

Het proces ik

Het zelfbewustzijn als logisch mechanisme

De Academische Boekengids 65, november 2007, pp. 16-17.

Onze wil wordt bepaald door neuronale onderbouwing. Wij zijn vrij op het niveau van een schakcomputer die van zijn fouten leert.

In het recent verschenen *I Am a Strange Loop* richt hoogleraar cognitieve wetenschappen Douglas Hofstadter zich op de indringende vraag waar ons (persoonlijke) 'zelfbewustzijn' vandaan komt. Vaststaat dat dit belangrijke aspect van ons leven iets met ons brein te maken heeft: een beschadiging daarvan kan duidelijk merkbare gebreken veroorzaken, zoals verlammingen, het niet meer kunnen begrijpen of kunnen formuleren van taal of deze drie handicaps tezamen. Ook chemische beïnvloeding van het brein, door bijvoorbeeld koffie, alcohol of paddo's, kan kleine of grote veranderingen in dit zelfgevoel veroorzaken. We weten echter niet hoe het zelfbewustzijn neurofysiologisch gezien in onze hersenen functioneert.

Heeft het dualisme dan misschien toch gelijk, dat beweert dat wij een ziel hebben die buiten de fysische werkelijkheid staat en die op een of andere mysterieuze manier verbonden is met ons brein? En die misschien na onze dood verdergaat naar hemel of hel of naar een ander lichaam? Of is het zo dat het zelf en ons bewustzijn daarvan een functie is van onze neuronen, met hun voortdurend vuren van pulsen? In dat geval zouden we eigenlijk geen vrije wil kunnen hebben. Ons bewustzijn is dan namelijk geen oorzaak van ons lichamelijk gedrag maar een gevolg daarvan, dat eigenlijk slechts 'achter de neurale activiteit aan hobbelt'.

In navolging van zijn bestseller *Gödel, Escher, Bach: An Eternal Braid* (afgekort als *GEB*) uit 1979 geeft Hofstadter in zijn nieuwste boek antwoord op deze vragen in de vorm van een these. *GEB* was geschreven in een briljante opwelling van creativiteit waarbij hij de onvolledigheidsstelling van Gödel uitlegde aan de hand van oefeningen, analogieën, *Alice in Wonderland*-achtige dialogen, illustratieve werken van Escher en minder toepasselijke passages van Bach. Tegen het einde van het boek vinden we voor het eerst Hofstadters these. Maar omdat *GEB* zo dik was (ca. 750 pag.) hebben weinig personen begrepen waarover het gaat. De auteur heeft daarom nu een nieuw boek geschreven (bijna de helft in omvang), waarin zijn these niet verstopt is, maar prominent als titel fungeert: 'I Am a Strange Loop'.

'HET GEVOEL VAN "ZELF" IS IN EEN EINDIG SYSTEEM "PROGRAMMEERBAAR", DUS IS HET DENKBAAR DAT ER OOIIT MACHINES ZULLEN KOMEN MET EEN "ZELF".'

De baanbrekende onzekerheidsstelling van Gödel (1931) zegt iets over de axiomatische methode van Aristoteles in de wiskunde. Die stelt dat er begrippen en uitspraken zijn. De meeste begrippen kunnen worden gedefinieerd uit andere begrippen, maar er zijn een paar ongedefinieerde 'primitieve' begrippen waarmee je begint. Evenzo worden de meeste uitspraken op zuiver logische gronden bewezen, maar is er een verzameling onbewezen uitspraken bestaande uit de axioma's. Deze geven impliciet de betekenis aan van de primitieve begrippen. In een bepaalde versie van de rekenkunde, bijvoorbeeld, heb je als primitieve begrippen onder andere de getallen 0 en 1 en de operaties optellen en vermenigvuldigen. Er zijn eenvoudige axioma's voor deze begrippen, die de getallen en hun basisoperaties volledig lijken te beschrijven. Een uitspraak als

Ieder priemgetal van de vorm $4n + 1$ is de som van twee kwadraten. (*)

lijkt waar te zijn. Van de verzameling viervouden plus één

5, 9, 13, 17, 21, 25, 29, ...

zijn de vette leden priemgetallen (er zijn zelfs oneindig veel getallen van die vorm priem). En inderdaad geldt $5=1+4$, $13=4+9$, $17=1+16$, $29=4+25$; steeds de som van twee kwadraten. De uitspraak dat *alle* onderstreepte getallen, dat wil zeggen: priemgetallen van de vorm $4n + 1$, de som van twee kwadraten zijn, blijkt uit de axioma's van de rekenkunde bewezen te kunnen worden. Daarom weten we dat (*) *waar* is. De vraag dringt zich nu op of het volgende klopt.

Iedere ware uitspraak kan bewezen worden uit de rekenkundige axioma's.

Dit wordt de *volledigheid* van de rekenkunde genoemd. Volledigheid is alleen interessant als de rekenkunde consistent is - dat wil zeggen: geen contradicties bewijst zoals $0=1$ - anders zou iedere uitspraak, waar of onwaar, bewezen kunnen worden (uitgaande van $0=1$ stort het bouwwerk van de rekenkunde in elkaar). Iedere wiskundige gelooft dat de rekenkunde inderdaad consistent is: de axioma's zijn duidelijk waar en de logische afleidingsregels behouden waarheid. Gödel toont aan dat, uitgaande van de consistentie van de rekenkunde, er een ware rekenkundige uitspraak *G* is, die niet bewezen kan

worden. Met andere woorden: de rekenkunde is *onvolledig*. Gödel doet dit door een uitspraak *G* te construeren die uitdrukt

G is onbewijsbaar in de rekenkunde. (G)

Dit is opmerkelijk om drie redenen. Ten eerste drukken rekenkundige uitspraken doorgaans iets uit over getallen, maar deze uitspraak zegt iets over een uitspraak (zichzelf nota bene!). Ten tweede drukken rekenkundige uitspraken meestal iets uit in termen van optellen en vermenigvuldigen, en niet in termen van bewijsbaarheid. Ten derde is dit een uitspraak die naar zichzelf verwijst. Dat rekenkundige uitspraken iets kunnen uitdrukken over (andere) uitspraken wordt bereikt via het adagium van Pythagoras: 'Alles is een getal.' Gödel bereikte dit door uitspraken te coderen tot getallen. Dat kan op zo'n manier gebeuren dat het begrip 'bewijsbaar' uit te drukken is in termen van optellen en vermenigvuldigen. Dit is de belangrijkste stap in het werk van Gödel. Heeft men eenmaal zoveel uitdrukingskracht, dan is het niet heel moeilijk meer om de zin *G* met zelfverwijzing te construeren.

Als *G* eenmaal geconstrueerd is, dan volgt onvolledigheid als volgt. Stel *G* is bewijsbaar. Dan volgt uit de betekenis van *G* een contradictie (uitspraak *G* zegt immers dat deze niet te bewijzen is). Maar we hadden aangenomen dat de rekenkunde consistent is. Dus onze aanname (*G* is bewijsbaar) is onjuist. Oftewel *G* is onbewijsbaar. Maar daarmee is *G*, per definitie, wel waar (!). *G* is dus waar en onbewijsbaar. Deze redenering geldt niet alleen voor de rekenkunde maar voor ieder consistent axiomatisch systeem dat een uitbreiding van de rekenkunde vormt. Dit noemen we de *essentiële onvolledigheid* van de rekenkunde. Het wijst op een zwakte van de axiomatische methode. Maar een betere methode hebben we niet en bovendien is ze juist zeer sterk: voldoende om ons naar de maan te kunnen brengen (en helaas ook om ons naar de maan te helpen).

'HET IS NIET VREEMD DAT IEMAND NA ZIJN OVERLIJDEN VAAK IN JE GEEST
TERUGKEERT. DAT IS GEEN MAGISCHE KRACHT, MAAR GEWOON DE VOORTZETTING
VAN HET PROCES DAT REEDS IN JE BEGONNEN WAS.'

Naar aanleiding van de stelling van Gödel is Hofstadter tot de these gekomen van de *strange loop*. Een *strange loop* is een structuur die op een of andere manier (door een 'vertaling') een kopie van zichzelf bevat. Denk bijvoorbeeld aan de afbeelding op het bekende Droste-cacaobusje. De vertaling bestaat uit een verkleining van de afbeelding van de zuster op het echte busje die geplaatst wordt op het getekende busje. Zoals bij vele *strange loops* is hier een duidelijke hint naar een oneindige structuur. Escher heeft veel van dergelijke grafieken gemaakt, waarbij werkelijkheid en afbeelding door elkaar heen lopen.

In *GEB* (p. 709) poneert Hofstadter de volgende these:

'Een vergelijkbare niveauoverschrijding [als bij *strange loops*] creëert ons bijna onanalyseerbare gevoel van zelf.'

Omdat we *strange loops* kunnen uitdrukken met eindig veel symbolen in relatief eenvoudige systemen, zoals blijkt uit de constructie van Gödel, volgt hieruit dat het gevoel van 'zelf' in een eindig systeem 'programmeerbaar' is. Dus is het denkbaar dat er ooit machines zullen komen met een 'zelf'. In zijn nieuwste publicatie gaat Hofstadter zelfs zover dat hij het bewustzijn wil verklaren uit de door middel van de *strange loop* verwerkte perceptie van het 'ik'.

Hofstadter geeft de volgende overwegingen die tot zijn these hebben geleid of die eruit volgen. In de rekenkunde wordt 'betekenis' toegekend aan de getallen en aan de operaties optellen en vermenigvuldigen. Er zijn in deze theorie niet slechts een paar wiskundige begrippen, maar zeer vele, die willekeurig kunnen groeien tot iedere gewenste complexiteit. Uit het werk van Gödel blijkt dat het mogelijk is om uitspraken in de rekenkunde ook over die rekenkunde zelf te laten gaan. Zo kan een uitspraak via betekenisvolle getallen iets zeggen over andere uitspraken (bijvoorbeeld dat ze bewijsbaar zijn). Verder blijkt het mogelijk dat uitspraken iets over zichzelf zeggen.

Analoog ziet Hofstadter de menselijke geest, die vol ideeën zit over van alles, ook over zichzelf. En zo wordt het 'ik' geboren. Waar wiskundigen de zelfreferentiële uitspraak van Gödel vaak zien als een anomalie, is die voor Hofstadter juist interessant. De rekenkunde, ontworpen voor het optellen en vermenigvuldigen, is voldoende complex om betekenisvol over logica en bewijsbaarheid te kunnen spreken, en daardoor ook over zichzelf. Doordat het semantische spel herhaald kan worden, krijgt de rekenkunde een gelaagdheid in niveaus. Overigens hebben J. Paris en L. Harrington (1977) een zuiver wiskundige uitspraak gevonden die ook waar maar onbewijsbaar is, dus zonder *strange loops*.

Ook binnen onze lichaam-geestcombinatie is er een hiërarchie van niveaus. Wanneer we te weinig gegeten hebben, zal ons neurale niveau minder goed functioneren. Dat zal echter naar het hogere niveau doorgegeven worden als 'ik heb trek'. Op een ander betekenisniveau kan men dan gaan denken: 'We hebben wat te vieren, dus laten we uit eten gaan.'

Bij sommige filosofen en hersenonderzoekers bestaat het steriele beeld dat de inhoud van onze gedachten 'slechts' veroorzaakt wordt door de vurende neuronen, en altijd als gevolg, nooit als oorzaak. Hofstadter weerlegt deze stelling overtuigend. Stel, we hebben een schakeling van neuronen. Als altijd

gaan ze vuren wanneer ze voldoende input krijgen. De neuronen zijn zo geschakeld dat neuron 641 nooit gaat vuren. Op de vraag hoe dat komt, kun je het flauwe antwoord geven dat deze neuron niet de juiste input krijgt. Er is echter een betere manier om naar de zaak te kijken als je weet dat de neuronen zó geschakeld zijn dat ze worden overgeslagen als ze een priemgetal als nummer hebben (iets wat eenvoudig te realiseren is door de samengestelde getallen weg te selecteren). Het is nu veel informatiever te zeggen dat neuron 641 niet gaat vuren omdat het een priemgetal als nummer heeft. Dus een globaal idee, het begrip 'priemgetal', heeft wel degelijk als oorzaak dat neuron 641 overgeslagen wordt in het eliminatieproces. Dit idee is vormgegeven als een schakelingspatroon van neuronen.

Uiteraard hebben de wetten van de neuronen ook met dit alles te maken. Deze dienen echter om het idee te 'implementeren'. Dit is de reden waarom Hofstadter pleit tegen het eenzijdige reductionisme, dat de hogere causale patronen niet nodig denkt te hebben.

'IN HET Z-UNIVERSUM ZITTEN ONZE ZOMBIEKOLONEN, DIE NET ALS WIJ REAGEREN, MAAR VERDER GEEN BEWUSTZIJN HEBBEN: ALLES GEBEURT IN HET DONKER.'

Hofstadter is het eens met de filosoof Derek Parfit dat het 'ik' een illusie is. Op de vraag van een denkbeeldige tegenstander wie die illusie dan heeft, antwoordt hij: 'Het is een hallucinatie gehallucineerd door een hallucinatie.' Overigens voegt hij eraan toe: 'Maar het is wel een illusie die we nodig hebben en waar we niet onderuit kunnen. Deze illusie noemen we het idee "ik".'

In een aantal hoofdstukken beschrijft de auteur zijn rouwproces nadat zijn geliefde vrouw van 42 opeens kwam te overlijden. Als het 'ik' dan bestaat uit gedachten die andere gedachten aansturen, volgt daaruit dat een andere persoon in iemands mentale wereld kan doordringen. Meestal gebeurt dat wederzijds. Daarom is het niet vreemd dat iemand na zijn overlijden vaak in je geest terugkeert. Dat is geen magische kracht buiten je die binnenkomt, dat is gewoon de voortzetting van het proces dat reeds in je begonnen was.

Hofstadter maakt een tweede hypothese: bewustzijn bestaat uit denken vanuit het symbool 'ik'. Hier onderscheidt zijn visie zich van die van zijn voormalige promovendus David Chalmers. Deze laatste heeft als gedachte-experiment twee universa: het Q- en het Z-universum (respectievelijk voor 'Qualia' en 'Zombie'). Het Q-universum is zoals het onze, met zijn bewuste wezens zoals wij. In het Z-universum zitten onze zombiekolonies, die net als wij reageren, maar verder geen bewustzijn hebben: alles gebeurt in het donker. Hofstadter vindt dit beeld van Chalmers onbegrijpelijk en plaagt hem dat hij zoekt naar een *élan vital* (in navolging van het *élan vital* gepostuleerd toen men het mechanisme van het leven nog niet begreep). In een ander hoofdstuk noemt hij dit gekscherend het element feelium. Ook stelt hij dat dit leidt tot dualisme (naast de fysische zaken bestaat er een onafhankelijke geest), met alle filosofische problemen van dien (welke materiële zaken hebben wel bewustzijn, welke niet?). Hofstadter stelt dat vrije wil niet bestaat. Hij ziet er geen plaats voor en weet zelfs niet wat het betekent.

Het is opmerkelijk dat de Britse wis- en natuurkundige Roger Penrose (1994) ook een conclusie trekt uit de stelling van Gödel, maar wel de omgekeerde, namelijk dat ons bewustzijn niet mechanisch kan zijn. Zijn argument is dat Gödel bewijst dat uitspraak *G* waar is terwijl deze niet door een gegeven formeel systeem afgeleid kan worden. Als we een machine zouden zijn, dan zouden de stellingen die we kunnen afleiden door een gegeven formeel systeem gegenereerd kunnen worden. Voor de Gödel-zin voor dit systeem kunnen wij echter wél inzien dat deze waar is, dus zijn wij niet mechanisch, aldus Penrose. (De fout die Penrose maakt is gelegen in het feit dat Gödel een aanname moet doen om tot onvolledigheid te komen, namelijk de consistentie van de rekenkunde. In een tweede onvolledigheidsstelling concludeert Gödel dan ook dat die consistentie als uitspraak binnen de rekenkunde ook onbewijsbaar is en zelfs equivalent is met *G*.)

Hofstadter merkt op dat de Gödel-zin heel erg lang is. En dat dit er niet toe doet omdat het een theoretisch resultaat is. Zelf heb ik wel eens uitgerekend dat via de codering van Gödel zelf, zijn zin zo groot is dat deze niet in het bekende universum past (zulke boven-astronomische getallen komen wel vaker voor in de wiskunde en informatica). Dat zou Hofstadter zich moeten aantrekken, omdat zijn *strange loop* van het 'ik' dan niet in het brein zou passen. Er is echter een veel efficiëntere codering van *strange loops* in de logische theorie genaamd *lambda calculus*, en deze is verbazingwekkend kort.

'ER ONTSTAAT MEESTAL ANGST, OF ZELFS WALGING, WANNEER MEN DE VISCERALE NIVEAUS ONDER HET DENKEN ZIET: HET "IK" IS WEG EN ALLES LOOPT VANZELF.'

Met de hoofdthese van Hofstadter kan ik het eens zijn. 'De illusie van het "ik" ' zou je zo kunnen formuleren: het 'ik' bestaat als proces, niet als ding. Een golf die naar de kust lijkt te rollen, bestaat uit waterdeeltjes die relatief hoog staan en die die hoogte doorgeven aan de naastgelegen waterdeeltjes. Maar er beweegt geen enkel watermolecuul naar de kust toe. Wij hebben de gave en de neiging processen te reïficeren, ze tot ding te maken. Dat gebeurt bij een golf, bij ons lichaam en ook bij ons 'ik'.

Hofstadter stelt een aantal keren dat het niet mogelijk is om onze 'ik'-illusie uit te zetten. Als op een gegeven moment in het boek een denkbeeldige tegenspreker oppert dat boeddhistische meditatie dat wel als doel heeft, antwoordt de auteur: 'Dat is gedoemd tot falen.' In *GEB* stonden nog informatieve stukken

over de zenmeditatie met haar koans die het denken tegengaan. In dit nieuwe boek wil de auteur, naar zichzelf zegt, niet van het denken loskomen. In de waarneming vóór het denken zit echter belangrijke emotionele informatie besloten. Naast het model van de menselijke cognitie, waarin wij met het logisch verwerken van betekenisvolle symbolen in de wereld staan, is er nog een ander model, dat van het connectionisme. Dit stelt dat ons gedrag bepaald wordt door talloze neuronen waarvan we de betekenis niet altijd kunnen bepalen (noch wij als mens noch wetenschappelijk), maar waardoor we vaak wel de juiste dingen doen.

Dit aspect van het menselijk leven dat zich niet bevindt op het niveau van het denken, maar op dat van de directe waarneming, is ondervertegenwoordigd in het boek. Op dat niveau worden onze gevoelens en emoties gemaakt op een niet-symbolische manier. Had Hofstadter deze kant van het bewustzijn meegenomen, dan had hij wel een kijkje kunnen nemen in de viscerale niveaus onder het denken. Er ontstaat meestal angst, of zelfs walging, wanneer men deze niveaus ziet: het 'ik' is weg en alles loopt vanzelf, zonder bijdrage van een persoon. Deze 'angst voor de leegte' is echter wel te overwinnen.

De bewustzijnstoestand *mindfulness* komt uit de inzichtsmeditatie en wordt tegenwoordig veel gebruikt in de psychotherapie – Mark Williams en coauteurs, bijvoorbeeld, publiceerden recent hun successen in het boek *Mindfulness en bevrijding van depressie* (2007). *Mindfulness* is een vorm van bewustzijn waarbij de input tussen aanhalingstekens wordt geplaatst, een 'metabewustzijn'. Elders heb ik gesteld (1996) dat *mindfulness* de benodigde afbeelding is voor de *strange loop* van Hofstadter om ons 'ik'-gevoel te verklaren.

Ten aanzien van het Q- versus Z-universum heb ik de volgende hypothese: in eerste instantie zijn mensen in het Q-universum gelijkwaardig met die in het Z-universum. Wanneer men echter via meditatie of anderszins de leegte tegenkomt, dan kan de 'Q', de fundering van het bewustzijn, causaal werkzaam zijn om de angst voor de leegte te overwinnen.

'WIJ HEBBEN DE GAVE EN DE NEIGING PROCESSEN TOT DING TE MAKEN. DAT GEBEURT BIJ EEN GOLF, BIJ ONS LICHAAM EN OOK BIJ ONS "IK".'

Met de tweede these van Hofstadter ben ik het niet eens, al kan ik mijn standpunt, evenmin als hij, met argumenten onderbouwen. Het gevoelsmatige argument van een paar eeuwen geleden dat er een *élan vital* moet zijn om iets tot leven te brengen, was onjuist. We begrijpen de cellulaire mechanismen; het verschil tussen complexe, levenloze, chemische reacties en reacties met zelfduplicatie is niet wezenlijk. Ik kan me echter niet voorstellen dat hersenprocessen, ook al bevatten ze *strange loops*, op den duur tot het ons bekende bewustzijn zullen leiden. Maar misschien ben ik te gehecht aan ons bewustzijn en moet ik meer mediteren om in te zien dat we geen Q, *feelium* of *élan mental* nodig hebben.

Hofstadter heeft naar mijn mening gelijk dat ons willen gedetermineerd is door de neuronale onderbouwing. Maar op het andere niveau zijn wij het die de gedachten aansturen en eventueel met ons voelen vergelijken. Op dat niveau heerst vrijheid. Zeg maar de vrijheid van een schaakcomputer die niet iedere keer op dezelfde manier een tegenzet doet, maar die van zijn fouten leert, terwijl hij toch geheel geprogrammeerd is. Hofstadter stelt dat we onszelf zouden frustreren als we iets doen wat we niet willen. Hij geeft wel toe dat er een metamotivatie voor kan zijn. Dat is de reden waarom we soms onze wil kunnen uitzetten. Een hulpmiddel daarbij is weer de *mindfulness*. Deze werkt veel sterker dan het geweten en de moraal, zelfs sterk genoeg om onze vrees en begeerte te ontstijgen.

Voor de lezer die goed bekend is met de hypothese uit *GEB*, is *I Am a Strange Loop* soms een beetje langdradig. Toch vond ik het uiteindelijk de moeite waard om gelezen te hebben. Het biedt een goede inleiding in de kracht van de 'reflectie' techniek van Gödel. Deze werd wel eens gezien als onnatuurlijk. In de moderne wiskunde begint zij echter een steeds grotere rol aan te nemen, zoals in het geheel mechanisch geverifieerde formele bewijs van Georges Gonthier van de vierkleurenstelling (2005).

Het boek geeft een bepaalde visie op de *theory of mind*, zonder veel andere visies te noemen. Een uitstekend beknopt overzicht daarvan biedt *Conversations on Consciousness* (2005) van psychologe Susan Blackmore, bestaande uit interviews met filosofen en hersenonderzoekers over hun meningen aangaande het ongrijpbare bewustzijn.

Henk Barendregt is hoogleraar grondslagen van de wiskunde en informatica aan de Radboud Universiteit Nijmegen.

Besproken boeken:

I Am a Strange Loop
door **Douglas Hofstadter**
Basic Books. New York 2007.
412 + xix pag., € 26,30

Literatuur:

- **H. Barendregt** (1996). 'Mysticism and Beyond. Buddhist Phenomenology, Part II', *The Eastern Buddhist, New Series*, vol. XXIX, pp. 262-287 (ook beschikbaar op <http://www.cs.ru.nl/~henk/BP/bp2.html>).
- **H. Barendregt** (1997). 'The Impact of the Lambda Calculus in Logic and Computer Science', *The Bulletin of Symbolic Logic* 3(2): 181-215.
- **G. Gonthier** (2005). 'A Computer-Checked Proof of the Four Colour Theorem', beschikbaar op <http://research.microsoft.com/~gonthier/4colproof.pdf>.
- **J. Paris** en **L. Harrington** (1977). 'A Mathematical Incompleteness in Peano Arithmetic', in: J. Barwise (red.), *Handbook of Mathematical Logic*. Amsterdam: North-Holland, pp. 1133-1142.
- **R. Penrose** (1994). *Shadows of the Mind*. Oxford: Oxford University Press.
- **M. Williams et al.** (2007). *Mindfulness en bevrijding van depressie. Voorbij chronische ongelukkigheid*. Amsterdam: Uitgeverij Nieuwezijds.