

door Rik Smits

Dr. A.H. Heinekenprijs voor de Cognitiewetenschap voor John Anderson

Hoe wij nieuwe dingen weten

Voor dingen als autorijden is er geen evolutionaire basis. John Anderson probeert te begrijpen en in een model te stoppen hoe onze hersenen dat dan toch kunnen leren. Dat je ook op een wetenschappelijke manier naar de menselijke geest kon kijken, was voor hem indertijd een ontdekking.

Het gebouw ligt tussen de bomen, achteraan op de verder hypermoderne centrale campus van de Carnegie-Mellon Universiteit in Pittsburgh. De voormalige staalstad is nog altijd doortrokken van Andrew Carnegie, de onmetelijk rijke staalbaron die ons het Vredespaleis gaf. Lange, rondbogige gangen in donkerbruine en gele steen, zwart-marmeren trappen, cirkelvormige gietijzeren radiatoren uit een andere tijd. Zo ongeveer moet het stadhuis van Rommeldam eruitzien, al zegt dat Amerikanen natuurlijk niets. Helemaal aan het eind op de tweede verdieping ligt het domein van John Anderson (59), de cognitief psycholoog die hier al sinds 1978 naarstig werkt aan zijn ACT-R theorie, het overkoepelende verhaal van hoe wij aan al onze wonderlijke, vaak onnatuurlijke vermogens komen. Het verhaal waarmee hij als eerste de nieuw ingestelde Dr. A.H. Heinekenprijs voor de Cognitiewetenschap heeft toegekend gekregen.

Het vermogen om snel willekeurige nieuwe dingen aan te leren is de kern

We praten in zijn werkkamer. Een schemerige, wat Spartaanse ruimte waar elke drie minuten de PC tevergeefs dingdongt dat er alweer e-mail is. We hebben het over van alles. Bijvoorbeeld over hoe Anderson in het afgelopen jaar al alles over de Heinekenprijzen gehoord had, omdat de instelling van de prijs voor de Cognitiewetenschap zijn wereldje in rep en roer had gebracht, met verhitte discussies over het effect dat zo'n prestigieuze prijs zou heb-

ben op een gemeenschap van wereldwijd zo'n duizend, hooguit tweeduizend mensen. Over zijn verbazing toen hij prompt ook nog de eerste winnaar bleek. Maar ook over zijn geesteskind, de ACT-R theorie, en over hoe het allemaal begon. Hoe de dingen soms anders lopen dan je bedacht had, en hoe Anderson ooit heel andere plannen had met de menselijke geest.

'Ik wilde wel schrijver worden. Daarvoor ging ik naar de universiteit van British Columbia in Canada, waar ik vandaan kom. Maar ik had pech, of misschien juist wel geluk. Ik kreeg meteen Margaret Atwood als docent, de schrijfster die veel later in 2000 de Booker Prize zou winnen. Dat was een nogal ontzagwekkende dame, en na een jaar of zo was mij wel duidelijk dat ik het ware romancier-stalent niet bezat.'

Maar wat moest ik dan wel doen? Ik neusde wat rond, tot ik tegen het idee aanliep van een *wetenschappelijke* benadering van de menselijke geest. Dat sprak me nou wel aan! De meeste psychologen op mijn universiteit waren zogenaamde Hollyaanse behavioristen, strikt gefocust op de rat. Dat strakke, die poging om een formele wetenschappelijke theorie te bouwen, dat vond ik mooi.'

'Dat ze misschien wat achterliepen bij de grote wereld, op een doodlopend spoor zaten, dat realiseerde ik me pas later. Maar ik zag wel dat wat zij deden ver afstand van de menselijke geest. Dus ging ik op zoek naar hoe je die geest zelf op een wetenschappelijke manier kon benaderen, las bijvoorbeeld vroeg werk van de taalkundige Noam Chomsky. Ik wilde mathematische psychologie gaan bedrijven, en sommige docenten steunden me in dat soort subversieve gedachten.'

Subversief?

‘Mijn universiteit was destijds nog een bastion van het ouderwetse behaviorisme, waarvoor alleen observeerbaar gedrag telde en dat van abstract getheoretiseer niets moest hebben. Hoe dan ook, met dat idee om de menselijke geest formeel aan te pakken ging ik in 1968 naar promotieplaatsen solliciteren, thuis in Canada maar ook in de Verenigde Staten. En de mooiste aanbiedingen uit Amerika rolden binnen! Wat bleek? In dat jaar was promoveren in dat land voor het eerst geen reden meer voor uitstel van militaire dienst – Vietnam, weet je wel. De universiteiten moesten dus wel meer buitenlandse promovendi en zelfs vrouwen binnenhalen – dat was wat hoor.’

‘Uiteindelijk kwam ik in Stanford terecht, bij de mathematisch psycholoog Gordon Bower, die me al bijna meteen vertelde dat de mathematische psychologie dood was en dat ik Artificial Intelligence (AI) moest gaan doen, kunstmatige intelligentie. Zo kreeg ik te maken met computationele modellen van allerlei cognitieve processen.’

Mensen leren van uitgewerkte voorbeelden, van instructies en door zelf te ontdekken

‘Samen met Bower maakte ik op Stanford een soort algemene simulatie van het associatieve geheugen van de mens, die we HAM noemden. Dat was in 1973. Daarna was ik, voordat ik in 1978 hier binnenkwam, onder meer drie jaar Junior Fellow aan de Universiteit van Michigan. Een fantastische tijd, je had geen neventaken, de meeste mensen wisten nauwelijks dat je bestond, dus je kon naar hartenlust je eigen ding doen. Mijn ding was het maken van een algemenere theorie van menselijke cognitie, en dat liep uit op de eerste versie van de *Adaptive Character of Thought* theorie, kortweg ACT, in grote lijnen nog altijd dezelfde waar ik nog steeds mee bezig ben, al heet hij nu ACT-R.’

Die theorie is het verhaal over hoe het kan dat mensen de meest uiteenlopende ingewikkelde vaardigheden kunnen verwerven waar geen evolutionaire basis voor is, van pianospelen via autorijden tot algebraïsche sommen oplossen.

‘Inderdaad. Dat vermogen om snel willekeurige nieuwe dingen aan te leren is de kern, en het is ook datgene wat ons onderscheidt van andere soorten. Kijk, onze hersenen bevatten een heel stel voor een bepaalde functie

specifieke modules. Grote delen van het brein verwerken bijvoorbeeld visuele informatie, er zijn de verschillende geheugens, er is een module die plant en doelen in de gaten houdt, en zo voort. Dat hebben we met andere soorten gemeen. En als je maar tijd van leven hebt, kun je een aap verbazend ingewikkelde dingen aanleren. Misschien zelfs algebra, het zou me alleen tweehonderd jaar kosten om het beest zo ver te krijgen. Maar kinderen maken zich zoiets in een paar honderd uur eigen: je laat ze wat voorbeelden zien, legt er wat bij uit en hup, daar gaan ze.’

Integratie is voor u een sleutelbegrip. Is dat niet wat hoog gegrepen? Zou je je niet beter op een bepaalde module kunnen richten?

‘Natuurlijk is het veld veel breder dan wie dan ook kan behappen. En met veel modules die een plaats hebben binnen de ACT-R theorie heb ik weinig of niets te maken gehad. Dat is één manier waarop integratie essentieel is: verdeling van werk. Wat we willen is juist de verschillende theorieën en modellen die al op allerlei deelreinen van de cognitieve psychologie bestaan, integreren tot één samenhangend geheel dat meer is dan de som der delen.’

‘Maar integratie is ook de sleutel tot ons vermogen om willekeurige nieuwe vaardigheden te verwerven. Als er voor een vaardigheid geen evolutionaire basis kan zijn, zeg autorijden, dan moet zo’n vaardigheid wel berusten op het herconfigureren van vaardigheden waarover we wel van nature beschikken. We nemen dus elementen van ons bestaande cognitieve systeem en verbinden die op nieuwe manieren tot iets compleet nieuws. Dat is integratie.’

Die integratie gebeurt door een centraal systeem met productieregels, dat ook weer nieuwe productieregels produceert. Wat moeten we ons daarbij voorstellen?

‘Wel, dat productiesysteem heb je in elk geval nodig, om gegevens die de verschillende modules opleveren, samen te brengen en daaruit gedragingen te laten ontstaan. Stel het je maar voor als een spin in een web van modules. Elk van die modules levert gegevens op die geparkeerd worden in een buffertje, en wat in het buffertje zit is zichtbaar en bruikbaar voor de regels van het productiesysteem.’

‘Daar zitten bijvoorbeeld regels in die dingen die je ziet, herkennen, lokaliseren en die vervolgens je hand instrueren om naar de gevonden coördinaten te gaan en samen te trekken. Dat hele proces noemen we pakken. Als je echter elke keer dat je iets pakt stapje voor stapje dat hele proces van elkaar oproepende regels moest

afwerken, zou je ontzettend traag zijn. Dus zorgt een proces van compilatie dat naarmate je het proces vaker doorloopt, telkens uit twee of meer van die regels een nieuwe gevormd wordt, zodat een soort afgekorte procedure ontstaat.'

Is dat wat er zit achter het feit dat je bij je eerste rijles ontzettend zit te schutteren, maar na verloop van tijd een auto 'vanzelf' bestuurt?

'Precies. En het interessante is dat daarbij van alles komt kijken – knoppen, meters, spiegels, pedalen – waar je brein van nature geen weet van heeft. Bij die nieuwe, gecompileerde regels wordt ook in het geheugen opgeslagen informatie geïntegreerd, zoals bijvoorbeeld de plaats en de vorm van de knop voor de richtingaanwijzer. Het eind van het liedje is dat je hersens regels bevatten die specifiek over auto's besturen gaan. Je brein weet voortaan van auto's, het heeft een vaardigheid verworven die er oorspronkelijk zelfs niet in aanleg was.'

Houdt dat in dat je geboren wordt met een beperkt setje algemene, misschien wel lege productieregels, die je bij het combineren als het ware invult met dingen die je in je geheugen opslaat?



John Anderson

'Goede vraag... Je lijkt inderdaad wel zo'n aangeboren kernset te moeten hebben, maar we hebben ons nog niet echt beziggehouden met waar de algemene regels vandaan komen. We nemen aan dat ze er zijn, simpele basisregels voor het interpreteren van instructies en voor het maken van analogieën bijvoorbeeld, en stoppen ze eenvoudig in onze modellen. En als we dat doen, kunnen we leerprocessen met succes modelleren. Maar wat de status van die basisregels is, of je ze bijvoorbeeld zou kunnen leren of dat ze gegeven moeten zijn...'

Dat modelleren, dat gebeurt op computers. Speelt de metafoor van het brein als computer daarbij nog mee?

'De computermetafoor is in de cognitieve psychologie wel zo'n beetje dood, die was vooral populair rond 1970. Maar wat wel overeind blijft, is de computationele metafoor: dat je hersenen in zekere zin berekeningen uitvoeren. En we gebruiken de computer natuurlijk als platform om simulaties van onze theorieën te maken. Computersimulaties zijn op allerlei terreinen van belang, tot aan onderzoek naar het gedrag van individuele neuronen toe.'

Hoe leert zo'n simulatiemodel dan? Hoe houdt het zijn doel in de gaten, weet het dat het op de goede weg zit?

'Zo'n simulatie leert in principe net als mensen op drie manieren: van uitgewerkte voorbeelden, van instructies en door zelf te ontdekken. Nou hebben we meestal wel gewerkt met leersituaties waarin expliciete instructies erg domineerden, dus waar minder te ontdekken valt dan in de echte wereld. De modellen overbruggen zelf ook nog niet veel redeneerstappen, ze volgen voornamelijk precieze instructies, of een voorbeeld dat min of meer alle stappen aangeeft die moeten worden afgelegd op weg naar het doel.'

'Maar er is al wel iets gebeurd op het gebied van taken die wat verder gaan dan wat is voorgezegd. Zo heb ik met enig succes het proces kunnen modelleren dat mensen doormaken als ze zich een nieuwe algebraïsche bewerking moesten eigen maken.'

Speelt al dat beroemde werk aan schaakcomputers nog een rol?

'Nou, nee. Die programma's leunen toch nog altijd vooral op brute kracht. Miljoenen en miljoenen spellen opslaan en doorrekenen. Daar heb je niet veel aan, een mens moet het met een fractie daarvan doen.'

'We hebben hier wel een student gehad die een tric-trac-programma maakte dat binnen een paar duizend spelletjes redelijk leerde spelen – geen wereldniveau, maar mij kon het verslaan. Het werkte door uit zijn



Op 28 september 2006 ontvangt Anderson de Dr. A.H. Heinekenprijs voor de Cognitiewetenschap. Het beeldje is ontworpen door Don Staakman

geheugen situaties op te diepen die op de voorliggende bordstand leken, te kijken welke daarvan het best was afgelopen, en zo verder te spelen. Zo'n systeem leert, al is het wel een heel oppervlakkige vorm van "denken": niet meer dan tripletjes van toestand, handeling en gevolg opslaan en raadplegen. Instance Based Learning noemen we dat, en het blijkt een heel gebruikelijke manier om ACT-R-modellen in te richten die van ervaring moeten leren. Zo zijn bijvoorbeeld modellen gemaakt om een energiecentrale te runnen. Ik denk niet dat wat die modellen doen, lijkt op wat echte ingenieurs daar doen – ik hoop het althans niet – maar het lijkt wel erg op wat er gebeurt als we een proefpersoon diezelfde dingen laten leren.'

De computermetafoor is in de cognitieve psychologie wel zo'n beetje dood

Over studenten gesproken, er zijn toch ook praktische toepassingen van uw theorie?

'Ja, dat is wel een raar verhaal. Begin jaren tachtig leek het alsof de theorie zo'n beetje klaar was. Ik had geen idee wat er verder nog aan gedaan zou kunnen worden. Dus zocht ik een flink zware test om de zwakke punten op te sporen. En ik bedacht: als we nou studenten bij hun wiskundeonderwijs eens dwingen om precies de stappen te volgen die zo'n model zet, het model dus als strenge begeleider inzetten, dan zul je de boel wel uit elkaar zien spatten.'

'Maar wat bleek? Het werkte prima, die studenten leerden met zo'n zogenaamd Intelligent Tutoring System meetbaar beter en sneller. En zo was ik het grootste deel van de jaren tachtig bezig om van die systemen te bouwen.'

'Inmiddels is de theorie alweer mijlen verder, maar rond die Intelligent Tutoring Systems heeft de universiteit hier een bedrijf ingericht, Carnegie Learning, dat zegt jaarlijks 350.000 middelbareschoolleerlingen te bedienen.'

'Iets anders is dat we veel werk maken van het ontwikkelen van goede interfaces tussen perceptie en motoriek. Die kunnen veel tijd en geld besparen bij het ontwerpen en testen van software en machines waar mensen vlot mee moeten kunnen werken. Maar dat zijn onderzoeksprojecten, geen producten die zo in de winkel liggen.'

Een theorie is uiteraard nooit echt af, wat staat er op de verlanglijst?

'De brain-imagingtechnieken. Ik zou graag dieper zoeken naar de neurale basis voor de elementen in de ACT-R-theorie. Nogal wat mensen twijfelen aan het nut van al die plaatjes met opgluoiende plekje in de hersenen, maar ik denk dat ze een belangrijke bijdrage kunnen leveren. Als je alleen maar naar gedrag kijkt, is de afstand tussen theorie en wat je waarneemt wel heel groot. Dat maakt je theorie lastig toetsbaar. Hersenscans gunnen ons tenminste een glimp van wat er op dat traject van beginsituatie naar waarneembaar gedrag gebeurt. Zo vertellen ze je iets over of je met je theorie op de goede weg zit.'

'We hebben zelfs al eens op basis van scans een module gesplitst. De theorie bevatte een doelmodule die niet helemaal lekker in elkaar zat. Hij moest zowel bijhouden met welk deel van een complexe taak je bezig was, als tussenresultaten beheren. Scans lieten zien dat er in werkelijkheid twee ver uiteenliggende centra actief waren, en dat gaf de doorslag om de module in de theorie dan ook daadwerkelijk te splitsen.'

En, besluiten we onbescheiden, wat doet een cognitief psycholoog met een prijs van 150.000 dollar?

'O, jee. Ik heb ooit al eens de Rumelhardt Prize gekregen, dat was 100.000 dollar, en die staan geloof ik ergens op een bankrekening. Tot ongenoegen van mijn vrouw, trouwens, die wat moeite heeft met mijn stijl van geen geld uitgeven. Maar misschien... mijn oudste zoon gaat dit jaar rechten studeren, en laat dat in dit land nou precies anderhalve ton kosten. Zou dat nou geen mooie bestemming zijn?'