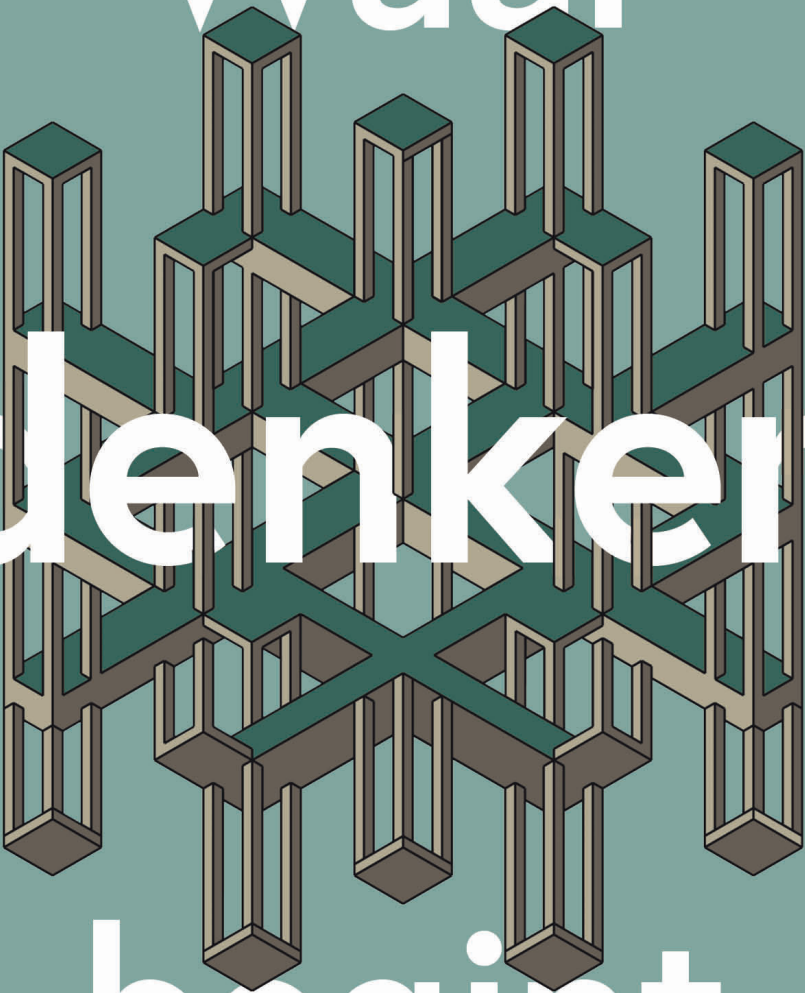


Gaurav Suri & Jay McClelland

Waar denken begint



Hoe intelligentie ontstaat in mensen en machines

Inhoud

Voorwoord	9
Deel 1: <i>Je geest is een neurale netwerk</i>	13
Hoofdstuk 1: Een uitnodiging	15
Hoofdstuk 2: Hoe de geest uit het brein is ontstaan	29
Hoofdstuk 3: Wat doet een neurale netwerk?	51
Deel 2: <i>Activatie produceert gedachten en handelingen</i>	83
Hoofdstuk 4: Een neurale netwerk als geheugensysteem	85
Hoofdstuk 5: Context doet ertoe	107
Hoofdstuk 6: De dingen die wij doen	131
Deel 3: <i>Kennis en leerprocessen: het zit in de verbindingen</i>	153
Hoofdstuk 7: Het ontstaan (en verlies) van betekenis	155
Hoofdstuk 8: De emergente denkmachine	187
Hoofdstuk 9: Door leren veranderen verbindingen	219
Deel 4: <i>Het raamwerk van het neurale netwerk uitbreiden en toepassen</i>	249
Hoofdstuk 10: Onze emergente gedachten	251
Hoofdstuk 11: Implicaties van het raamwerk van het neurale netwerk voor onszelf en voor AI	287
Woord van dank	309
Verder lezen en noten	311
Verantwoording illustraties	319
Index	321
Over de auteurs	331

Voorwoord

Onze hersenen bestaan uit enorme netwerken van cellen, de neuronen, die door patronen van almaar wisselende elektrische en chemische activiteit hun functies uitvoeren. Onze waarnemingen, gedachten, besluiten en handelingen – processen die we samenvatten als de geest – komen voort uit deze activeringspatronen.

Hoe? Hoe kan de geest ontstaan uit activeringspatronen in de hersenen?

Dit is een van de oudste en intrigerendste vragen die er zijn. De vraag gaat over onze essentie en onze plaats in het heelal. Hij gaat ook over de mogelijkheid – en mogelijkwerijs de aard – van kunstmatige geesten. Wij hebben *Waar denken begint* geschreven om een nieuw type antwoord te geven op deze vraag naar het hoe.

Wij zijn twee wetenschappers die onze carrières hebben gewijd aan het begrijpen van de menselijke geest. Toen we begonnen aan onze in eerste instantie gescheiden loopbanen, wilden we onderzoeken of de werking van de geest mechanistisch kon worden begrepen – de manier waarop je bijvoorbeeld een vliegtuig zou willen onderzoeken, of hoe een virus een ziekte veroorzaakt. We kwamen erachter dat de bestaande mogelijkheden om onze vragen te adresseren te vaag omschreven waren en vaak te ver afstonden van relevante feiten over de onderliggende hersenactiviteit. We dachten dat als we onze verklaringen baseerden op de mechanistische werkingen van het brein, dit betere antwoorden over onszelf zou opleveren.

De natuurkundige Richard Feynman liet na zijn dood het volgende op zijn schoolbord achter: ‘Wat ik niet kan maken, begrijp ik niet.’ Zijn stelling vat een essentieel aspect van onze benadering samen. Wij willen breinachtige systemen bouwen die de verschijnselen reproduceren van de geest die we willen begrijpen. Het is echter ondoenlijk een brein in al zijn detail te re-

produceren. In plaats daarvan bouwen we modellen die vele details weglaten zodat we vooruitgang kunnen boeken.

De modellen die we gebruiken – zogeheten neurale netwerken – zijn gebaseerd op de enorme neuronennetwerken in de hersenen. Hiermee kunnen we onderzoeken hoe onze menselijke vermogens uit neurale activiteit kunnen voortkomen. In deze modellen wordt bewust voorbijgegaan aan talloze complexiteiten van het brein, zodat we ons kunnen richten op de basisprocessen waarmee we kunnen begrijpen hoe onze geest werkt. In dit boek beschrijven we neurale netwerken die een verklaring bieden van hoe mensen waarnemen, beslissen, concepten vormen en doelen nastreven.

Met deze modellen kunnen we ook de antwoorden begrijpen op vragen over onszelf die ons en vele anderen voor een raadsel hebben gesteld. Deze vragen beginnen vaak met ‘waarom’, ‘wat’ en ‘waar’. Waarom lukt het ons soms niet om onze bedoelingen te volgen? Waarom vertonen wij en anderen diepgewortelde vooroordelen? Wat maakt dat we soms makkelijk inzien dat iets waar is, terwijl we het een andere keer niet begrijpen? Waar komen onze intenties vandaan en waarom kunnen ze vaak verkeerd zijn?

Neurale netwerken – geïmplementeerd als computerprogramma's – hebben opvallend genoeg de basis gevormd van kunstmatige intelligentie. Modellen die wij en anderen oorspronkelijk hebben ontwikkeld om de menselijke geest te begrijpen, blijken de basis te vormen voor het bouwen van kunstmatige geesten. Hoe neurale netwerken in staat zijn onze denkvermogens na te bootsen, werpt ook een licht op de huidige AI-systemen. In dit boek bespreken we de belangrijkste principes van de neurale netwerken van onze eigen geest die ten grondslag liggen aan de AI-systemen die de kenmerken van onze menselijke cognitieve vaardigheden benaderen en soms overschrijden. Hoewel onze focus ligt op wat wij beschouwen als ideeën die standhouden, bespreken we ook bepaalde innovaties binnen AI-systemen die zich vanaf halverwege de jaren twintig van deze eeuw snel ontwikkelen – met name in de laatste twee hoofdstukken. Deze innovaties zijn interessant en veelbelovend en kunnen zich de komende jaren nog veel verder ontwikkelen. Waar mogelijk hebben we deze ideeën beschreven op een manier die hun onderliggende principes benadrukt en niet hun tijdelijke toepassingen. Ons doel is de lezer een basis te geven die ook nuttig blijft naarmate zich specifieke toepassingen ontwikkelen.

Dit boek bestaat uit vier delen. In deel 1 beschrijven we hoe een systeem eigenschappen kan hebben die niet in een van de onderdelen aanwezig zijn. Dit verschijnsel, dat we emergentie noemen, vormt de kern van het raam-

werk voor het neurale netwerk van de geest. Hier komt de geest voort uit de interacties tussen eenvoudige verwerkingseenheden, analoog aan hersencellen. De hersencellen zelf kunnen niet denken, maar hun interacties doen een systeem ontstaan dat wel kan denken. In deel 2 laten we zien hoe het raamwerk voor het neurale netwerk van de geest een verschillend aantal menselijke gedragingen kan verklaren. Eerst kijken we naar de werking van het geheugen op basis van het neurale netwerk, waaronder de feilbaarheid ervan. Vervolgens beschouwen we onze afhankelijkheid van context om de wereld om ons heen te kunnen begrijpen, waaronder de manier waarop onze verwachtingen vormgeven aan onze gedachten. Daarna nemen we de besluitvorming onder de loep, inclusief hoe onze keuzes dan weer rationeel, dan weer irrationeel zijn. In deel 3 gaan we dieper in op hoe neurale netwerken – zowel biologische als kunstmatige – leren van ervaring. We beschrijven hoe leren onze kennis over objecten en hun eigenschappen mogelijk maakt en hoe dit het gebruik van taal kan ondersteunen, met name in grote taalmodellen (*large language models*, LLM's). In deel 4 ten slotte breiden we het gezichtspunt van de geest uit gezien vanuit het neurale netwerk en passen dit toe. We beschrijven hoe neurale netwerken handig kunnen zijn bij het begrijpen van verschijnselen als formeel redeneren, gemotiveerd gedrag en bewustzijn – kenmerken van de geest die nog niet volledig binnen het raamwerk voor het neurale netwerk zijn vastgelegd. We sluiten af door enkele implicaties van het perspectief van het neurale netwerk te behandelen – zowel voor de mens als voor machines.

In het boek zijn intermezzo's opgenomen die, tenzij anders vermeld, merkwaardige en fictieve gesprekken bevatten. Zo is er een gesprek tussen Sigmund Freud en Adam Smith en een ander waarin de redacteur van dit boek in een bar in New York City tegen een (fictieve) klant praat. We stellen ons deze gesprekken en alle woorden die de personen gebruiken – historisch dan wel fictief – voor als hersenspinsels. We hopen dat deze intermezzo's de kijk van de lezer op de besproken kwesties net zozeer stimuleren als het ons heeft gedaan bij het schrijven van dit boek.

Wij hebben dit boek geschreven voor iedereen die nieuwsgierig is naar de geest – zowel de menselijke als de kunstmatige. Er wordt geen wiskundige kennis verondersteld. Afgezien van enkele sommen en vermenigvuldigingen, staan er geen vergelijkingen in het boek. Ook wordt er geen voorafgaande kennis van cognitieve, psychologische, neurale of computerwetenschappen verwacht. Aan het eind van het boek zijn noten opgenomen met bronnenmateriaal voor degenen die meer willen weten.

Ons begrip van de mechanismen van de geest blijft zich ontwikkelen. Er is nog veel wat op ontdekking wacht, maar wat er bekend is, is opwindend en heeft verstrekkende gevolgen.

We nodigen de lezer uit deze onderneming met ons aan te gaan. Mogelijk verrijkt het de manier waarop je jezelf en je plaats in het heelal ziet.

Gaurav en Jay

Hoofdstuk 1

Een uitnodiging

Toen een van de auteurs van dit boek, Gaurav, veertien jaar oud was, kreeg hij van zijn ouders voor zijn verjaardag het equivalent van 50 dollar. Het geld was voldoende om een spijkerbroek met wijde pijpen te kopen – in die tijd een must-have voor een tiener – of voldoende om een plek te reserveren voor een langverwacht schoolreisje met zijn vrienden. Het probleem was dat hij er maar één kon kiezen, terwijl hij ze allebei wilde hebben. Hij moest een beslissing nemen. Die avond beet hij door de zure appel heen en maakte zijn keuze. Hij zou op reis gaan. Die broek kon wel wachten. Hij wist zeker dat hij de juiste keuze had gemaakt.

Maar de volgende ochtend was er iets geks aan de hand: Gaurav werd wakker met het gevoel dat hij de broek moest kiezen. Er was niets veranderd aan de twee opties, maar zijn beslissing was veranderd. Deze twijfel, die de dagen erna een aantal keren omsloeg, verbaasde hem enorm. In die periode was hij net met computers in de weer en hij stelde zich de geest voor als een soort computer die op basis van logische principes werkte. Welke computer gaf 's morgens het ene antwoord en 's avonds een ander? Hoe konden op logica gebaseerde antwoorden veranderen terwijl er geen duidelijke reden voor was? En als zijn gedachten en voorkeuren niet voortkwamen uit logica en verstand, wat waren het dan voor dingen?

Gaurav was toevallig gestuit op grote, blijvende vragen van de mensheid: hoe ontstaan onze gedachten? Waarom doen we wat we doen? Kunnen we onze gedachten vertrouwen? En meer algemeen: wat is de geest en hoe werkt hij?

Algemene opvattingen over de geest (en hun beperkingen)

Wat is de geest? We kunnen deze beschouwen als iets binnen in ons dat leidt tot gedachten, waarnemingen, herinneringen, gevoelens, beslissingen en handelingen. Maar wat is het eigenlijk echt? Waar komt het vandaan? Hier gaan we kort in op enkele algemene opvattingen en de beperkingen ervan.

Een van de opvattingen over de geest is gebaseerd op onze godsdienstige tradities, waarvan er vele zich toespitsen op het idee dat onze geest is afgeleid van goddelijke materie of geestelijke stof die menselijk vlees tot een persoon kan bezielen. De herkomst van dit idee is niet moeilijk te begrijpen: menselijk vlees lijkt zo gewoon en geestloos. Hoe is het mogelijk dat het intelligentie kan voortbrengen? De geest moet wel ergens anders uit voortkomen, iets onbekends, iets van buiten de aarde. Iets dat ons verbindt met het onsterfelijke, het heilige en het zinvolle. Iets goddelijks.

Als we aan de geest iets eeuwigs toeschrijven, is dat weliswaar een opwindend en schitterend idee, maar het is niet echt een verklaring van wat de geest is en hoe gedachten ontstaan. Op deze manier wordt de geest beschouwd als een onnoembare entiteit die een nader begrip onmogelijk maakt. Willen we begrijpen hoe de geest gedachten voortbrengt en de andere dingen die we eraan toeschrijven, dan is dit geen acceptabele conclusie.

Een tweede opvatting suggereert dat de geest een verzameling overtuigingen en wensen is in de vorm van een soort zinnen. Zo is 'Mensen die zijn afgestudeerd, verdienen meer geld' een voorbeeld van een overtuiging. En zo is 'Ik wil later een baan hebben die goed betaalt' een voorbeeld van een wens. Het lijkt redelijk te veronderstellen dat overtuigingen en wensen kunnen samenwerken om intenties en handelingen te laten ontstaan. Als iemand vraagt waarom we hebben besloten om op een bepaalde manier te handelen, kunnen we verwijzen naar de overtuigingen en wensen die blijkbaar onze beslissing hebben gestuurd. Waarom zijn we gaan doorleren? Het antwoord kan zijn: 'Om een goedbetaalde baan te krijgen.' Misschien dat de werking van de geest niet meer inhoudt dan interacties van overtuigingen en wensen om doelen te vormen die vervolgens ons gedrag sturen.

Een beperking van het overtuiging-wensmodel van de geest is dat het niet verklaart waar deze overtuigingen en wensen vandaan komen. Hoe kunnen zulke abstracte zaken als overtuigingen en wensen het gevolg zijn van fysische processen in ons brein, en hoe kunnen ze aanleiding geven tot fysieke handelingen, waaronder bewegen, handelen en het produceren van taal? Bovendien verklaart het model niet waarom mensen niet handelen

naar hun overtuigingen en wensen. Patiënten nemen bijvoorbeeld vaak geen medicijnen die goed voor hen zijn en werknemers bouwen geen pensioen op, ondanks dat het van groot belang is voor hun financiële toekomst. Dit gebeurt ondanks dat dergelijke personen daadwerkelijk geloven in de werkzaamheid van hun medicijnen of de zekerheid willen van een pensioen. En toch doen ze niets.

Volgens een andere opvatting lijkt de geest op software die zijn input uit de buitenwereld krijgt en hierop een reeks regels toepast, mogelijk door de evolutie verstrekt. In sommige gevallen lijken er voor de werking van de geest inderdaad regels te worden toegepast: als een dier vleugels bezit en kan vliegen, wordt het als een vogel geclassificeerd; we kiezen iets van een menu omdat we denken dat we hiermee onze waarde maximaliseren vergeleken met andere dingen op de kaart; en we voorspellen dat de verleden tijd van een recent bedacht werkwoord als 'faxen', 'gefaxt' is, in overeenstemming met de eenvoudige regel dat de voltooid tijd van een werkwoord gewoon op een 'd' of een 't' eindigt afhankelijk van de laatste letter van de stam.

Een probleem met de zienswijze geest-als-software is dat we er niet ver mee komen voor een begrip van de geest. Ja, talloze vogels kunnen vliegen, maar we herkennen ook vogels die niet kunnen vliegen. Ja, we kiezen van een menu iets wat we lekker vinden, maar onze keuzes zijn vaak beïnvloed door variabelen die gepaard gaan met de inherente waarde, bijvoorbeeld of iets op het menu bovenaan de pagina staat. En ja, we voegen vaak een 'd' of een 't' toe aan een woord in de voltooid tijd, maar er zijn vele onregelmatige werkwoorden waarvoor dat niet opgaat ('slapen' wordt bijvoorbeeld 'geslapen'). Een andere belangrijke beperking van deze modellen is dat ze niet hebben geleid tot kunstmatige intelligente systemen die werkten. De benadering die we in *Waar denken begint* presenteren, blijkt veel succesvoller te zijn.

Ten slotte is er de veelgehoorde opvatting waarin verschillende aspecten van de geest als afhankelijk beschouwd worden van afzonderlijke specialisten, waarbij elke specialist zich in een eigen gespecialiseerd gebied in de hersenen bevindt. Volgens dit standpunt bewegen we vanwege hersengebieden die in beweging zijn gespecialiseerd, zien we vanwege hersengebieden die in het zien zijn gespecialiseerd, raken we gemotiveerd vanwege hersengebieden die in motivatie zijn gespecialiseerd, spreken we vanwege hersengebieden die in taal zijn gespecialiseerd. In sommige gevallen van dergelijke verklaringen denken we doordat we afhankelijk zijn van gespecialiseerde gebieden die denken.

We kunnen er niet omheen dat sommige hersengebieden een zekere mate van specialisatie vertonen. De vraag is, wat is de oorzaak van deze specialisatie? Een benadering die door de filosoof Jerry Fodor wordt bepleit in zijn boek *The Modularity of Mind*, is dat deze specialisatie voortkomt uit de specifieke interne eigenschappen van die hersengebieden die door de evolutie geselecteerd zijn om berekeningen uit te voeren die gespecialiseerd zijn voor de taken die ze uitvoeren. Hoewel hersengebieden in bepaalde mate intern van elkaar verschillen, is de zienswijze die we in *Waar denken begint* presenteren dat een dergelijke specialisatie grotendeels het gevolg is van verschillen in verbindingen naar en vanuit verschillende hersengebieden. Zo speelt bijvoorbeeld de visuele cortex een belangrijke rol in de visuele waarneming omdat dit hersengebied een sterke input krijgt vanuit de ogen. Dit suggereert dat als de input naar een hersengebied verandert, dit een verandering zal veroorzaken in de functie die door dat hersengebied wordt uitgevoerd. Het blijkt inderdaad dat bij personen van wie de visuele cortex geen input krijgt omdat ze vanaf de geboorte blind zijn, dit gebied een nieuwe, niet-visuele functie krijgt, zoals het verwerken van geluid of aanrakingen, die beide enige input geven aan dit hersengebied.

In plaats van de geest te beschouwen als een verzameling onveranderlijke, gespecialiseerde modules, kunnen we de geest volgens dit standpunt zien als een adaptief systeem dat door ervaring, leren en omgevingsinvloeden zijn vorm krijgt. Met deze benadering kan een verklaring worden gegeven voor het bestaan van wat sommige wetenschappers de *visual word form area* noemen; een hersengebied dat gespecialiseerd lijkt voor het lezen van geschreven woorden die visueel worden gepresenteerd. Het is niet plausibel dat dit gebied door de evolutie is geselecteerd om zich in lezen te specialiseren. Lezen en schrijven zijn immers zo'n vijfduizend jaar geleden ontstaan. Vijfduizend jaar is evolutionair een veel te korte tijd om dankzij natuurlijke selectie een specifieke leesmodule te produceren.

Toch is dit hersengebied gespecialiseerd in lezen bij mensen die hebben leren lezen. Waarom? Lezen is sterk afhankelijk van het kunnen onderscheiden van kleine visuele details, zoals de verschillen tussen letters die op elkaar lijken. De *visual word form area* ontvangt sterke input van andere neuronen in visuele hersengebieden die de hoogste gevoeligheid hebben voor details. Op deze manier wordt het gebruikt in het leesproces. Bij mensen die niet lezen raakt dit gebied gespecialiseerd in andere taken die ook afhangen van het kunnen onderscheiden van kleine details, zoals het herkennen van gezichten. Zulke ontdekkingen ondersteunen ons standpunt dat specialisatie

in het brein gevormd wordt door ervaring en afhankelijk is van de input en output van hersengebieden. Het zijn dus geen specifieke, uit evolutie voortgekomen functies.

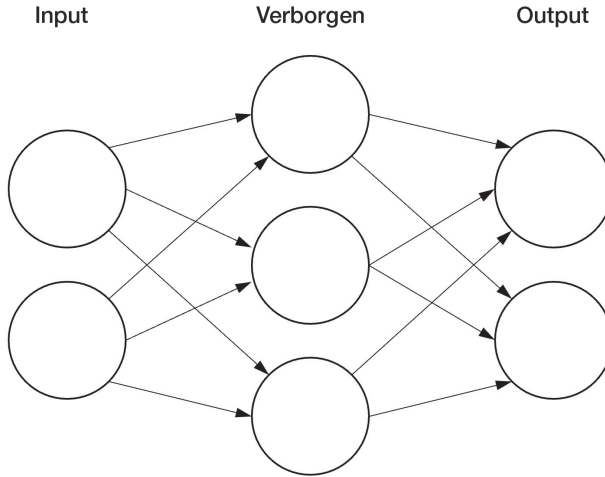
We hebben een paar opvattingen van de geest besproken die alle in hun verklaringen tekortschieten. We gaan nu kijken naar een opvatting die een route biedt naar een dieper begrip van onze geest en die de opkomst van kunstmatige intelligentie mogelijk heeft gemaakt: de opvatting dat de geest naar voren treedt in een neurale netwerk.

Wat is een neurale netwerk?

Gezien de beperkingen van de gangbare opvattingen over de geest hebben wetenschappers en wiskundigen een totaal andere benadering gezocht om de geest te begrijpen. Ze dachten dat het productief zou kunnen zijn om eerst een goed begrip te krijgen van de output van de geest (onze gedachten en handelingen) door de signalen in de neuronennetwerken in het brein te volgen. Het brein bestaat uit miljarden cellen, neuronen genaamd. Een neuron is de fundamentele bouwsteen die verantwoordelijk is voor het verwerken en doorgeven van informatie in het brein en de rest van het zenuwstelsel.

Op basisniveau doen neuronen dingen die we goed begrijpen: (1) ze kunnen geactiveerd worden – dat wil zeggen dat ze kleine elektrische ont-ladingen genereren, de zogenaamde ‘actiepotentialen’; (2) ze kunnen signalen doorgeven naar andere neuronen waarmee ze verbonden zijn; en (3) ze kunnen verbindingen met andere neuronen maken en de sterkte ervan aanpassen.

Laten we ons een voorstelling van een netwerk van neuronen proberen te maken. Het is lastig om echte neuronen weer te geven, dus we kunnen hiervoor cirkels gebruiken die we ‘eenheden’ noemen. Sommige van deze eenheden zijn met andere verbonden en deze verbindingen kunnen we weergeven met pijlen tussen de eenheden. We kunnen eenheden die door input van de buitenwereld geactiveerd worden aangeven als ‘input-eenheden’, eenheden die output naar de buitenwereld sturen als ‘output-eenheden’, en eenheden die geen contact met de buitenwereld hebben als ‘verborgen eenheden’. Op die manier is een neurale netwerk ontstaan (afbeelding 1.1).



Afbeelding 1.1: Een prototype van een neurale netwerk.

Hoe kunnen we zo'n neurale netwerk gebruiken om de geest proberen te begrijpen? Om te laten zien hoe dit mogelijk is, stellen we ons een bus voor die met hoge snelheid op een voetganger op een zebrapad afkomt. Als we het brein van de voetganger zien als een neurale netwerk, kunnen we ons voorstellen dat de signalen die een groot bewegend voorwerp detecteren via de ogen naar de inputeenheden worden doorgestuurd. Deze kunnen een signaal aan verborgen eenheden afgeven, die weer een signaal aan de output-eenheden kunnen afgeven. Deze kunnen ervoor zorgen dat de beenspieren van de voetganger samentrekken zodat deze uit de weg gaat. De pijlen geven de richting van de signalen aan. Zo beïnvloeden bijvoorbeeld inputeenheden verborgen eenheden en wijzen de pijlen dus van links naar rechts.

Deze verklaring is gebaseerd op het soort dingen waarvan we weten dat neuronen ze doen. Het suggereert dat onze neuronen – die elk op zich niet kunnen bedenken om voor een aanstormende bus weg te lopen – een dergelijke actie voor elkaar kunnen krijgen door aan elkaar signalen af te geven en elkaar te activeren. Deze voorstelling van een neurale netwerk suggereert dat de gedachten en handelingen in het brein veroorzaakt worden als gevolg van de interactie tussen onze neuronen. Iets anders is er niet.

Natuurlijk is dit afgeven van signalen door neuronen naar elkaar complexer dan in dit vereenvoudigde voorbeeld. Maar je kunt vooruitgang boeken door je op slechts enkele aspecten van dit proces te richten. In de eerste plaats verschillen neuronen in de frequentie waarmee ze hun actiepotentialen

genereren, iets wat we ons kunnen voorstellen als een glijdende schaal van activaties. Ten tweede kan elk neuron signalen ontvangen van vele andere neuronen. Bij benadering kunnen we stellen dat in neuronen de van andere neuronen ontvangen activaties bij elkaar worden opgeteld. Hoe groter de som van deze input, hoe sterker de activatie. Ten derde kunnen neuronen verbindingen aanleggen en de sterkte ervan wijzigen als een functie van de ervaring. Als we bijvoorbeeld iemand ontmoeten die we niet kennen, kunnen er verbindingen worden gemaakt tussen neuronen die het gezicht van die persoon representeren en neuronen die de naam representeren. Deze verbindingen kunnen bij herhaaldelijke ontmoetingen worden versterkt. Een sterke verbinding tussen twee neuronen zorgt ervoor dat ze elkaar sterker beïnvloeden.

Het is ondoenlijk om de vele neuronen die op verschillende niveaus worden geactiveerd en elkaar tegelijkertijd beïnvloeden via verbindingen van verschillende sterkte, in een diagram vast te leggen. Maar een van ons, Jay McClelland, kwam een voorbeeld tegen waarmee hij de diverse gelijktijdige werkingen van een neuraal netwerk kon vormgeven (afbeelding 1.2).



Afbeelding 1.2: Een riviertje dat door poeltjes stroomt is een goed voorbeeld van enkele van de gelijktijdige werkingen van een neuraal netwerk.

In die tijd was Jay een jonge universitair docent die onderzoek deed naar de geest, niet als een systeem dat afhangt van afzonderlijke ideeën die stapsgewijs worden bekeken, maar als een continu werkend systeem dat van allerlei invloeden van verschillende sterkte afhankelijk is. Waar anderen tijdens een wandeltocht in het Yosemite National Park een riviertje zagen dat trapsgewijs in verschillende poeltjes stroomde, zag Jay een metafoor voor deze continue

processen in de geest. Hij maakte zich een voorstelling van de diverse poeltjes als evenzovele gedachten. De hoeveelheid water erin komt dan overeen met hun activaties. Elk poeltje kan water ontvangen van meerdere poeltjes en het waterniveau is afhankelijk van de som van deze input. Sommige poeltjes waren door diepe of brede stroompjes met veel water verbonden met andere poeltjes, terwijl weer andere verbonden waren met stroompjes die nauwelijks water bevatten. Deze verbindingen riepen verbindingen op van verschillende sterkte die de gedachten met elkaar verbonden. Hij realiseerde zich dat deze metafoor hem hielp om over de geest na te denken als een neurale netwerk. Al snel begon hij modellen te bouwen die deels door deze intuïtie werden ingegeven.

Zoals alle metaforen is Jays metafoor niet perfect. Maar behalve dat ze de processen in een neurale netwerk illustreert, suggereert de mechanistische aard ervan een intrigerende mogelijkheid: de werking van de geest kan als een fysisch proces worden gezien – gelijk aan water dat omlaag stroomt of planten die zich naar het licht buigen – in plaats van de uitdrukkingsvorm van een raadselachtige essentie. Zou dit een beginpunt kunnen zijn om alle mysteriën van de geest te begrijpen?

De emergentie van de geest in een neurale netwerk

Wellicht begin je nu in te zien dat een neurale netwerk mogelijk een route biedt naar een begrip van de geest. Maar je bent mogelijk nog onzeker over het idee dat het denken voortkomt uit de interacties tussen neuronen, die elk op zich niet tot gedachten in staat zijn.

Ons centrale idee in dit boek is dat het denken ontstaat uit componenten die niet zelf kunnen denken.

Voorbeelden kunnen helpen om emergentie te begrijpen. Een van de indrukwekkendste manifestaties van emergent gedrag is de gecoördineerde vlucht van een zwerm vogels, vaak spreeuwen of zwaluwen. De zwerm gaat alle kanten op, golft en pulseert, als een fascinerend ballet in de lucht (afbeelding 1.3). Wie een dergelijke groep vogels ziet, krijgt makkelijk het gevoel dat de zwerm één geheel is, met een eigen geest die op een of andere manier de bewegingen van zijn componenten coördineert.

Maar er is geen overkoepelende intelligentie. Elke vogel is in interactie met een klein aantal buurvogels, meestal vlak in de buurt. Elke vogel reageert op de veranderingen in de richting en snelheid van de vlucht van zijn meest nabije burens. Deze lokale interacties zorgen voor een domino-effect waarbij een kleine wijziging in de vlucht van één vogel zich door de hele zwerm kan