

De geest ontzenuwd

Bert G. J. Frederiks

23 juli 2001

Om uit te kunnen leggen wat bewustzijn is, wil ik in leken-taal de principes van enkele van de belangrijkste neurale netwerken uitleggen, te weten “error backpropagation” (backprop), welke door publicatieproblemen 4 keer opnieuw uitgevonden moest worden, en “Adaptive Resonance Theory“ (ART), ruim 30 jaar oud, maar nu pas op zijn waarde geschat. Oud nieuws dus, maar het mag desondanks toch eindelijk wel eens in de krant. De stap van zenuwcellen naar de geest en vice versa is vol valkuilen, maar neurale netwerken zijn erg saai zonder een blik in die richting.

Wat is een neuraal netwerk?

Een neuraal netwerk is een netwerk van zenuwcellen, ook wel neuronen genaamd, zoals die in onze hersenen zitten. Neuronen hebben veel ingaande signalen en (gewoonlijk) een uitgaand signaal, dat naar veel andere neuronen gaat. De cellen in onze cerebrale cortex hebben ieder gemiddeld zo'n 10.000 ingangen. Kunstmatige modellen hebben er meestal slechts een handje vol. Cellen kunnen gekoppeld zijn aan elkaar, aan sensoren (ogen, oren, huid, enz.) of aan spieren en klieren.

Via zijn invoer kan een neuron geprikkeld worden. Indien een neuron over een bepaald niveau komt qua prikkeling dan wordt hij geactiveerd, waarmee hij zelf andere neuronen kan gaan prikkelen. Belangrijk hierbij is, wiskundig gezien, dat proces dit niet lineair is. Dat wil zoiets zeggen als dat de prikkeling een drempel moet nemen. Over het algemeen moet een behoorlijk deel van de ingangen flink geprikkeld worden wil een neuron zelf geactiveerd raken.

Een neuraal netwerk leert uit zichzelf

Bijzonder is verder dat de gevoeligheid van een neuron aangepast kan worden door het neuron zelf, waardoor een neuron zichzelf dingen kan leren. Het algemene principe daarbij is dat indien een neuron geactiveerd had moeten worden en ook inderdaad geactiveerd werd, dan worden alle invoer-kanalen die daartoe positief hebben bijgedragen iets gevoeliger, en de anderen juist niet, enz. Okee, zult u zeggen, maar hoe weet een neuron nu of hij geactiveerd had moeten zijn? En indien hij dat weet, waarom moet hij dat dan nog leren? Het simpele antwoord is dat een neuron dit leert voor die momenten waarop hem dit niet verteld wordt. Met andere woorden, er kan een onderscheid gemaakt worden tussen leren en het tonen van het geleerde. In de leerfase spreekt men van een

leer-invoer (teaching-input), die op de uitvoer wordt gezet. Door leer-invoer plus gewone invoer “weet” een neuron hoe zichzelf aan te passen.

U kunt zich voorstellen dat, wanneer een neuron op zichzelf al de neiging heeft om bij een bepaalde invoer een bepaalde uitvoer te genereren, een netwerk van neuronen als geheel de neiging heeft om bij een bepaald ingevoerd patroon een bepaald patroon als uitvoer te genereren, met andere woorden om patronen met elkaar te associëren. Patroon-associatie, dat wil zeggen het automatisch aan elkaar koppelen van tegelijkertijd voorkomende patronen, is inderdaad één van de dingen waartoe een neurale netwerk geconstrueerd kan worden.

Mijn favoriete voorbeeld van een nuttig gebruik hiervan is dat van een olie-maatschappij die naar olie zoekt. Dat doet men door explosieven onder de grond tot ontploffing te brengen om vervolgens de terugkaatstende geluidsgolven met microfoons te registreren. Deze geluiden worden door experts geanalyseerd en aan de hand daarvan wordt bepaald of de kans dat er olie aanwezig is groot genoeg is om te gaan boren. Dit is een ingewikkelde analyse, waarbij intuïtie, of tenminste een natte vinger, niet gemist kan worden. Dit is een analyse die ook heel goed door een neurale netwerk gedaan kan worden.

Maak hiertoe een neurale netwerk bestaande uit lagen. Onderaan bevindt zich een invoerlaag met neuronen die gekoppeld zijn aan de signalen zoals geregistreerd door de microfoons. Dan een aantal tussenliggende lagen (die in het jargon “hidden”, dat is “verborgen”, worden genoemd), en tenslotte een uitvoerende laag. Activerings- of prikkelingssignalen lopen van invoerlaag, door de tussenliggende lagen, naar de uitvoerlaag.

In ons geval zal de uitvoerende laag uit slechts twee neuronen bestaan. Activering van het ene neuron betekent: “Ja, er zit olie in de grond”, en activering van de andere betekent: “Nee, er zit geen olie in de grond”. We trainen het neurale netwerk door het oude geluidsopnamen uit het archief te tonen, waarvan we door boring of door experts weten of er olie in de grond zat of niet. Daarbij vertellen we het netwerk steeds of er olie in de grond zat of niet. Aldus zal het neurale netwerk bepaalde patronen associëren met “Ja, er zit olie in de grond”, en anderen juist niet. Voor de meeste neurale netwerken geldt bovendien dat ze dit zodanig leren dat ze ook van nieuw aangeboden geluidspatronen meer dan toevallig kunnen bepalen of er olie in de grond zit—daarover dadelijk.

Het aanpassen of leren van zo’n neurale netwerk als geheel is natuurlijk een stuk ingewikkelder dan dat van een enkel neuron, maar het principe is gelijk. Aan uitvoerzijde verandert er niets. Daar vertellen we het neuron gewoon of het geactiveerd had moeten zijn of niet. Maar dan kan dit neuron weer aan ieder neuron in de laag daaronder vertellen of het geactiveerd had moeten zijn of niet, enzovoorts. Dit mechanisme heet “error-backpropagation”, of kortweg “backprop”, dat is “fout-terugkoppeling”. Het is door een aantal mensen steeds opnieuw uitgevonden—de geschiedenis van neurale netwerk theorieën kenmerkt zich ook buiten deze krant door publiciteitsproblemen. Paul Werbos was waarschijnlijk de eerste, maar het is beroemd geworden door de PDP-groep.

Merk op dat bij vrijwel elke associatie het complete netwerk betrokken is omdat alle neuronen elkaar beïnvloeden. Maar juist omdat ze ieder op zich slechts een klein beetje bijdragen, weet een neurale netwerk zich ook nog vrij veel te herinneren wanneer een deel van zijn neuronen sterft. De kennis is dus gedistribueerd. Het is gedistribueerd in de gevoeligheden tussen neuronen.

Een probleem met backprop is dat het in biologische systemen niet in deze vorm bestaat. Een ander nadeel is dat dit soort netwerken vrij langzaam leren.

Je moet ze de te associëren patronen honderden keren laten zien, het liefst door elkaar, zodat het netwerk de kans krijgen om al deze patronen zo ideaal mogelijk in zichzelf op te slaan. Een bijkomend nadeel is aldus dat dit soort netwerken oude associaties kunnen vergeten door nieuwe te leren.

Een neurale netwerk abstraheert uit zichzelf

Behalve voor patroon-associatie kunnen neurale netwerken ook worden gebruikt voor het abstraheren van eigenschappen uit patronen (feature abstraction). Anders dan hierboven maken we een neurale netwerk hiertoe zodanig dat we het niet middels een leer-invoer hoeven te vertellen wat de te abstraheren eigenschappen zijn. Die ontdekt het juist zelf.

We doen dit door in het neurale netwerk lagen neuronen aan te brengen die elkaar trachten uit te schakelen. Met andere woorden, in plaats van elkaar te activeren, proberen deze neuronen elkaar uit te doven. Een neuron welke uitgedoofd is kan geen ander neuron meer uitdoven. In het sterkste geval blijft er één winnaar over. Deze winnaar zorgt dan weer voor de leer-invoer voor onderliggende lagen neuronen. Van invoerlaag naar uitvoerlaag is dit netwerk niet anders dan het eerder genoemde neurale netwerk, alleen dwars op deze activeringsrichting proberen neuronen elkaar uit te schakelen.

Het gevolg van deze neuronenschakeling is dat, indien het aantal mogelijke winnaars kleiner is dan het aantal aangeboden patronen, het netwerk deze patronen op den duur zal gaan groeperen op te onderscheiden kenmerken. Ergo, deze winnende neuronen gaan ieder staan voor bepaalde eigenschappen van getoonde patronen.

Hoe werkt dit? Neem, om het simpel te houden, een netwerk dat alleen in de bovenste laag winnaar-neem-alles neuronen heeft. Stel we laten dit neurale netwerk afwisselend twee patronen zien, A en B . Beginnen we met patroon A , dan zal er in de bovenste laag een winnaar zijn. Welke dat is, is de eerste keer puur toeval, maar er kan er maar één winnen. Noem dit neuron a . Volgens het backprop-principe levert uitvoer-neuron a vervolgens een leer-invoer aan onderliggende lagen. Aldus leert het neurale netwerk patroon A te associëren met uitvoer-neuron a . Indien patroon A voldoende verschilt van patroon B , dan zal dit vanzelf juist niet uitvoer-neuron a neigen te activeren, maar wel een ander uitvoer-neuron, welke we b zouden kunnen noemen. Merk op dat we hier een leer-invoer creëren zonder leraar. Overigens, in dit simpele geval zou het nog best gemakkelijk kunnen gebeuren dat zowel patroon A als patroon B een zelfde uitvoer-neuron activeren, maar dat kan met allerlei trucs worden ondervangen.

Beschouw nu de situatie waarbij er meer patronen zijn dan er uitschakel-neuronen zijn. Het is duidelijk dat dan niet alle patronen één op één aan een uitvoer-neuron gekoppeld kunnen zijn. Uitvoer-neuronen kunnen echter wel gekoppeld zijn aan overeenkomende delen van patronen, bijvoorbeeld een ronde vorm of een bepaalde kleur, met andere woorden (algemene) kenmerken of eigenschappen van deze patronen. Zoiets zal vanzelf ontstaan, want niet alleen kan er slechts één neuron winnen, er zal er altijd één winnen, en indien we alle patronen aan het netwerk blijven tonen dan zal het zich blijven aanpassen totdat alles het beste past.

Indien we toestaan dat meer dan één neuron wint, dan kunnen ook bepaalde combinaties van uitvoer-neuronen geactiveerd raken, welke tezamen een soort

analyse of deconstructie van het ingevoerde patroon vormen; waar je als mens ook naar kijkt, het valt onmiddellijk uiteen in een aantal delen.

Het principe van onderlinge competitie van neuronen kan in meerdere of mindere mate in ieder neuraal netwerk worden ingebouwd. Het maakt netwerken bijna vanzelf een beetje “intelligenter”. Het zorgt dat een neuraal netwerk zichzelf structureert door belangrijke dingen te ordenen en te onthouden en onbelangrijke dingen wat sneller te vergeten. Het behelst abstractie in heel abstracte zin—ik noem het zelf graag immanicering, met als tegendeel van het immanente het transcendente. Het helpt bijvoorbeeld ook bij het eerder genoemde voorbeeld van het analyseren van geluidsgolven om te bepalen of er ergens olie in de grond zit. In onze hersenen zitten velerlei neuronen die dergelijke functies toegegaan lijken te zijn.

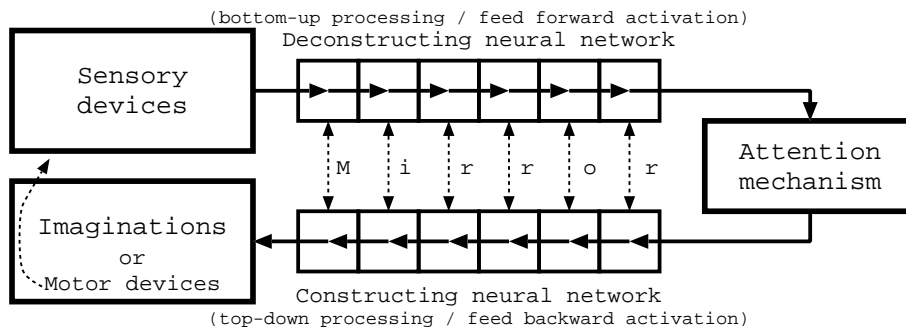
Een biologisch meer waarschijnlijk model: aandacht en spiegelen

Een probleem met backprop is dat fout-terugkoppeling in onze hersenen niet op die manier kan werken. Stephen Grossberg bedacht ruim 30 jaar geleden al een oplossing die dit probleem niet heeft. Hij noemt zijn theorie “Adaptive Resonance Theory”, of kortweg “ART”, waarbij hij resonantie nadrukkelijk verbindt met bewustzijn. In totaal andere woorden dan hij gebruikt wil ik dat hier trachten uit te leggen. Ik heb daarbij bovendien een beeld van de menselijke hersenen voor ogen. ART zoals Grossberg het beschrijft is logisch en wiskundig waterdicht, maar voor mij te gedetailleerd en ingewikkeld om een paar belangrijke grote lijnen te zien. Het feit dat Grossberg wiskunde en bewustzijn in één wetenschappelijke theorie bij elkaar heeft gebracht blijft magnifiek, wat je daar verder ook precies van denkt. Dat hij het zo lang geleden al opschreef en gebouwd heeft, maar nu pas erkenning krijgt is, vind ik, triest en geeft mijns inziens ook aan hoe het in de praktijk van alle dag met de wetenschap gesteld is als het gaat om echt nieuwe ontwikkelingen.

Neem twee neurale netwerken, ieder bestaande uit een zevental lagen, en leg deze achterstevoren op elkaar, dus zodanig dat de uitvoer van het ene netwerk op de invoer van het andere ligt. Laag 7 ligt aldus op laag 1, laag 6 op laag 2, enz. Neuronen in deze netwerken zijn op een speciale, spiegelende manier, één op één met elkaar verbonden en wel zo dat ze elkaars gevoeligheid verhogen zonder dat ze elkaar echt kunnen activeren. Eigenlijk vormen een neuron uit de ene laag en een neuron uit de andere laag samen één neuron (in de uitleg van Stephen Grossberg is het precies dat). Het uiteindelijke resultaat moet zijn dat de twee netwerken ertoe neigen om elkaar te spiegelen, maar dan wel met een neurale activering die ten opzichte van elkaar in tegengestelde richting loopt.

Het idee is dat je via het ene netwerk dingen kunt waarnemen, abstraheren en in delen opsplitsen (deconstructie), terwijl je je via het andere netwerk dingen kunt voorstellen en construeren. Daartoe koppelen we de invoer van het eerste netwerk aan onze zintuigen. Dit netwerk noem ik het deconstructieve netwerk. De uitvoer van dit netwerk koppel ik aan een winner-neem-bijna-alles neurale laag van elkaar uitschakelende neuronen.

Dit winner-neem-alles netwerk functioneert feitelijk als een aandachtsmechanisme. Dit neurale aandachtsmechanisme doet niet meer dan dat het een klein



Figuur 1: Een hiërarchisch neurale netwerk gesplitst in twee delen

aantal winnaars toelaat in de uitvoer van het deconstructieve netwerk.

De winnaars van het aandachtsmechanisme leveren de input voor het tweede neurale netwerk, welke ik het constructieve netwerk noem. Vanuit het aandachtsmechanisme worden steeds enkele neuronen in het constructieve netwerk geactiveerd. Van daaruit vloeit neurale activering terug richting de zintuigen, iets wat wij bij afwezigheid van waarneming ervaren als een mentale voorstelling of beeld van iets.

Aandachtsmechanismes bevinden zich overigens in elke laag, maar in lagere lagen is er niet een enkel aandachtsmechanisme als wel vele honderden tot honderdduizenden, die allemaal een klein gebiedje in het neurale netwerk bestrijken. Hun functie is de eerder genoemde abstrahering. Stelt u zich voor: U kijkt naar een schilderij. Dat is een complex patroon. Dat komt via uw ogen uw neurale netwerk binnen. Elkaar uitschakelende neuronen zorgen voor abstrahering. In de eerste laag worden lijnen, kleuren en beweging gedetecteerd, hogerop worden steeds complexere vormen gedetecteerd. Uiteindelijk laat het aandachtsmechanisme slechts vijf winnaars toe. Vanaf deze vijf winnaars vloeit neurale activering door het constructieve netwerk terug richting de ogen.

Omdat het deconstructieve en het constructieve netwerk elkaar neigen te spiegelen, zal deze terugvloeiende activering ertoe neigen om een (re)constructie te zijn van het waargenomen. Wanneer dit lukt is er sprake van wat Grossberg “resonantie” noemt. Resonantie is een relatief stabiele situatie van het neurale netwerk die, gegeven de eigenschappen van neuronen, vanzelf tot leren leidt om de eenvoudige reden dat neuronen elkaar in deze situatie relatief langdurig prikkelen of juist niet prikkelen. Bij een embryo zal de werkelijke activeringsrichting nog vrij willekeurig zijn, en de resonantie gering, maar ook daar wordt de activering zodanig gestuurd en in banen geleid dat de twee netwerken elkaar *neigen* te spiegelen. Aldus vormen we ons een voorstelling bij wat we zien. Met andere woorden, op deze wijze trekt iets onze aandacht. Andersom geldt dat, omdat het constructieve netwerk ook het deconstructieve netwerk beïnvloed, onze aandacht zich ook neigt te richten op dingen die met onze gedachten samenhangen. We hebben dan als het ware eerst een aandachtspatroon en van daaruit worden, door spiegeling, automatisch keuzes in het waargenomen gemaakt. Langs beide wegen, maar vooral langs de laatste weg, zal een aandachtspunt tot steeds weer een ander aandachtspunt leiden. Dit wordt mede geholpen doordat specifieke aandacht van nature snel is uitgeput. Zie hier een basis voor ‘denken’ die dieren

en mensen met elkaar delen.

Merk op dat de vrij ‘transcendente’ informatie zoals die op het netvlies van mijn ogen valt uiteindelijk wordt omgezet in de activering van slechts enkele neuronen, die samen een soort aandachtspatroon vormen. Aandacht is dus eigenlijk het summum van abstrahering. De spiegeling tussen het constructieve en het deconstructieve netwerk is bovendien een biologisch mogelijke manier van fout-terugkoppeling (backprop). Spiegeling houdt namelijk in feite in dat elk verschil in activering tussen het ene en het andere netwerk een ‘fout’ is, welke aan beide zijden op cel-niveau gedetecteerd en gecorrigeerd wordt, geheel volgens mechanistische principes.

Alhoewel mijn uitleg totaal afwijkt van die van Stephen Grossberg, is dit mijns inziens grotendeels toch een samenvatting van zijn theorie, aangevuld met ideeën van mijzelf. Grossberg heeft onder de noemer van “Adaptive Resonance Theory” (ART) bewezen dat deze neurale netwerken goed werken, veel sneller leren dan backprop netwerken en niet lijden onder massaal geheugenverlies.

Tijdelijkheid van verbeelding en bewustzijn

Ik ben het op één punt niet helemaal met Stephen Grossberg eens. Grossberg koppelt bewustzijn aan resonantie. Volgens mij bestaat er geen bewustzijn zonder tijdelijkheid.

Aangaande ons voorstellingsvermogen spelen veel zaken een rol. In het bijzonder is tijd enorm belangrijk. We stellen ons dingen voor door ze in de tijd te verbeelden. Onze aandacht schuift van het ene naar het andere en we doen dit niet zonder structuur. Deze structuur is aldus tenminste gedeeltelijk op een niet tijdelijke manier vastgelegd. Aandacht speelt hierin een rol, omdat in het aandachtspatroon opeenvolgende zaken met elkaar kunnen worden geassocieerd. Maar speciale plannings- en taalmechanismen zijn minstens zo belangrijk, zeker voor mensen. Al met al zorgen de middelste en hogere neurale lagen voor samenhang in onze mentale beelden. Ze vormen een “voor”-bewustzijn. Ons eigenlijke bewustzijn bestaat alleen in de tijd. Dat wij de illusie hebben ons van meer bewust te zijn komt eenvoudig hierdoor dat wij ons niet bewust kunnen zijn van iets waarvan wij ons niet bewust zijn. Wij zijn ons pas bewust van de kleur van de ogen van een eekhoorn op het moment dat wij ons bewust zijn van de kleur van de ogen van een eekhoorn.

Een aardig voorbeeld en tevens een metafoor van tijdelijke beelden, en ooit de reden en aanleiding voor mij om dit alles te ontdekken, is de werking van een film of televisieprogramma. Kijkend naar een film ontstaat elke 2 tot 5 seconden een nieuw aandachtspatroon—u heeft misschien de illusie meer te zien, maar knijp uw ogen dicht en tel wat u zag. Op verschillende momenten in de tijd worden dergelijke patronen in het hier en nu met elkaar verbonden, bijvoorbeeld wanneer een bepaald persoon, of een zelfde huis, enz. opnieuw onze aandacht trekt. Een film is een beeld in de tijd, oftewel een tijdelijk beeld, maar door dit soort associaties van terugkerende elementen vormt zich ook een meer permanent beeld van waaruit we ons dingen kunnen herinneren. Indien associaties en betekenissen voldoende verweven zijn ontwikkelen ook lagere neurale lagen associatie-patronen. Ons bewustzijn is een beetje als zo’n film. Ons bewustzijn is een beeld (een beweten-zijn) in de tijd.

Een belangrijke vorm van spiegeling die ik, in weliswaar heel andere be-

woordingen, heb genoemd is die welke in de semiotiek, semiologie en linguïstiek “referentie” of “verwijzing” wordt genoemd. Zo verwijst het woord “koe” naar een koe. In mijn uitleg verwijst de activering van een neuron uit een aandachtspatroon naar iets in de wereld. Onze neuronen zijn mechaniekjes die deel uitmaken van de wereld. Neuronen worden dus volgens ‘natuurkundige principes’ aangestuurd. Geheugen in de vorm van aanpassing van gevoeligheden van neuronen speelt bij deze spiegeling een fundamentele rol—met “spiegeling” bedoel ik hier dus het spiegelen van de wereld in het neurale netwerk. Het van moment tot moment spiegelen van de wereld (referentie) is op zichzelf weinig waardevol. Van belang is dat een structuur in de tijd aangaande deze spiegelingen wordt onthouden, en wel zodanig dat het later tot een vergelijkbare structuur in de spiegende materie, dat is het neurale netwerk, kan leiden. Daarmee gaan we verder dan ‘natuurkundige principes’. Niet dat het ermee strijdig is. Het wordt slechts een ander onderwerp doordat geheugen en structuren in de tijd hun intrede in het begrijpen ervan doen. Niet alleen gaan en moeten we op een ander aggregatieniveau praten omdat we anders door de bomen het bos niet meer kunnen zien, het systeem op dit niveau is ook een realiteit in die zin dat het een structureel relatief stabiele situatie is van de wolk moleculen die wij een neurale netwerk noemen, welke als geheel interacteert het met zijn omgeving. Dat het als geheel acteert heeft te maken met de structuur zoals hierboven omschreven. Aldus is het daadwerkelijk zo dat ‘de inhoud’ van het neurale netwerk, c.q. jij/ikzelf, actor in oorzaak- en gevolg-relaties is.

Motoriek

Als we nog even een paar flinke denkstappen maken en spieren en botten aan ons neurale netwerk knopen, die worden aangestuurd door onze eigen voorstelling van onze eigen bewegingen (met andere woorden, ze zijn gekoppeld aan de uitvoer van een deel van het constructieve neurale netwerk), met daarbij wat slimme plannings- en energie-mechanismen, plus instincten en een onderliggend reptielenbrein, dan hebben we al iets dat qua intellect redelijk op een dier lijkt. Rest ons nog de werking van taal en zelfbewustzijn wat beter uit de doeken te doen om bij de mens uit te komen.

Uitweiding over tijdelijkheid

Alles bestaat in de tijd. Niks bijzonders aan. Maar in hoeverre speelt iets wat eerder gebeurd is later weer een rol? Dan kunt u zeggen: Ik heb een geheugen en daardoor herinner ik me dingen, wat heeft dat met bewustzijn te maken? Dan zeg ik: Okee, maar u spreekt steeds over “ik”. Formuleer dat eens zonder “ik” en maak er een machine van, want dat hebt u nodig om tot een verklaring van bewustzijn te komen. Zonder “ik” hebt u het over het meer tijdloze geheugen, en de meer tijdelijke verbeelding, dat is het meer tijdelijke bewustzijn. Een belangrijke vooronderstelling mijnerzijds is dus dat bewustzijn niet stil kan staan; ik kan met niets voorstellen bij een bewustzijn dat bevroren is, maar ik kan me ook niets voorstellen bij een bewustzijn zonder geheugen. Het geheugen verandert natuurlijk ook door het leren, maar minder snel.

Stel er dwarrelt een blaadje over straat van punt A naar punt B. Vervolgens

dwarrelt het naar punt C. Het feit dat het eerst op B lag is belangrijk, want anders had het later nooit precies op C kunnen belanden. Dat is ook een soort geheugen. Maar zodra het op C ligt, zijn A en B verdwenen. Het bijzondere van bewustzijn is dat het wel een meer tastbaar spoor achterlaat van zijn eigen gang, en wel zodanig dat elementen uit dat spoor ook steeds weer vanzelf terugkeren (wanneer dat nodig is). De kern van mijn verhaal zit hem in de relatie tussen dit vastgehouden spoor en het steeds veranderende bewustzijn. Dit vasthouden noem ik ook wel “spiegelen”, waarbij ik dan niet het spiegelen tussen het constructieve en deconstructieve neurale netwerk bedoel, als wel het in zichzelf spiegelen door het neurale netwerk van de wereld. Deze spiegeling is wat men in kunstmatige intelligentie theorieën (AI) probeert te beschrijven, maar het spiegelen zelf, daar doen zij weinig mee. Ik probeer juist wel het spiegelen, dat zijn de relaties tussen deze tijdloze spiegeling, de wereld en het bewustzijn—de laatste zijn beide tijdelijk van aard—te begrijpen.

Waarom is dit een verklaring voor bewustzijn? Ik denk dat u eerst tot u door moet laten dringen dat mijn idee eigenlijk een veel te simpele verklaring is voor een gezond denkend individu. Ten tweede is mijn idee ongelofelijk abstract. Er zijn potentieel wellicht vele implementaties denkbaar, alhoewel het nog best tegenvalt concrete voorbeelden te vinden. Een derde belangrijk punt is dat bewustzijn zoals hierboven bedoeld eigenlijk alleen maar interessant is indien het een bewustzijn *van iets* is. Daarvoor heb je iets nodig zoals mijn neurale netwerk, maar mijn bepaling van bewustzijn is in principe dus veel abstracter.

Mierenhoop “Miera Hoop”

Een bewustzijn dat geen bewustzijn *van iets* is vindt u bijvoorbeeld in Hofstadters’ boek “Gödel, Escher, Bach” in de vorm van Miera Hoop, dat is een mierenhoop. Mieren vormen in hun loopbewegingen en activiteiten bepaalde, terugkerende patronen omdat ze elkaars sporen en signalen volgen. Deze patronen worden wel beïnvloed door de omgeving en een mierenhoop heeft de neiging om zich te handhaven, dus enig bewustzijn van iets kan het niet ontzegd worden, maar wat stelt dat inhoudelijk voor? Zo goed als niets. Zolang u zich maar beseft wat dit bewustzijn inhoudt, is het niet vreemd hier bewustzijn aan toe te kennen.

Dit gezegd zijnde is het misschien de vraag of men een bewustzijn dat geen bewustzijn *van iets* is wel “bewustzijn” moet noemen? Misschien niet, maar ik los dat liever op door erop te wijzen dat systemen die geen bewustzijn *van iets* zijn allemaal dusdanig armetierig zijn dat die vraag helemaal niet interessant is. De enige uitzondering vormt wellicht AI, maar ook daar geldt: Het onderwerp van AI is datgene wat zich slechts en alleen in de bovenste lagen van mijn systeemje afspeelt.

Heb ik nu de vraag waarom dit een verklaring voor bewustzijn is beantwoord? Wel als u het met mijn definitie/beschrijving van bewustzijn eens bent. In abstracto voldoet mijn beschrijving van bewustzijn aan mijn idee ervan. Aangezien ik daarbij tevens het mechaniek ervan heb kunnen beschrijven, heb ik het verklaard. Het idee is zo ontzettend abstract dat het niet simpel meer lijkt, maar dat is het wel.

Geestesziekten

Bij dit simpele idee van bewustzijn komen dan duizend vragen op die allemaal getoetst moeten worden. Neem het belang van de zinnigheid van gedachten. Bepaalde geestesziekten maken mensen dusdanig verward dat hun bewustzijn er erg door beperkt wordt. Gedachten kunnen bij psychoses als een dwarrelend blaadje worden en het geheugen heel erg ontwrichten (desintegratie). Zinnigheid heeft dus te maken met samenhang en structuur, en dat heeft weer alles te maken met bewustzijn, want als er samenhang is in het tijdloze “geheugen”, dan zal het bewustzijn dat daaruit voortvloeit die samenhang ook hebben. Maar als die samenhang te sterk is, dan kan ze weer vernauwend zijn. Dat is precies het kenmerk van een neurose (fixatie).

Een ander belangrijk aspect van mijn bewustzijnstheorie is dat ik de illusie wil ontcrachten dat we ons van heel veel dingen tegelijk bewust zijn, zoals de meeste mensen volgens mij denken. Bewustzijn op een bepaald moment is voor mij totaal onzinnig. Dit is meer een aannname dan dat ik dit bewijzen kan, maar ik wil wel uitleggen hoe het lijkt alsof we ons op elk moment van heel veel meer bewust zijn.

Zelfbewustzijn van dieren en mensen

Geen enkel bewustzijn is helemaal zonder zelfbewustzijn. Dat zit hem in het relatief tijdloze “geheugen”. Het verleden van het bewustzijn speelt altijd een rol in een toekomstig bewustzijn. Alleen de manier waarop en de directheid waarmee het bewustzijn daar zelf deel van kan uitmaken kan heel verscheiden zijn. Mensen kunnen systematisch hun eigen bewustzijn registreren (ons geestes oog of “the mind’s eye”) middels mechanismen die ten grondslag liggen aan de taal en de taal zelf. Dieren associëren gebeurtenissen met hun eigen bewustzijn (middels hun aandachtspatronen) maar ze kunnen dit niet of moeilijk planmatig doen. Ze kunnen moeilijker hun eigen bewustzijns-“inhouden” met elkaar associëren anders dan via hun handelingen, omdat een eerder bewustzijn voor hen grotendeels weg is zodra het volgende er staat, met uitzondering van de tijdloze neerslag ervan. Dat zou netzo voor de mens gelden, ware het niet dat wij met taal ons eigen en andermans bewustzijn kunnen manipuleren. Bij zowel mensen als dieren echter, zullen opeenvolgende aandachtspatronen elkaar meestal overlappen, en aldus zullen, direct en indirect, toch associaties ontstaan, op dezelfde manier zoals wij die leggen wanneer wij naar een stomme film kijken. Volgens mij verschillen dieren wat dit betreft behoorlijk in hun vaardigheden; zoogdieren en vogels kunnen wel degelijk tijdelijke patronen waarnemen en onthouden (in de basal ganglia en/of de prefrontale, motorische cortex) en zo een zeker zelfbewustzijn ontwikkelen zoals mensen dat hebben. Bij Miera Hoop kan ik wat dit betreft helemaal niets ontdekken.

Noodzaak van aandacht voor bewustzijn

In mijn uiteindelijke bepaling van bewustzijn heb ik het belang van aandacht een beetje weggelaten. Mijn credo dat het belang van aandacht vervangt is: „Vraag niet of iets bewustzijn heeft, vraag hoe het eruit ziet”. Dit gezegd zijnde kan ik mij eigenlijk helemaal niet bedenken hoe er zonder aandacht een

samenhangend bewustzijn zou kunnen zijn waarin alles samenkomt, waardoor bewustzijn een waardige structuur zou kunnen krijgen. Als Popperiaan zou ik dit dus toch moeten toevoegen als essentieel voor bewustzijn. Popper zou namelijk stellen dat ik hier een hypothese van moet maken opdat anderen de kans krijgen dit te falsifiëren. Ik weet misschien wel een soort vervanger voor het aandachtsmechanisme, namelijk instinct—denk aan katten—maar in feite is dat ook een vorm van aandacht.

Aandacht bewerkt dat ook de grootste atheïsten uiteindelijk tot één belangrijkste, ordenend principe neigen te komen. Voor zover dat niet lukt blijven iemands' gedachten, of zelfs persoonlijkheden, als los zand, bijvoorbeeld met ethiek losgekoppeld van wetenschappelijk of praktisch denken, wat mij heel kwalijk lijkt.

Pijn, angst en superego

Bijzonder intrigerend in verband met bewustzijn zijn qualia zoals pijn en gevoelens in het algemeen. Begrip hiervan is ten dele complex omdat oude hersenstructuren en aangeboren reflexen een belangrijke rol spelen. Instincten zijn vermoedelijk iets anders aan onze nieuwe hersenen gekoppeld dan spieren en zintuigen. Ze zijn weinig specifiek, waarmee ik bedoel dat een instinct relatief veel neuronen prikkelt. Aldus kunnen ze onze waarneming globaal veranderen. Stel, bijvoorbeeld, dat een baby het instinct heeft om wanneer het twee donkere vlekken ziet een bepaald deel van de hersenen iets meer te activeren. Die twee donkere vlekken zullen meestal de ogen van de moeder of vader zijn. Aldus neigt het menselijke neurale netwerk ertoe om alles wat vader en moeder doen op een bepaald plekje op te slaan. Andersom kan het instinct deze informatie later ook “gebruiken” door dit deel van de hersenen te activeren. Dit zouden we dan kunnen analyseren als zijnde ons “superego”.

Aldus is instinct een soort aangeboren, vage aandacht. Een bijzonderheid van pijn is dat het tegelijkertijd een enorme aandachtstrekker is als ook een enorme prikkelaar van het neurale netwerk als geheel. Alhoewel pijn zelf niet echt als zodanig onthouden kan worden, moet langdurige pijn daarom haast wel leiden tot inhoudelijke reconstructie van hersenen. Het is in ieder geval heel vermoeiend. De tijdelijkheid van pijn zit hem erin dat het gewoon niet weggaat. Tegelijkertijd werkt het door zijn aandachtstrekkerij ook bewustzijnsvernaauwend. Voor angst geldt eigenlijk hetzelfde, maar daar is geen aanwijsbare bron van pijn anders dan een gedachte.

Een deel van de ervaring van qualia lijkt te liggen in de interactie met oudere hersenstructuren. Het is in ieder geval iets dat we met veel dieren delen. De enige, maar wel belangrijke bijzonderheid van mensen is dat ze zichzelf actief veel meer zaken kunnen voorstellen en aldus ook meer angsten kunnen hebben. Aldus is een boerderij voor dieren niet altijd een concentratiekamp, maar vivisectie blijft onozel of hufterig—ook al spelen bijvoorbeeld katten zelf ook graag met de dood van hun slachtoffers, en ook al is de afweging van belangen vaak lastig. Een mens in paniek is qua bewustzijn echt niet anders dan een dier in paniek, en opgejaagde of grof overheerste dieren gedragen zich ook niet heel veel anders dan opgejaagde of grof overheerste mensen. Dat is dan ook precies het rottige van dit vakgebied der hersenen en neurale netwerken; dat er zoveel beulen in rondlopen.

Dit gezegd zijnde leven we natuurlijk niet in het Paradijs, dus wij moeten ook overleven. Je zal maar een manisch-depressieve dochter hebben en je hoop koesteren op een medische ontdekking—niet dat ik bij veel wetenschappers iets anders vermoed dan nieuwsgierigheid, hang naar roem en brood op de plank; het blijft toch een argument.

Gemis en verbondenheid

Om er een beetje leuker einde aan te breien... Ik mis iets in mijn uitleg van gevoelens. Wellicht is dit gewoon het gemis zelf; het weten dat ik iets niet weet; het gat in mijn geest en verlangen naar iets dat ik niet heb, al is dat maar de oplossing om van mijn pijn af te komen. Toch is dit niet alles. Of ik nu denk aan kiespijn of aan gevoelens van schoonheid, er lijkt meer te zijn. Merkwaardig is dat ik dit juist ervaar ten aanzien van heel aardse zaken. Bewustzijn en individualiteit, dat begrijp ik wel, maar... dat gevoel en die verbondenheid, en het toch zowel vredige als wrede bestaan, daar begrijp ik geen bal van. Het lijkt wel alsof dit al datgene is wat in mijn bewustzijn komt maar wat toch buiten mijn macht ligt, en buiten mijn voorstellingsvermogen. Het stuurt en het duwt, en ik kan meegaan of me verzetten, of een list bedenken om het toch in mijn macht te krijgen. Het is merkwaardig en magnifiek.

Literatuur

Op de inhoud van dit artikel wordt dieper ingegaan in een Internetboek van de schrijver getiteld: “The Time Machine, Prototype of a Conscious Machine”, te vinden op <http://tm.bedenkerij.nl/>

Publicaties van Stephen Grossberg staan eveneens grotendeels op het Internet, helaas verwijzend naar de nodige vivisectie, maar voor de kern van zijn verhaal is dat eigenlijk overnodig. Zie <http://www.cns.bu.edu/Profiles/Grossberg/>

James Anderson en Edward Rosenfeld schreven (in 1998 alweer) een prachtig boek over de mensen achter de ontwikkeling van neurale netwerken: “Talking Nets. An Oral History of Neural Networks”, The MIT Press: Cambridge, London.