de hand van enkele voorbeelden een drietal vragen proberen te beantwoorden: wat is sterke plasticiteit precies, leidt sterke plasticiteit tot supersterke qualia, en waarom zijn sterke qualia te verkiezen boven zwakke?

Wat behelst het beginsel van sterke plasticiteit precies? Philipse geeft de volgende samenvatting die als aanknopingspunt kan dienen.

According to the strong plasticity thesis, learning neurophysiological theory and being trained to use the sentence 'The bipolar cells in my retinae are firing in the pattern XYZ' on all occasions of our seeing a cat, will enable us literally to see or otherwise perceive these firing patterns on these very same occasions.⁴²

Bovenstaande bewering dient naar het rijk der fabelen te worden verwezen. Wanneer ik word getraind om bij het zien van H.M. de Koningin spontaan en non-inferentieel 'President Lincoln' te zeggen (d.w.z. wanneer ik 'President Lincoln' onder deze omstandigheden als waarnemingsuitspraak ga gebruiken), volgt uiteraard niet dat ik President Lincoln ook werkelijk 'zie of anderszins ervaar'. Dat zou immers een bovennatuurlijk waarnemingsvermogen vergen. Als dit een juiste weergave van sterke plasticiteit is, is Churchlands positie van meet af aan onhoudbaar.

Het heeft er alle schijn van dat het beginsel van sterke plasticiteit onjuist is weergegeven. Ik noem hier een drietal punten waarop commentaar mogelijk is. In de eerste plaats wordt in bovenstaande twee voorbeelden (dat van Philipse en dat van mij) Churchlands beginsel uiterst *onvolledig* toegepast. Gesuggereerd wordt dat het aanleren van een geïsoleerde nieuwe waarnemingsuitspraak, bijv. 'President Lincoln' bij het zien van Beatrix, *ceteribus paribus* (m.n. met instandhouding van *alle overige taaidisposities*), een kwalitatief nieuwe waarneming en gewaarwording oplevert. Het beginsel dat onze waarnemingsinhouden een functie van de theorie zijn wordt hier wel zeer selectief toegepast! Al mijn gedragsdisposities wijzen erop dat ik Beatrix kan onderscheiden van Lincoln, presidenten van vorstinnen, mannen van vrouwen en zo voort, met uitzondering van deze ene hard-

⁴² Philipse 1990, p. 133; cf. pp. 132 en 171-172, n. 15.

nekkige vergissing. Iemand die voor de taak staat mijn uitlatingen te interpreteren zal concluderen dat ik Beatrix waarneem en dat ik 'Beatrix' bedoel, ook al zeg ik 'President Lincoln'. Met sterke plasticiteit heeft dit niets te maken.

In de tweede plaats valt op dat in het geciteerde voorbeeld *uitwendige* waarneming wordt vervangen door *introspectie*. Bij het zien van een kat zouden wij een neurofysiologisch rapport als waarnemingsuitspraak kunnen aanleren, en dat zou op zijn beurt betekenen dat wij het beschreven fysiologische proces gewaarworden in plaats van de distale kat. De *uitwendige* waarneming van een kat zou worden vervangen door een kwalitatief andere *inwendige* waarneming.⁴³ Deze mogelijkheid wordt weliswaar niet uitgesloten door sterke plasticiteit, maar lijkt er evenmin een direct gevolg van te zijn.⁴⁴ Ter vergelijking een manier om hetzelfde resultaat te bereiken *zonder* plasticiteit. Ik zou in een quasi-Buddhistische bui kunnen besluiten mij niet meer bezig te houden met uitwendige waarneming; voortaan wijd ik mij uitsluitend aan introspectie. Katten neem ik niet meer waar, enkel inwendige toestanden. Wellicht kruist mijn geestesoog nu en dan het pad van (wat normale mensen zouden noemen) een 'kat-achtig* quale, maar van de distale oorzaken daarvan blijf ikzelf noodzakelijk onkundig. Sterke plasticiteit *dicteert* niet dat uitwendige waarneming wordt vervangen door inwendige, al kan zij wel *verklaren* waarom ik na de vervanging geen katten meer waarneem: onze gewaarwordingen veranderen met ons theoretische begrip van hun oorzaken. Vervang ik mijn kattheorie door een introspectieve theorie, dan word ik dus niet langer katten gewaar (ik heb niet meer een theorie over de distale oorzaken), maar blijft mijn bewustzijn steken in de detectie van de proximale toestand van mijn zintuigen.

Er schuilt nog een derde adder onder het gras. Naar aanleiding van zijn weergave van het beginsel van sterke plasticiteit stelt Philipse dat niemand "will seriously think that any training whatever of our skill of visual perception will enable us to see or otherwise perceive the firings of the bipolar cells in our retinae".⁴⁵ Indien waar. is deze bewering niet ter zake; is zij ter zake, dan is zij dubieus. Gelezen als een uitspraak over de mogelijkheden van visuele waarneming, waarin wij een zekere handigheid zouden moeten zien te verwerven, of over een van de andere modi van uitwendige waarneming, is de bewering beslist waar. Ik zal mijn fysiologische processen nooit 'van binnenuit* zien. Maar dit staat ook niet ter discussie. Wat het beginsel van sterke plasticiteit beweert is dat neurofysiologische processen, als interne processen, m.b.v. theorie en training *inwendig* zouden kunnen worden waargenomen in een introspectief *bewustzijn* van onze fysiologie.⁴⁶ Zo opgevat is Philipse's stelling aanmerkelijk minder plausibel. Wij kunnen ons van zoveel zaken in ons lichaam onmiddellijk bewust worden, zoals iedere sporter meteen zal beamen. Waarom dan niet ook van de fysiologische details van onze

⁴³ Uiteraard laboreert dit voorbeeld ook aan het als eerste genoemde bezwaar.

⁴⁴ Wij zullen verderop zien dat Churchland zelf groot belang hecht aan de mogelijkheid om introspectie 'aan te passen' aan uitwendige waarneming. Wat dit betreft komt Philipse's voorbeeld bepaald niet uit de lucht vallen. Hier volg ik echter vooralsnog zo strikt mogelijk het beginsel van sterke plasticiteit.

⁴⁵ Philipse 1990, p. 132.

⁴⁶ Over zelfbewustzijn of introspectief bewustzijn als een vorm van inwendige waarneming, zie bijv. Churchlands persuasieve 1988a, pp. 73vv.

neurale toestand? "It must be a dull man indeed whose appetite will not be whet", zoals Churchland zegt over de mogelijkheden die zo'n expansie van ons bewustzijn belooft."

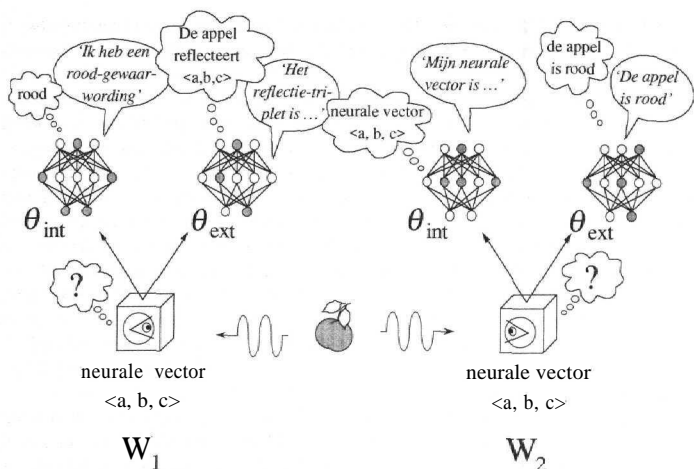
Ofschoon er bezwaren kleven aan deze eerste uitleg van het beginsel van sterke plasticiteit, zit er ook een kern van waarheid in. Al het verklarende werk werd zojuist verricht door het onderscheid tussen uitwendige waarneming en introspectie. Maar hoe verhouden deze zich eigenlijk tot elkaar en tot hun gemeenschappelijke of verschillende qualia? Een analogie werkt hier wellicht verhelderend. Om aan te geven dat introspectie niet een *speciaal* probleem oplevert ten opzichte van uitwendige waarneming vergelijkt Churchland ons waarnemingsapparaat met een ampèremeter. De schaalverdeling van de meter is geijkt om de stroomsterkte in het aangesloten circuit te meten en geeft uiteraard een verdeling in Ampères. Exact dezelfde meter, voorzien van een geschikte andere schaalverdeling, ditmaal in Gauss, kan worden gebruikt om de sterkte van het magnetisch veld in de spool die de wijzer aandrijft weer te geven. De meter werkt dan "in introspectieve mode", zoals Churchland het plastisch uitdrukt.⁴⁸

Latén wij deze analogie proberen toe te passen op het eerder besproken voorbeeld van de kleurwaarneming (zie figuur 3). Bij de waarneming van een appel, waarvan het oppervlak a% rood licht, b% groen licht en c% blauw licht weerkaatst, wordt een neurale vector $\langle a, b, o \rangle$ geactiveerd in de chromatische *phase space*. De waarnemer heeft tevens een interne theorie die deze detectietoestand conceptueel exploiteert, een 'schaalverdeling'. Stel dat dit in het geval van waarnemer W_1 de theorie over kleuren als reflectietripletten is; W_1 verwoordt de waarneming spontaan en non-inferentieel als: 'De appel weerkaatst licht met samenstelling $\langle a, b, c \rangle$ '. Een andere waarnemer W_2 met dezelfde detectietoestand is conceptueel minder bedeed; zijn waarnemingsuitspraak luidt: 'De appel is rood'. Intussen richten beide waarnemers zich ook introspectief op dezelfde detectietoestand. Zij doen dit aan de hand van hun theorieën over de interne mentale organisatie van de mens. W_1 bedient zich van eenvoudige volkpsychologie. Zijn introspectierapport luidt: 'Ik heb een rood-gewaarwording'. W_2 , daarentegen, beschikt over gedetailleerde kennis van de neurofysiologie van kleurwaarneming, en heeft geleerd deze te gebruiken in zijn spontane, non-inferentiële introspectierapporten. Zijn introspectierapport luidt: 'Er is een vector $\langle a, b, o \rangle$ actief in mijn chromatische *phase space*'.

Welke qualia spelen in dit voorbeeld een rol? Als wij zouden uitgaan van *zwakke* qualia is er maar één enkel quale: de rood-gewaarwording die neuraal geïnstantieerd wordt als vector $\langle a, b, o \rangle$ in de chromatische *phase space*. Deze ene gewaarwording wordt op vier verschillende manieren conceptueel geëxploiteerd.

⁴⁷ Cf. Churchland 1979, p. 7.

⁴⁸ Churchland 1979, p. 40. Ofschoon dit voorbeeld stamt uit Churchlands periode van 'zwakke qualia', accepteert hij het duidelijk ook in zijn latere werk. Dit blijkt uit de beschrijving die hij geeft van mogelijke reducties van introspectieve waarnemingen, die gelijke pas houden met de reductie van uitwendige gewaarwordingen. Cf. de eerder geciteerde passage uit Churchland 1986 (1989, p. 106): "This more penetrating conceptual framework might even displace the common-sense framework as the vehicle of *intersubjective description and spontaneous introspection*" (cursief JS).



Figuur 3: proliferatie van sterke qualia?

Toestanden van een detector D worden voorgesteld als een vector $\langle x_1, x_2, \dots, x_n \rangle$ in een n -dimensionale neurale phase space. Deze vectoren worden afgelezen door telkens twee neurale netwerken 0 , waarbij het ene netwerk een introspectieve waarneming oplevert, het andere een uitwendige waarneming.

Gaan wij er evenwel van uit dat qualia *sterk* zijn, dan dienen zich in beginsel vier verschillende qualia aan, voor elke theorie één. Elke nieuwe theorie brengt immers *ex hypothesi* een eigen, kwalitatief nieuwe gewaarwording met zich mee, gebaseerd op dezelfde detectietoestand. Dit leidt op het eerste gezicht tot ongerijmde gevolgen. Intuïtief gezien zijn de vier qualia in bovenstaand voorbeeld gewoon *identiek*. Deze intuïtie wordt gestaafd door de alledaagse ervaring: het aanleren van de theorie van kleuren laat je kleuren *qualia* intact. Bovendien zijn sterke qualia, zo opgevat, onverenigbaar met Churchlands vereenzelviging van qualia en detectietoestanden. In het voorbeeld leveren exact dezelfde 'causal middle-men' immers vier verschillende sterke qualia op. De enige uitweg zou zijn om sterke qualia te identificeren met detectietoestanden *plus* theorie - een oplossing die qualia *conceptueel* (en daarmee *supersterk*) maakt.

Deze conclusie is echter te overhaast. Weliswaar volgt uit sterke plasticiteit dat nieuwe theorieën nieuwe qualia met zich mee *kunnen* brengen, maar niet dat dit ook noodzakelijk het geval is. Zouden sterke qualia ondanks de exploitatie door nieuwe theorieën identiek blijken te zijn, dan is hiervoor een extra verklaring nodig. Churchland kan zo'n verklaring geven: zij volgt uit de reductie van qualia tot detectietoestanden. In wezen gaat het om een verfijning van wat wij eerder bij zwakke qualia hebben gezien. Zwakke qualia waren ostensief gedefinieerde detectietoestanden, waarbij impliciet bleef of deze toestanden moesten worden voorge-

steld als een unieke vector dan wel als een partitie (verzameling vectoren) in de desbetreffende vectorruimte. De notie van sterke qualia kan m.i. worden gezien als een poging om deze bepaling preciezer te maken. Ik stel de volgende samenhangende definities voor.

5. Twee theorieën θ_1 en θ_2 , die dezelfde neurale vectorruimte V conceptueel exploiteren hebben dezelfde sterke qualia wanneer zij dezelfde partitionering van V gebruiken.
6. θ_1 heeft andere sterke qualia dan θ_2 wanneer θ_1 -partities verschillen van θ_2 -partities, in het bijzonder wanneer θ_1 -partities een subpartitie zijn van θ_2 -partities, of omgekeerd.
7. Of het perceptierooster van θ_1 verschilt van dat van θ_2 , blijkt uit het discriminatieve gedrag dat zij mogelijk maken, m.n. uit de waarnemingsuitspraken.

Omwille van de eenvoud en aanschouwelijkheid zal ik mij hier concentreren op gevallen waarin de perceptieroosters van θ_1 en θ_2 zich verhouden als partities en subpartities, d.w.z. waarin θ_1 een verfijning biedt t.o.v. θ_2 , of omgekeerd.

Laten wij nu teruggaan naar figuur 3 en de *uitwendige* waarnemingen van W_1 en W_2 vergelijken.⁴⁹ W_1 heeft een '<a, b, o'-quale, W_2 een 'rood'-quale. In *abstracto* doen zich hier vier mogelijkheden voor:

8. Het 'rood'-quale en het '<a, b, c>'-quale zijn identiek en stellen W_1 en W_2 in staat tot hetzelfde discriminatieve gedrag.
9. Het 'rood'-quale en het '<a, b, o'-quale zijn identiek, maar het discriminatieve gedrag van W_1 en W_2 kan verschillen.
10. Het 'rood'-quale en het '<a, b, o'-quale zijn verschillend maar desondanks kunnen W_1 en W_2 geen verschillen in discriminatief gedrag vertonen.
11. Het 'rood'-quale en het '<a, b, o'-quale zijn verschillend, terwijl ook het discriminatieve gedrag van W_1 en W_2 kan verschillen.

Mogelijkheid (8) is intuïtief plausibel; het is de optie die aanvankelijk als objectie tegen sterke plasticiteit werd ingebracht. Ook al bevat W_2 's theorie in beginsel de *conceptual resources* om meer fijngemaasde distincties aan te brengen (de variabelen in een omschrijving als '<a, b, c>' zijn continu), deze onderscheidingen vindt hij niet terug in zijn spontane waarnemingen. Dit blijkt objectief uit het gedrag van W_2 , m.n. uit zijn waarnemingsuitspraken. Kras gezegd, W_2 kan <a, b, o niet beter onderscheiden van <a+ Δ a, b+ Δ b, c+ Δ c> dan W_1 .

Mogelijkheden (9) en (10) worden door de hier beschreven vorm van sterke plasticiteit uitgesloten, terwijl mogelijkheid (11) het werkelijk *nieuwe* van sterke qualia is: dat dezelfde zintuigen en detectoren op kwalitatief nieuwe wijze kunnen worden gebruikt. Hoe plausibel is deze suggestie? Laten wij ter vergelijking een voorbeeld nemen van de fervente antifysicist Frank Jackson. Jackson introduceert Fred, een normaal mens met één bijzondere gave: hij kan twee soorten rood onderscheiden waar wij maar één tint zien.

⁴⁹ Voor de inwendige waarnemingen van W_1 en W_2 kan een vergelijkbare analyse worden gegeven.

To him (Fred, JS) red, and red, are as different from each other and all the other colors as yellow is from blue. And his discriminatory behavior bears this out: he sorts red₁ from red, tomatoes with the greatest ease in a wide variety of circumstances. Moreover, an investigation of the physiological basis of Fred's exceptional ability reveals that Fred's optical system is able to separate out two groups of wavelengths in the red spectrum as sharply as we are able to sort out yellow from blue. I think that we should admit that Fred can see, really see, at least one more color than we can; red₁ is a different color from red₂. We are to Fred as a totally red-green color-blind person is to us.⁵⁰

Fred ziet één kleur meer dan wij (of eigenlijk: hij mist een kleur en krijgt er twee voor in de plaats). Het is plausibel dat ook zijn beide rood-qualia verschillen van elkaar en van de onze. Deze consequentie stemt volledig overeen met de theorie van sterke qualia. Sterker nog, sterke qualia kunnen dit verschijnsel verklaren in termen van Freds fijnere perceptieooster: de rood-partitie in onze qualia-kubus valt bij Fred uiteen in twee afzonderlijke partities. Fred heeft van nature wat volgens Churchland door het leren van een nieuwe theorie kan worden verworven: nieuwe sterke qualia. Het belangrijke punt is in beide gevallen echter hetzelfde, nl. dat de waarnemer in staat moet zijn om de relevante onderscheidingen te maken in zijn waarnemingsgedrag. Gesteld dat aan deze voorwaarde is voldaan, dan is het aannemelijk dat de oorzaken van het discriminatieve gedrag (verschillende detectietoestanden) samenvallen met verschillende sterke qualia.⁵¹

Een vraag van een heel andere orde is of de fysiologie van het menselijke zenuwstelsel dergelijke meer fijnmazige percepties toelaat. Onze receptoren en detectiemechanismen kennen ongetwijfeld endogene beperkingen. Sterke plasticiteit speelt zich af binnen deze marges. Het is echter onwaarschijnlijk dat ons huidige begrippenapparaat de capaciteit van de beschikbare detectoren reeds volledig uitput. Churchland geeft diverse rekenvoorbeelden van het indrukwekkende representatievermogen van neurale vectorruimtes, zelfs bij netwerken van zeer bescheiden omvang.⁵² Tien distincte posities langs elke as van de qualia-kubus in figuur 2a stellen de waarnemer al in staat om maximaal 10^3 kleuren te onderscheiden en evenzovele 'atomaire' sterke qualia te activeren. Waarschijnlijk is het aantal distincte posities langs elke as in werkelijkheid nog veel groter. In veel gevallen van waarneming is bovendien het aantal dimensies van de vectorruimte groter. Dertig posities langs zeven assen leveren 22 miljard distincte toestanden op. Dergelijke getallen doen vermoeden dat endogene bepaalde marges geen beletsel vormen voor de mogelijkheid van sterke plasticiteit.⁵³

⁵⁰ Jackson 1982 (1990, p. 470).

⁵¹ Jackson is het overigens ie doen om een andere vraag, nl. *welk soon gewaarwording* Fred heeft als hij rood, of rood, ziet. "What is the new color or colors like? We would dearly like to know but do not; and it seems that no amount of physical information about Fred's brain and optical system tells us. (...) It follows that Physicalism leaves something out" (a.w., pp. 470-471).

⁵² Zie bijv. Churchland 1986 (1989, pp. 102vv), 1989, pp. 131-132.

⁵³ Een andere vraag is of Churchland de plasticiteit van de exploiterende theorieën kan verklaren. Het connectionisme geeft ook hertoe een aanzet. Begrippen, theorieën en andere suprazinziug-

Bovenstaande uitleg mag hier volstaan ter verduidelijking van Churchlands beginsel van sterke plasticiteit. In het bijzonder slagen definities (5)-(7) erin om sterke qualia te begrijpen vanuit een reductionistisch perspectief. Twee belangrijke vragen staan nu nog open: leidt sterke plasticiteit tot supersterke qualia, en waarin onderscheiden sterke qualia zich van zwakke? De twee vragen hangen samen en kunnen zelfs in de vorm van een dilemma worden gebracht. Onderscheiden sterke qualia zich niet van zwakke, dan zijn zij overbodig; Churchlands nieuwe plasticiteitstheorie wordt daarmee een loze exercitie. De enige manier waarop sterke qualia zich kunnen onderscheiden lijkt evenwel door zich *supersterk* te maken. In dat geval is het nog maar de vraag of sterke plasticiteit zich leent om Churchlands empirisch realisme te combineren met een netwerktheorie van betekenis. Met een bespreking van dit dilemma besluiten wij hier.

De nauwe koppeling tussen sterke qualia en theorieën wekt de indruk dat qualia een soort van eigen conceptuele identiteit bezitten. Dit laat zich illustreren aan de hand van definities (5)-(7), waarin sprake is van partities die door theorieën worden 'opgepakt'. Nieuwe sterke qualia partitioneren de oude vectorruimte op een nieuwe manier, overeenkomstig het conceptuele raster van de nieuwe theorie. Daarmee krijgen qualia echter een verkapte conceptuele identiteit toegemeten. Freds sterke quale 'rood₁' is een zelfstandige partitie die als zodanig volstaat om het oordeel 'rood₁' te activeren; de betekenis van het waarnemingsoordeel is dus primair een functie van het quale en niet van de theorie.

Een zwak punt in deze redenering is dat voorbij wordt gegaan aan het *metaforisch* karakter van uitdrukkingen als 'vectorruimte' en 'partitie'. Partities zijn niet meer dan verzamelingen mogelijke detectietoestanden, d.w.z. grafische of wiskundige representaties van activatiepatronen in een neurale netwerk. Nieuwe sterke qualia hanteren een nieuwe taxonomie van detectietoestanden, maar maken uiteraard gebruik van dezelfde oude vectoren en patronen. Op dit punt verschillen sterke qualia allerminst van zwakke. De wijze waarop de vectoren worden gegroepeerd wordt niet 'bottom-up' bepaald door de vectoren zelf, maar 'top-down' in de conceptuele exploitatie door de theorie.⁵⁴ Er is niet zoiets als een 'intrinsieke'

lijke constructen hebben evenals qualia een reëel fundament in het zenuwstelsel; zij kunnen algemeen gesproken worden voorgesteld als partities op neurale vectorruimten (zie fig. 2b en de toelichting aldaar). Dit betekent dat zij in beginsel bepaalde eigenschappen van het neurale substraat kunnen 'overnemen', waaronder de aantoonbare fysiologische plasticiteit. In het neurale netwerk van de hersenen is niet alleen de waarde van de bestaande connecties veranderlijk (graad van inhibitie of excitatie), maar ook is de structuur van *het net zelf* aan verandering onderhevig (vorming van nieuwe connecties). Beide vormen van neurale plasticiteit kunnen bijdragen aan ons begrip van leerprocessen en openen aldus ook een mogelijkheid om conceptuele plasticiteit te verklaren. Een bespreking van leerprocessen vanuit dit perspectief, zie Kandel & Hawkins 1992 en Churchland & Sejnowski 1992, pp. 239-330.

⁵⁴ Voor een blik op het micromechanisme van conceptuele exploitatie, zie Lehky & Sejnowski 1988 en Churchland & Sejnowski 1992, pp. 183vv. "A unit's representative significance may be revealed only when we look at its projective as well as its receptive field: cells are not just reactors, they are also agents in a larger economy". Een sprekend voorbeeld is het onderzoek van Lehky en Sejnowski naar de extractie van informatie over 3-D kromming uit de verdeling van grijswaarden. Dat een gegeven knoop een hoek- of streepdetector is, bijvoorbeeld, wordt niet alleen *bottom-up* bepaald door het receptieve veld. De knoop ontvangt als input immers geen boeken en strepen, maar enkel grijswaarden. De streep/hoek-betekenis ontleent de knoop aan zijn

partitionering van mijn vectorruimte. Als nieuwe sterke qualia anders 'aanvoelen' dan oude, zoals intuïtief naar voren kwam uit Jacksons voorbeeld van Fred, dan kan hierover alleen worden gezegd dat dit verschil zijn grond vindt in verschillende partitioneringen. Net zoals zwakke qualia zijn sterke qualia voor het overige uitsluitend ostensief te definiëren. Hieruit volgt echter niet dat deze partities een 'zelfstandig bestaan' hebben en kunnen worden 'waargenomen' door het subject.

Laten wij nog een laatste maal terugkeren naar Philippe's weergave van het beginsel van sterke plasticiteit, geciteerd aan het begin van deze paragraaf. Daarin werd de mogelijkheid geopperd dat Churchlands sterke qualia (neurale partities) hun identiteit wellicht onlenen aan een *introspectieve* theorie. Na het leren van een nauwkeurige theorie over de menselijke fysiologie kunnen wij een introspectierapport van onze eigen detectietoestanden geven. Wij zouden ons bewust kunnen worden van de individuele vectoren en hun mogelijkheden tot partitionering. Churchland laat zich op tal van plaatsen in deze zin uit. Hij benadrukt dat onze huidige introspectieve theorie (volkpsychologie) achterloopt bij ons beeld van de rest van de wereld. De achterstand zou kunnen worden ingelopen m.b.v. een geavanceerd neurowetenschappelijk zelfbeeld ter vervanging van de volkpsychologie. Het gevolg hiervan is een dramatische expansie en verfijning van ons zelfbewustzijn.

This more penetrating conceptual framework might even displace the commonsense framework as the vehicle of intersubjective description and spontaneous introspection. Just as a musician can learn to recognize the constitution of heard musical chords, after internalizing the general theory of their internal structure, so may we learn to recognize, introspectively, the *n-dimensional* constitution of our *subjective* sensory qualia, after having internalized the general theory of *their* internal structure.⁵⁵

Gesteld nu dat wij de beschikking krijgen over zo'n fijnmazig introspectieraster, kan dit dan niet worden gebruikt om ook onze *uitwendige* waarneming te verbeteren? Door studie van supersterke kattequalia zou ik meer over katten aan de weet komen.

Churchlands theorie verzet zich echter tegen deze interpretatie. Stel dat ik honden niet kan onderscheiden van katten. Mijn detector veegt 'hond'-vectoren en 'kat'-vectoren in één neurale partitie bijeen. De vraag is nu of het mogelijk is dat ik door nauwkeurige introspectie van mijn vectorruimte erin slaag deze partitie opnieuw te partitioneren in een honden- en een kattenruimte, zodanig dat ik op grond hiervan *vervolgens* in staat ben ook in de uitwendige waarneming honden van katten te onderscheiden. Het antwoord luidt ontkennend. Hoezeer ik ook mijn best doe, neurofysiologische introspectie alleen levert slechts 'naamloze', neurofysiologische vectoren op. Vectoren presenteren zich niet 'geprepartitioneerd', in introspec-

plaats in een collectief van *projectieve* velden. Het hier geschetste beeld van de semantiek van een netwerk is het exacte tegendeel van de klassieke karikatuur van multidimensionale empirische 'compounders'. Een vergelijkbare conclusie heb ik in een eerder artikel over Churchland bepleit n.a.v. een methodologische evaluatie van het werk van Hubel en Wiesel (Sleutels 1988).

⁵⁵ Churchland 1989, p. 106.

tie noch in uitwendige waarneming. Partities zijn slechts het resultaat van conceptuele exploitatie door een theorie. Dit laat onverlet dat ik mij volgens Churchland introspectief bewust kan worden van 'hond'-partities en 'kat'-partities *nadat* ik deze distinctie m.b.v. een geschikte theorie in de uitwendige waarneming heb leren maken. In deze zin gaat uitwendige waarneming, zoals Kant al meende, transcendentaal vooraf aan introspectie.⁵⁶

Sterke qualia zijn niet supersterk. Zij onderscheiden zich in wezen niet van zwakke qualia. Hun schijn van eigen identiteit is een abstractie *post factum*; zij zijn causaal inert; zij zijn reduceerbaar tot detectietoestanden. De punten van verschil tussen sterke en zwakke qualia zijn stellingen (9) en (11), hierboven. In een van de eerder besproken voorbeelden hadden infrarood-waarnemers hetzelfde zwakke 'wit'-quale bij waarneming van warme voorwerpen als wij hebben bij waarneming van witte voorwerpen. Als qualia sterk zijn is deze mogelijkheid op grond van stelling (9) uitgesloten. Dit lijkt mij geen groot verlies, maar veeleer een plausible correctie van een intuïtie die toch a! wankel was.⁵⁷ Het andere verschil schuilt in stelling (11): doen er zich *nieuwe* qualia voor, die een systematisch effect op het waarnemingsgedrag hebben, dan kan dit worden verklaard in termen van een nieuwe partitionering van detectietoestanden. Ook in dit opzicht is de theorie van sterke qualia een correctie op de zwakke opvatting. Zij onderstreept de rol van conceptuele exploitatie in het perceptieproces, echter zonder daarmee in het labyrint van supersterke qualia met een intrinsieke inhoud te geraken.

6. Conclusie

Churchlands theorie van zwakke qualia doorbreekt het dilemma van superzwakke en supersterke qualia. Zwakke qualia geven een plausible, reductieve verklaring van tal van aspecten van ons kwalitatief bewustzijn. De theorie van sterke qualia, een gevolg van het beginsel van de sterke plasticiteit van perceptie, onderscheidt zich in dit opzicht niet wezenlijk van zijn voorganger, maar brengt er enkele correcties op aan. Het dilemma waarmee critici Churchlands filosofie confronteren - *of* cognitief irrelevant (superzwakke qualia), *of* circulaire (supersterke qualia) — kan door het beginsel van de plasticiteit van perceptie adequaat worden beantwoord. Bezien vanuit dit perspectief blijkt Churchlands perceptieeler in veel opzichten Kantiaans aan te doen. Zo komt het idee van 'conceptuele exploitatie van natuurlijke informatie' overeen met Kants categoriale exploratie van de 'reine Mannigfaltigkeit' van aandoeningen; in beide gevallen wordt ontkend dat de primaire 'gegevens' van het bewustzijn een intrinsieke identiteit bezitten op grond waarvan zij cognitief zouden kunnen worden verwerkt. Churchland zelf gebruikt dikwijls de Kantiaanse notie van een 'sensory manifold' voor zijn idee van een neurale vectorruimte. Dit wijst op mogelijke andere overeenkomsten tussen Churchlands positie en die van Kam, hetgeen echter stof is voor een ander artikel.⁵⁸

⁵⁶ Cf. Kant 1789, 'Widerlegung des Idealismus' (B 274vv).

⁵⁷ Zie ook noot 24.

⁵⁸ Philippe betoogt o.m. dat "Churchland's naturalism or scientific realism collapses into a 'neo-Kantian' transcendental idealism" (1990, p. 169). Voor een vergelijking van Churchlands idee van exploitatie van natuurlijke informatie met een andere vorm van neo-Kantianisme, nl. een 'hermeneutische ontologie' verwant aan Putnam's 'intern realisme', zie Sleutels & Corbey 1992.

Bibliografie

- Brown, C., P. Hagoort & T.C. Meijering, red., 1989, *Vensters op de geest. Cognitie op het snijvlak van filosofie en psychologie*. Grafiet, Utrecht.
- Campbell, K., 1982, 'The implications of Land's theory of color vision'. Herdrukt in Lycan 1990, pp. 567-576.
- Churchland, P.M., 1979, *Scientific realism and the plasticity of mind*. Cambridge, Cambridge UP.
- Churchland, P.M., 1985, 'Reduction, qualia, and the direct introspection of brain states'. *Journal of Philosophy* 82. Herdrukt in: Churchland 1989, pp. 47-66.
- Churchland, P.M., 1986, 'Some reductive strategies in cognitive neurobiology'. *Mind* 95. Herdrukt in: Churchland 1989, pp. 77-110.
- Churchland, P.M., 1988a, *Malter and consciousness. A contemporary introduction to the philosophy of mind*. Revised edition (oorspr. druk 1984). Cambridge, MA. MIT Press.
- Churchland, P.M., 1988b, 'Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor'. *Philosophy of Science* 55, pp. 167-187. Herdrukt in Churchland 1989, pp. 255-279.
- Churchland, P.M., 1989, *A neurocomputational perspective. The nature of mind and the structure of science*. Cambridge, MA, MIT Press.
- Churchland, P.S. & T.J. Sejnowski, 1992, *The computational brain*. Cambridge, MA, MIT Press.
- Dennett, D.C., 1990, 'Quining qualia'. Herdrukt in Lycan 1990, pp. 519-547.
- Dennett, D.C., 1991, *Consciousness explained*. Harmondsworth, Allen Lane/Penguin Press.
- Devitt, M., 1990, 'A narrow representational theory of the mind'. In: Lycan 1990, pp. 371-398.
- Fodor, J. & E. Lepore, 1992, *Holism. A shopper's guide*. Oxford, Basil Blackwell.
- Fodor, J.A. & Z.W. Pylyshyn, 1988, 'Connectionism and cognitive architecture. A critical analysis'. *Cognition* 28, pp. 3-71.
- Fodor, J.A., 1988, 'A reply to Churchland's "Perceptual plasticity and theoretical neutrality"'. *Journal of Philosophy* 55, pp. 188-198.
- Jackson, F., 1982, 'Epiphenomenal qualia'. *Philosophical Quarterly* 32, pp. 127-136. Herdrukt in: Lycan 1990, pp. 469-477.
- Kandel, E.R. & R.D. Hawkins, 1992, 'The biological basis of learning and individuality'. *Scientific American* 267, nr.3, pp. 52-61.
- Kant, I., 1781 (A), 1789 (B), *Kritik der reinen Vernunft*.
- Kripke, S., 1972, 'Naming and necessity'. In: D. Davidson & G. Harman, eds., *Semantics of natural language*. Dordrecht/Boston, Reidel.
- Land, E., 1977, 'The retinex theory of color vision'. *Scientific American* 237 no. 6, pp. 108-128.
- Lehky, S.R. & T.J. Sejnowski, 1988, 'Network model of shape-from-shading: neural function arises from both receptive and projective fields'. *Nature* 333, pp. 452-454.
- Lycan, W.G., ed. 1990, *Mind and cognition. A reader*. Oxford, Basil Blackwell.
- Mandelbaum, M., 1964, 'Locke's realism'. In: *Philosophy, science, and sense perception. Historical and critical studies*. Baltimore, Johns Hopkins Press.
- Nagel, Th., 1974, 'What it is like to be a bat'. *Philosophical Review* 83, pp. 435-450.
- Philipse, H., 1990, 'The absolute network theory of language and traditional epistemology. On the philosophical foundations of Paul Churchland's scientific realism'. *Inquiry* 33, pp. 127-178.
- Quine, W.V., 1990, *The pursuit of truth*. Cambridge, MA, Harvard UP.
- Richards, W., ed., 1988, *Natural computation*. Cambridge, MA, MIT Press.
- Robinson, H., 1982, *Matter and sense*. Cambridge, Cambridge UP.
- Searle, J., 1980, 'Minds, brains, programs'. *Behavioral and Brain Sciences* 3, pp. 417-424.
- Sellars, W., 1963, *Science, perception and reality*. London, Routledge & Kegan Paul.
- Sleutels, U.M., 1988, 'Eliminatief materialisme en de autonomie van de bottom-up benadering.' *Algemeen Nederlands Tijdschrift voor Wijsbegeerte* 80, pp. 41-62.
- Sleutels, Jan, 1989, 'Zakjapanners, superchinezinnen en hersenschimmen in de hedendaagse filosofie van de cognitiewetenschappen.' In: Brown, Hagoort & Meijering 1989, pp. 252-290.
- Sleutels, J.J.M., 1991, 'Natuurlijke teleologie. Het probleem van misrepresentatie in de fysicistische philosophy of mind.' In: G. Debrock, red., *Rationaliteit kan ook redelijk zijn. Bijdragen over het probleem van de teleologie*. Van Gorcum, Assen/Maastricht, 1991, pp. 129-146.

- Sleutels, Jan & Bart Geurts, 1989, 'Knopen en connecties. Filosofische aspecten van het connectionisme.' In: Brown, Hagoort & Meijering 1989, pp. 220-251.
- Sleutels, Jan & Raymond Corbey, 1992, 'Darwin, Dilthey, and beyond. Science, literature, and hermeneutical ontology'. *Tractrix. Yearbook for the History of Science, Medicine, Technology and Mathematics*, 4 (1992), 114-125.
- Wilkes, K., 1988, *Real people. Personal identity without thought experiments*. Clarendon Press, Oxford.
- Yolton, J.W., 1987, 'Representation and realism. Some reflections on the Way of Ideas.' *Mind* 96, pp. 318-330.

Abstract

The philosophy of Paul Churchland is marked by a tension between elimination and reduction of mental representations. This also applies to Churchland's account of qualia, which are said to be cognitively dispensable, while at the same time being reducible to neural states. In this paper I try to alleviate the tension and remove some of the ambiguities inherent in Churchland's philosophy of mind. A distinction is made between four possible constructions of the phenomenally 'given', viz., as superweak, weak, strong, or superstrong qualia. Churchland endorses weak and strong qualia, corresponding to two versions of his thesis of the plasticity of perception. I defend his position against the criticism that qualia are either superweak (hence cognitively irrelevant) or superstrong (endowed with a proper conceptual identity).