Liesbeth Flobbe

Recensie van:
Roger Penrose: *The Emperor's New Mind*
Oxford University Press, Oxford, 1989, 466 p.

Tussen quantumdeeltjes en geesten...

Computers nemen een steeds grotere plaats in in het dagelijks leven. Niemand kan ontkennen dat computers steeds slimmer worden. In veel science fiction wordt dan ook een toekomstbeeld geschetst van computers die op zijn minst gelijkwaardig aan mensen zijn. Hoe moeilijk het ook kan zijn om zulke computers te *maken*, volgens de aanhangers van 'strong AI' (AI = Artificiële Intelligentie) is dit niet onmogelijk. Er is *in principe* niets dat computers niet zouden kunnen en mensen wel. Het is alleen een kwestie van tijd en moeite voordat computers een volwaardige geest hebben.

Roger Penrose gaat in zijn boek *The Emperor's New Mind* hard tegen dit standpunt in. Als hoogleraar wiskunde is hij op het gebied van de Artificiële Intelligentie slechts een buitenstaander. Dat weerhoudt hem er geenszins van om zich in de discussie te mengen. Een computer zal volgens hem nooit een geest hebben. De keizers van de AI hebben helemaal geen kleren!

Om zijn punt duidelijk te maken, heeft Penrose een lange argumentatie nodig. In meer dan vierhonderd bladzijden behandelt hij onderwerpen uit de informatica, de wiskunde, de klassieke natuurkunde, de quantummechanica, de kosmologie, de neurowetenschap en de psychologie. Al deze onderwerpen zijn volgens hem relevant als we willen begrijpen hoe de geest werkt. Overal legt hij verbanden tussen verschillende disciplines. De manier waarop al deze disciplines met elkaar te maken hebben, is vanuit wetenschapsfilosofisch oogpunt zeer interessant. Heeft het hele 'lage' niveau van de quantummechanica werkelijk invloed op de werking van het brein? En is kennis van de quantummechanica nodig om de geest te kunnen doorgronden? Hoe zal een theorie van de geest er uit komen te zien?

Noodzakelijkerwijs moet ik mij bij de bespreking van het boek enigszins beperken. Ik wil daarom vooral aandacht besteden aan drie verschillende disciplines, op drie verschillende aggregatieniveaus: de quantummechanica, de (cognitieve) psychologie, maar vooral ook het tussenliggende niveau van de neurowetenschap, waarvan het belang volgens mij niet onderschat moet worden. Dit betekent niet dat de andere disciplines die Penrose aan bod laat komen niet interessant zijn. De wiskunde is van groot

belang voor de argumentatie, maar daarnaast heeft Penrose uitgesproken ideeën over de werkelijkheid van wiskundige waarheden en concepten; ideeën die ook van belang zijn bij de vraag "Waarom komen de ideeën in onze geest overeen met de wereld om ons heen?" Ook kosmologische vraagstukken - over de oorsprong van het universum en de richting van de tijd - zijn relevant als we willen weten waarom we *tijd* op een bepaalde manier ervaren. Maar over deze aspecten van de geest zal ik het verder niet hebben.

Ik zal beginnen met een korte weergave van de lijn van Penrose's boek. In het boek zelf is het vanaf het begin duidelijk waar Penrose heen wil met zijn argumentatie; en dat is geen overbodige luxe. Penrose begint met de stelling dat een computer geen geest heeft en ook nooit zal hebben. Als we willen begrijpen wat de verschillen tussen mens en computer zijn, dan moeten we kijken naar de fysieke realisatie van deze 'systemen': de hardware van een computer versus het brein van een mens. Vervolgens gaat Penrose in op de beperkingen van computers, vanuit een abstracte beschrijving van computers als algoritmische systemen. Hij laat zien dat mensen slimmer kunnen zijn dan algoritmes. Dit kan alleen maar mogelijk zijn, als ook het brein zelf op een of andere manier niet-algoritmisch werkt. Een groot deel van het boek wordt daarom in beslag genomen door een zoektocht naar bruikbare, niet-algoritmische systemen in de fysische wereld. Deze zoektocht komt uit bij de quantummechanica. Na een uitgebreide verhandeling over de huidige en de toekomstige quantummechanica vraagt Penrose zich af of de wondere wereld van de quantummechanica werkelijk van invloed kan zijn op ons brein. Hij beantwoordt dit met een twijfelend "ja": het zou kunnen, maar we weten nog te weinig om te bepalen hoe. Maar tegelijkertijd *moet* het wel zo zijn, als we aannemen dat het brein een niet-algoritmisch systeem is. Penrose besluit met een aantal ideeën over hoe een toekomstige 'physics of mind' eruit zou kunnen zien. Het boek presenteert dan ook geen dichtgetimmerde theorie. Penrose probeert vooral te laten zien waar onze huidige kennis tekort schiet, en in welke gebieden verder onderzoek wellicht nog onverwachte en interessante resultaten kan opleveren.

Kan een computer denken?

De vraag of een computer een geest heeft, leidt onvermijdelijk tot een uitgebreide discussie over de betekenis van het woord geest. Bij de aanhangers van 'strong AI' gaat het vooral om het denkvermogen, en Penrose is bereid hierin mee te gaan. Wel merkt hij meteen op dat denken gekoppeld moet zijn aan bewustzijn, wil er sprake zijn van echt denken en niet van simulatie. Maar de vraag "Heeft een computer een geest?" is vereenvoudigd tot "Kan een computer denken?"

Een traditioneel antwoord op deze vraag is de Turing test: als een menselijke ondervrager niet meer in staat is een computer en een mens van elkaar te onderscheiden, kan de computer denken. Dit is een zinvol operationeel criterium, waarvan Penrose niet verwacht dat een computer er ooit aan zal

voldoen als de test goed wordt afgenomen. Hij accepteert de test dan ook, maar alleen bij gebrek aan iets beters. Dat betere zou moeten bestaan uit een goede theorie over wat denken en bewustzijn inhoudt, waaruit dan vanzelf volgt of computers daar wel of niet aan kunnen voldoen.

De aanhangers van 'strong AI' hebben al een theorie over denken: denken is het uitvoeren van een algoritme. Of beter gezegd: alle mentale activiteit - waaronder gevoel, begrip en bewustzijn - staat gelijk aan het uitvoeren van een algoritme. Kenmerk van een algoritme is, dat het door verschillende fysische systemen kan worden uitgevoerd. Daarom is de fysische realisatie onbelangrijk. Wat een brein kan, kan een computer in principe ook. Het enige dat we moeten doen is doorgronden wat voor algoritme het brein gebruikt.

Een bekend argument tegen deze positie is het argument van de Chinese Kamer van John Searle. De conclusie van dit gedachte-experiment is, dat het uitvoeren van een algoritme niet tot *begrip* leidt. Een computer heeft geen enkel besef van de *betekenis* van de symbolen die hij manipuleert. Penrose heeft een verklaring voor waar begrip bij mensen vandaan komt: denken is *meer* dan het uitvoeren van een algoritme. En als we willen begrijpen waarom menselijk denken verschilt van computers, moeten we kijken naar de onderliggende fysische systemen.

Algoritmes

Tot nu toe is het begrip 'algoritme' nog niet verder toegelicht. Het idee van een algemeen algoritme is pas in de 20^{e} eeuw ontstaan en is verbonden met het idee van een Turing machine. Het woord 'algoritme' is zelfs gedefinieerd als datgene wat door een 'mechanische procedure' uitgevoerd kan worden - precies wat een Turing machine kan. Die 'mechanische procedure' is overigens wel vrij abstract - bewegende onderdelen hebben er niets mee te maken. Interessanter is, dat men er achter kwam dat deze mechanische procedures precies hetzelfde doen als de operaties die in de formele wiskunde - met Church's lambda calculus - uitgedrukt kunnen worden. Begrippen als 'berekenbaar', 'algoritmisch' en 'recursief' betekenen dan ook allemaal ongeveer hetzelfde. Het blijft moeilijk om te omschrijven wat deze categorie nu precies *in*houdt; de termen hebben dan ook vooral betekenis vanwege het besef dat ze ook bepaalde dingen *uit*sluiten.

Een van de problemen met Turing machines is, dat ze soms niet stoppen maar oneindig lang door kunnen gaan. Op zich is dit begrijpelijk, maar het is erg vervelend dat niet goed te *voorspellen* is, wanneer dit het geval zal zijn. Dit staat bekend als het 'halting problem'. In een uitgebreide argumentatie - die ik onmogelijk kort kan samenvatten - laat Penrose zien waarom het 'halting problem' onoplosbaar is. Deze argumentatie loopt geheel parallel aan de Stelling van Gödel: in ieder (interessant) systeem zijn er *onbeslisbare* stellingen, oftewel stellingen waarvan de waarheid niet met behulp van het systeem kan worden beslist. Turing machines hebben, omdat ze algoritmische systemen zijn, onvermijdelijk enkele beperkingen. Wat voor Penrose echter nog veel belangrijker is, is

dat mensen, door *buiten* het systeem te denken, van sommige onbeslisbare stellingen wel de waarheid(swaarde) kunnen inzien. Mensen kunnen slimmer zijn dan algoritmes.

Op zoek naar een niet-computationeel systeem

Penrose heeft laten zien dat computers - als fysieke realisatie van Turing machines - onvermijdelijk beperkt zijn, en dat mensen daardoor altijd slimmer zullen zijn dan computers. Maar dat betekent dat er iets bijzonders moet zijn in het onderliggende fysische systeem - het brein - dat mensen in staat stelt op een niet-computationele manier te denken. Er moet iets niet-computationeels zijn aan het brein zelf. Wat voor mechanisme in de fysische wereld zou dat kunnen zijn?

In de klassieke natuurkunde is zo'n systeem niet te vinden. In het mechanistische wereldbeeld - een voorstelling waarbij alles bestaat uit massieve deeltjes die volgens duidelijke wetten bewegen en tegen elkaar botsen - is de situatie op ieder moment berekenbaar[1] als de situatie op één moment exact bekend is. Natuurlijk zal dit in de praktijk haast onmogelijk zijn: kleine onzekerheden in de beginsituatie kunnen grote gevolgen hebben, en om zo'n berekening uit te voeren moet je op een goede (exacte) manier met irrationele getallen omgaan. Maar in principe is zo'n systeem algoritmisch te beschrijven en kan zo'n systeem ook alleen maar dingen doen die algoritmisch beschreven kunnen worden; het kan dus geen bouwstenen leveren voor een niet-algoritmisch brein.

Zo komt Penrose uiteindelijk bij de quantummechanica uit. Deze wereld wijkt nogal af van het mechanistische 'biljartbalmodel'. Van een deeltje in quantumtoestand kun je niet meer zeggen dat het zich op een bepaalde plaats bevindt. De quantumtoestand is een soort combinatie van verschillende mogelijkheden die allemaal een bepaalde waarschijnlijkheid hebben. Dit geldt niet alleen voor de plaats; ook grootheden als moment en 'spin' zijn onderdeel van deze combinaties. Hier is echter niks *vaags* aan. Deze toestand is volgens Penrose een objectief, reëel iets waarvan de ontwikkeling nauwkeurig door wiskundige formules beschreven kan worden. Zolang een systeem in de quantumtoestand is, is het volledig berekenbaar.

Iets merkwaardigs gebeurt op het moment dat onderzoekers aan zo'n quantumsysteem iets proberen te meten. Bij zo'n 'waarneming' springt het systeem namelijk in een van de mogelijke toestanden van de complexe quantumtoestand. Welke toestand dat wordt, hangt af van de waarschijnlijkheden van die mogelijkheden, maar is verder niet te voorspellen. Dit springen heet 'reductie' (en heeft niets te maken met het reduceren van wetten en theorieën). Na een 'waarneming' is het systeem *niet* meer in de quantumtoestand met verschillende waarschijnlijkheden, maar is de

[1] 'berekenbaar' is niet hetzelfde als 'gedetermineerd'. Het begrip 'determinisme' levert problemen op met causaliteit en relativiteit, zodat Penrose het zoveel mogelijk vermijdt. Een probleem kan wel gedetermineerd zijn maar niet berekenbaar, zoals bij de vraag "Zullen A en B *ooit* tegen elkaar botsen?"

waargenomen toestand de *enige* toestand geworden. Omdat het ‘springen’ een kansproces is, is de toestand van een quantumsysteem na een sprong *niet* berekenbaar.

Een opmerkelijk verschijnsel in de quantummechanica is *beïnvloeding op afstand*. Een quantumsysteem hoeft namelijk niet ruimtelijk beperkt te zijn. Een systeem kan ook bestaan uit twee tweelingdeeltjes die ooit samen zijn ontstaan en vervolgens allebei een andere kant op zijn bewogen. Een meting aan een van die deeltjes heeft ook gevolgen voor de toestand van het andere deeltje. Deze beïnvloeding is problematisch, omdat ze in strijd is met de relativiteitstheorie, die zegt dat informatie niet sneller kan dan het licht.

Een onopgelost probleem in de quantummechanische is de vraag *wanneer* reductie optreedt. Waarom merken we wel quantumverschijnselen op de schaal van elektronen en fotonen, maar niet op het niveau van golfballen en katten? Oftewel, waarom zien we in het dagelijks leven om ons heen geen objecten die op twee plaatsen tegelijk zijn? Het staat vast dat reductie, waarbij een quantumtoestand gereduceerd wordt tot een ‘gewone’, klassieke toestand, plaatsvindt zodra onderzoekers proberen iets interessants aan die quantumtoestand te meten; maar dat geeft weinig duidelijkheid over de precieze oorzaak van deze sprong.

Een toekomstige theorie van quantummechanica zal nog een ander verschijnsel moeten verklaren: materie en zwaartekracht. Veel onderzoekers verwachten daarom, dat voor zo’n toekomstige theorie van ‘quantum gravity’ de relativiteitstheorie op een of andere manier aangepast zal moeten worden aan de huidige quantummechanica. Penrose beweert echter dat het juist de huidige quantummechanica is die volledig op de schop zal gaan. Hij heeft al een idee voor zo’n toekomstige theorie, dat in een keer een verbinding legt met de zwaartekracht en het schaalprobleem (de vraag wanneer er reductie optreedt) oplost: reductie komt doordat de effecten van de verschillende quantumtoestanden zo ver worden uitvergroot dat het verschil van die effecten groter wordt dan één graviton (een zwaartekrachtsdeeltje). Dit uitvergroten is noodzakelijk als we waarnemingen willen doen, maar het feit dat er een *bewuste* waarnemer aanwezig is, maakt voor het quantumsysteem niet uit.

Quantummechanica en breinen

Als de quantummechanica niet-berekenbaar is, is het dan mogelijk dat de quantummechanica iets bijdraagt aan de specifieke niet-algoritmische mogelijkheden van ons brein? Om deze vraag te beantwoorden, geeft Penrose eerst een overzicht van de neurowetenschap. Hij noemt een aantal punten waar mogelijk de quantummechanica de werking van het brein zou kunnen beïnvloeden.

Ten eerste is van netvliesneuronen bekend, dat één foton (lichtdeeltje) genoeg is om ze te laten vuren (een signaal afgeven aan andere neuronen). Een foton is typisch een deeltje dat zich in een quantumtoestand kan bevinden. Maar dat betekent dat een foton *met een bepaalde waarschijnlijkheid* een netvliescel kan treffen, terwijl het tegelijkertijd met een bepaalde waarschijnlijkheid in een

toestand is waarin het die cel niet treft. Betekent dit dat het neuron ook in een quantumtoestand is waarin het zowel wel als niet aan het vuren is? En hoe zit het met al die andere neuronen die hierdoor ook gaan vuren (of niet)? Hier biedt Penrose's '1-graviton-theorie' uitkomst: als een neuron gaat vuren, worden hierdoor zoveel deeltjes verstoord, dat de grens van één graviton overschreden wordt. Volgens Penrose's theorie moet er op zo'n moment reductie plaatsvinden en kan een compleet neuron zich niet in een quantumtoestand bevinden. Op het niveau van neuronen kan er geen sprake zijn van quantumverschijnselen, maar neuronen kunnen wel door quantumverschijnselen gestimuleerd worden.

Ook is er het concept van *quantum computers* van Deutsch. Uiteraard zijn zulk soort computers nog toekomstmuziek, maar het concept is al uitgedacht. In tegenstelling tot Turing machines, die alle handelingen serieel uitvoeren, maken deze computers gebruik van parallellisme: het verschijnsel dat in de quantummechanica verschillende mogelijke toestanden tegelijkertijd evolueren. Hierdoor kan een quantum computer veel sneller berekeningen uitvoeren dan een gewone computer. Quantum computers zoals Deutsch ze heeft bedacht, hebben echter uiteindelijk precies dezelfde mogelijkheden en beperkingen als Turing machines.

Tot slot oppert Penrose dat quantumverschijnselen wellicht een cruciale rol kunnen spelen bij de groei van synapsen. Synapsen zijn de verbindingen tussen neuronen, en de flexibiliteit waarmee deze synapsen kunnen groeien en veranderen is waarschijnlijk wat mensen in staat stelt om te leren en herinneringen op te slaan. Volgens Penrose moet hier, net als bij de quasi-kristallen die hij bestudeerd heeft, het niet-lokale karakter van quantumsystemen (oftewel de mogelijkheid van beïnvloeding op afstand) een rol spelen. Als ieder atoom alleen op basis van lokale informatie een bepaalde positie 'kiest', kunnen zulke structuren niet ontstaan. Alleen als een veel groter systeem van vele atomen verschillende quantumposities kan 'uitproberen', kan een structuur gevormd worden die op macroniveau optimaal is.

Een theorie over bewustzijn

In het laatste hoofdstuk van het boek geeft Penrose een aantal ideeën over hoe een succesvolle theorie over bewustzijn, een 'physics of mind', er uit zou moeten zien. Echte intelligentie kan volgens hem niet zonder bewustzijn. Verder moet bewustzijn, vanwege de evolutietheorie, een functie hebben en is het niet alleen maar een passieve toeschouwer. Voor het uitvoeren van algoritmes is geen bewustzijn nodig, maar voor het selecteren en evalueren van algoritmes wel. Algoritmes zelf zeggen namelijk helemaal niks over *waarheid*, want je kunt net zo makkelijk algoritmes bedenken die onwaarheden opleveren als algoritmes die waarheden opleveren. Het uiteindelijke oordeel over waarheid kan niet door algoritmes zelf worden gemaakt, en moet dus gebeuren door het bewuste denken. Ook *begrip*, het geven van betekenis aan de symbolen die in algoritmes worden gemanipuleerd, is het terrein van het bewustzijn. Als mensen zich een algoritme eenmaal eigen hebben gemaakt, kunnen ze dit echter vaak onbewust uitvoeren. Soms is de uitkomst van een waarheidsoordeel vervolgens algoritmisch te

beschrijven en kan zo'n beslissing later automatisch en onbewust gemaakt worden. Het is echter het bewustzijn dat zo'n algoritme ontdekt en beoordeelt.

Omdat bewust denken *over* algoritmes gaat, kan het niet zelf een algoritme zijn. Als dat wel zo was, zou dat algoritme immers ook z'n eigen onbeslisbare Gödelstellingen hebben, en het is nou juist een kenmerk van menselijke intelligentie dat we zulke problemen op niet-algoritmische manier kunnen oplossen.

En hoe zit het met de quantummechanica? Penrose trekt enkele parallellen tussen quantummechanica en bewustzijn. Zo verbaast hij zich over de eenheid van het bewustzijn. Waarom ervaren we maar één bewustzijn, terwijl er zich in ons brein allemaal processen parallel afspelen? Dit lijkt op de quantumtoestand, die immers bestaat uit meerdere mogelijke toestanden die zich parallel ontwikkelen, maar uiteindelijk ook weer één toestand is. Een verschijnsel dat lijkt op de eenheid van het bewustzijn, is de 'globality' van ideeën: dat we een hele complexe gedachte in één moment kunnen overzien en begrijpen. Voorbeelden van zo'n complexe gedachte zijn een muziekstuk of de opzet van een wiskundig bewijs. Ook dit komt overeen met de quantumtoestand, die immers een combinatie van een groot aantal verschillende toestanden is. De suggestie is dus dat deze verschijnselen rechtstreeks door een of ander quantummechanisme worden veroorzaakt. Maar de belangrijkste reden waarom de quantummechanica een rol zou moeten spelen in een theorie over bewustzijn, blijft natuurlijk, dat bewustzijn een niet-algoritmisch systeem nodig heeft om mogelijk te zijn.

Commentaar

Aan het begin van deze recensie heb ik aangegeven dat ik vooral aandacht wilde besteden aan drie aggregatieniveaus: de quantummechanica, de neurowetenschap en de cognitieve psychologie. Opvallend is, dat voor Penrose het verschil tussen neurowetenschap en psychologie niet heel belangrijk is. In mijn samenvatting heb ik deze disciplines zoveel mogelijk gescheiden, maar in de laatste twee hoofdstukken van het boek zelf wisselen deze onderwerpen elkaar steeds af. Penrose's stelling aan het begin van het boek is: we kunnen het bewustzijn / het denken niet begrijpen zonder een goed begrip van het onderliggende fysische systeem. Omdat bewustzijn niet-algoritmisch is, moeten quantummechanische verschijnselen in dit systeem een cruciale rol spelen. Hoewel hij enkele manieren noemt waarop quantummechanische verschijnselen van invloed zouden kunnen zijn op neuronen, vertelt hij niet wat voor rol dit neurowetenschappelijke niveau zou moeten spelen in een theorie over bewustzijn. Dit is jammer, aangezien hij wel duidelijk maakt dat (eventuele) quantummechanische verschijnselen zich op of onder, maar *niet* boven dit niveau zullen moeten manifesteren: een compleet neuron is immers al te groot om nog in een quantummechanische toestand te kunnen zijn.

Uiteraard is een verklaring in termen van een onderliggend aggregatieniveau een uiterst krachtige verklaring. Wat dat betreft kan ik Penrose begrijpen als hij vindt dat onze kennis van het bewustzijn

niet volmaakt is, zolang we het niet kunnen verklaren in termen van het allerlaagste fysische niveau - de quantummechanica. Maar ik denk niet dat we, totdat we een volmaakte theorie van 'quantum gravity' hebben, geen zinnige vooruitgang in ons begrip van het bewustzijn kunnen boeken. Het lijkt me dat men vooral moet onderzoeken of en hoe quantummechanische verschijnselen het niveau van de neurowetenschap beïnvloeden. Intussen kan men *ook* zinvolle dingen onderzoeken over de relatie tussen de neurowetenschappen en de psychologie. Sterker: als onze beschrijving van het neurowetenschappelijke niveau klopt - ook al kunnen we misschien niet alle verschijnselen verklaren - dan kunnen we daarmee uitstekend het bewustzijn proberen te verklaren. Latere ontwikkelingen in de quantummechanica zullen weliswaar onze verklaringen van de neurowetenschappen veranderen, maar dat hoeft geen consequenties te hebben voor onze theorie van het bewustzijn.

Het 'tussenniveau' van de neurowetenschap lijkt me dan ook heel zinnig, en ik denk dat Penrose onderschat hoeveel we nog kunnen leren zonder op een alomvattende 'quantum gravity'-theorie te wachten. Het direct proberen te verbinden van psychologie met quantummechanica is ongelooflijk moeilijk en leidt volgens mij niet tot nuttige, nauwkeurige onderzoeksvragen. Penrose's ideeën over quantummechanica en de eenheid van het bewustzijn, kunnen op geen enkele zinvolle manier onderzocht worden. Zijn ideeën over de invloed van quantummechanica op neuronen en synapsen zijn wat dat betreft een stuk interessanter.

Hoe zal een toekomstige theorie van bewustzijn eruit moeten zien? Als we dat wisten, waren we al een heel eind op weg! Penrose claimt nergens dat hij een doortimmerde theorie presenteert, en het valt hem dan ook niet kwalijk te nemen dat het op dit punt een beetje bij losse ideeën en suggesties blijft. Toch valt het me een beetje tegen. Penrose is vooral *verwonderd* over de kracht van het menselijk denken / bewustzijn, en probeert aannemelijk te maken waarom zoiets überhaupt mogelijk zou zijn. Ik krijg daardoor een beetje de indruk dat hij vooral probeert 'bewustzijn' intact te houden. Het gedachte-experiment van de Chinese Kamer roept de vraag op: "Wie of wat *begrijpt*¸ wie of wat is *bewust*?" Deze vraag geldt niet alleen voor computers, maar ook voor breinen. Penrose biedt een mogelijkheid: een of ander quantummechanisch mechanisme maakt begrip en bewustzijn mogelijk. Dit komt op mij over als een manier om *iets* (een mechanisme) aan te wijzen dat bewust is. Ik vraag me echter af, of in een uiteindelijke theorie van het denken er werkelijk zoiets aangewezen zal kunnen worden. Hoewel ik het moeilijk vind om me voor te stellen dat zo'n begrip compleet uiteen zou vallen, lijkt het me niet onwaarschijnlijk. Andere begrippen 'verdwijnen' ook vaak op een lager aggregatieniveau. Het bekendste voorbeeld in deze situatie is het begrip 'zien'. Een vrij naïeve verklaring van 'zien' luidt als volgt: "Lichtstralen vallen op het netvlies, waar een beeld ontstaat van de wereld. Dit beeld wordt door de oogzenuw getransporteerd naar de occipitaalkwab (in het achterhoofd). Daar wordt opnieuw een beeld gevormd, en dit wordt door iets in het brein 'gezien'." Dit is echter helemaal geen bevredigende verklaring. We hebben het probleem verplaatst, maar we kunnen niet uitleggen *wat* of *wie* er nou eigenlijk ziet. Als we proberen het verschijnsel *zien* goed te verklaren, vinden we nergens een

homunculus die ziet, maar vinden we alleen neuronen die gevoelig zijn voor een heel specifieke gebeurtenis in een specifiek deel van het gezichtsveld. Er is niet iemand aan te wijzen dat het hele beeld begrijpt. Ik stel me voor dat bij een uiteindelijke verklaring van bewustzijn hetzelfde gebeurt.

Penrose's argument dat bewust denken een niet-algoritmisch proces moet zijn, is heel sterk. De vraag die overblijft, is dan: wat is het dan wel? Als AI'er vind ik het moeilijk om me een nauwkeurige, bevredigende beschrijving voor te stellen die niet (algoritmisch) implementeerbaar zou zijn. Maar dat zou betekenen dat een goede beschrijving onmogelijk is, juist vanwege het niet-algoritmische karakter van bewustzijn. Een belangrijkere vraag vind ik echter: wat hebben we bereikt met het introduceren van een niet-algoritmisch systeem? Natuurlijk hebben we een of ander niet-algoritmisch systeem *nodig*, maar hoe kan een *toevalsproces* (quantummechanische reductie) ons in staat stellen dat te doen wat we met algoritmes niet kunnen, namelijk het vormen van een waarheidsoordeel? Hoe kunnen we *toeval* op een zinnige manier gebruiken, als we er geen controle over hebben? Met dit soort vragen kom je eigenlijk automatisch terecht bij het vraagstuk van determinisme en vrije wil, en het valt Penrose natuurlijk niet kwalijk te nemen dat hij dat niet voor me heeft kunnen oplossen. Aan de andere kant: misschien valt het allemaal wel mee, en kunnen we inderdaad door de introductie van toeval onze algoritmische systemen verbeteren. Er wordt op het moment met dergelijke probabilistische 'algoritmen' geëxperimenteerd. Hoewel computers geen toeval kunnen genereren, kunnen ze het wel onttrekken aan de buitenwereld (bijvoorbeeld een klok, het Internet of een radioactieve bron). Dat zou wel eens heel interessant kunnen zijn.

Tot slot wil ik nog iets zeggen over de manier waarop Penrose theoretiseert. Ik denk dat hij er compleet naast zit als hij parallellen trekt tussen de quantummechanica en de manier waarop hij bewustzijn ervaart. Maar ik denk niet dat het leggen van dit soort verbanden tussen disciplines altijd een zinloze bezigheid is. De manier waarop wiskundige ontdekkingen van belang kunnen zijn in de natuurkunde is verbazingwekkend. Zijn hoofdstukken over wiskunde, natuurkunde (en quantummechanica) en kosmologie zijn dan ook in zichzelf al interessant en de moeite van het lezen waard. Bij het ontdekken van fundamentele, 'allesomvattende' theorieën denk ik dat generalistische wiskundigen nog een belangrijke rol kunnen spelen. Maar op het complexe niveau van biologische wezens hebben we andere soorten theorieën nodig...