

I am a strange loop. Douglas Hofstadter, Basic Books, 412 + xix pp.

Dit boek gaat over de indringende vraag waar ons (persoonlijke) ‘zelf bewustzijn’ vandaan komt. Dit belangrijke aspect van ons leven heeft zeker iets met ons brein te maken: beschadigingen van delen ervan kunnen duidelijk merkbare gebreken veroorzaken, zoals verlammingen, het niet meer kunnen begrijpen of kunnen formuleren van taal of al deze drie handicaps tezamen. Ook chemische beïnvloeding van het brein, door bijvoorbeeld koffie, alcohol of paddo’s, kunnen kleine of grote veranderingen in dit zelfgevoel veroorzaken. Neurofysiologisch is er echter onbekend hoe dat zelf bewustzijn in de hersenen functioneert. Heeft het dualisme dan toch gelijk dat beweert dat wij een ziel hebben die buiten de fysieke werkelijkheid staat en op een of andere mysterieuze manier verbonden is met ons brein? En die misschien na onze dood verder gaat naar hemel of hel of naar een ander lichaam? Of is het zo dat het zelf en ons bewustzijn een functie is van onze neuronen, met hun voortdurend vuren van pulsen. Dan zouden we eigenlijk geen vrije wil kunnen hebben en ons bewustzijn is dan geen oorzaak van ons lichaaamlijk gedrag maar een gevolg, dat eigenlijk maar ‘achter de neurale activiteit aan hobbelt’.

In navolging van zijn bestseller ‘Gödel Escher Bach’ (GEB) uit 1979 geeft Hofstadter antwoord op deze vragen in de vorm van een these. GEB was geschreven in een briljante opwelling van creativiteit waarbij de stelling van Gödel uitgelegd wordt met oefeningen, analogiën, *Alice in Wonderland*-achtige dialogen, illustratieve werken van Escher en minder toepasselijke passages van Bach. Tegen het einde van het boek stond Hofstadter’s these. Maar omdat GEB zo dik was (ca. 750 p) hebben weinig personen begrepen waarover het gaat en heeft de auteur een nieuw boek (ca. 400 p) geschreven, waarin zijn these niet verstopt, maar op de voorkant als titel staat: “Ik ben een *strange loop*”.

Gödel

De baanbrekende onvolledigheidsstelling van Gödel (1931) zegt iets over de axiomatische methode van Aristoteles in de wiskunde. Hierin wordt gesteld 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 er is een verzameling onbewezen uitspraken bestaande uit de *axioma*’s. Deze geven impliciet aan wat de betekenis is van de ongedefinieerde begrippen. In een bepaalde versie van de rekenkunde bijvoorbeeld heb je als primitieve begrippen o.a. de getallen 0 en 1 en de operaties optellen en vermenigvuldigen. Er zijn eenvoudige axioma’s voor deze begrippen, die de getallen en hun basis operaties volledig lijken te beschrijven. Een uitspraak als

$$\text{Ieder priemgetal van de vorm } 4n + 1 \text{ is de som van twee kwadraten} \quad (1)$$

lijkt waar te zijn. Van de verzameling viervouden plus een

$$\underline{5}, 9, \underline{13}, \underline{17}, 21, 25, \underline{29}, \dots$$

zijn de onderstreepte 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 is, blijkt uit de axioma's van de rekenkunde bewezen te kunnen worden. Daarom weten we dat (1) *waar* is. De vraag is nu of *iedere* ware uitspraak bewezen kan worden uit de rekenkundige axioma's. Dit wordt de volledigheid van de rekenkunde genoemd. Deze vraag is alleen interessant als de rekenkunde consistent is, d.w.z. 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 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 notabene!). Ten tweede drukken rekenkundige uitspraken iets uit in termen van optellen en vermenigvuldigen, en meestal 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." Met name kunnen we uitspraken coderen tot getallen. Dat kan op zo'n manier gebeuren dat het begrip 'bewijsbaar' uit te drukken is in termen van optellen en vermenigvuldigen. Dat 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 meteen. Stel G is bewijsbaar. Dan volgt uit de betekenis van G een contradictie (G zegt immers dat hij niet te bewijzen is). Maar we hadden aangenomen dat de rekenkunde consistent was. 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 systeem¹ dat een uitbreiding van de rekenkunde is. Dit wordt de *essentiële onvolledigheid* van de rekenkunde genoemd. Het wijst op een zwakte van de axiomatische methode. Maar een betere methode hebben we niet en bovendien is ze voldoende sterk om ons naar de maan te kunnen brengen (en helaas ook om ons naar de maan te helpen).

¹Waarbij de axioma's wel herkend moeten kunnen worden; anders zou je als axioma verzameling alle ware zinnen kunnen nemen.

Hofstadter's These

Een *strange loop* is een structuur welke op een of andere manier (door een 'vertaling') een kopie van zichzelf bevat. De afbeelding op het bekende Droste cacao-busje (waarop een zuster staat die het cacao-busje zelf op een dienblad draagt) is een voorbeeld van een *strange loop*. De vertaling bestaat uit een verkleining van de tekening van de zuster op het echte busje welke geplaatst wordt op het getekende busje.



Zoals bij vele *strange loops* is hier een duidelijk hint naar een oneindige structuur. Escher heeft veel van dergelijke grafieken, waarbij werkelijkheid en afbeelding door elkaar heen lopen.

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

“Ons gevoel van zelf wordt veroorzaakt door een *strange loop*.”

Omdat *strange loops* uitgedrukt kunnen worden 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 het huidige boek gaat Hofstadter zelfs zo ver dat hij het bewustzijn wil verklaren uit verwerkte perceptie door middel van de *strange loop* van het ‘ik’.

Gevolgen

Hofstadter geeft naar aanleiding van de stelling van Gödel de volgende overwegingen die tot zijn these hebben geleid of die er uit volgen.

Niveaus van betekenis. In de rekenkunde wordt ‘betekenis’ aan de getallen en de operaties optellen en vermenigvuldigen toe gekend waarmee adequaat om gegaan wordt. Er zijn niet slechts een paar wiskundige objecten met betekenis, 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 over andere uitspraken iets zeggen (bijvoorbeeld dat ze bewijsbaar zijn). Verder bleek het mogelijk dat uitspraken iets over zichzelf kunnen zeggen. Analooq ziet Hofstadter de menselijke geest, die vol ideeën zit over van alles, maar óók over zichzelf en zo wordt het ‘ik’ geboren. Wiskundigen zien de uitspraak van Gödel vaak als een anomalie². Maar voor Hofstadter is de zelfreferentiële uitspraak van Gödel juist interessant. De rekenkunde, ontworpen voor het optellen en vermenigvuldigen, is voldoende complex om zinvol over logica en bewijsbaarheid te spreken en daardoor over zichzelf. Zo krijgt de rekenkunde een gelaagdheid in niveaus omdat het semantische spel herhaald kan worden.

Ook binnen ons lichaam-geest combinatie 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 door gegeven worden als ‘ik heb trek’. Op een ander betekenis niveau kan men dan gaan denken: “We hebben wat te vieren, dus laten we uit eten gaan.”

Downward causation. Er is een steriel beeld van sommige filosofen en herenonderzoekers dat de inhoud van onze gedachten ‘slechts’ veroorzaakt wordt door de vurende neuronen, maar altijd als gevolg en nooit als oorzaak. Dit wordt door Hofstadter als volgt goed weerlegd. 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. Als je vraagt 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 wanneer je weet dat de neuronen zo geschakeld zijn dat als ze een priemgetal als nummer hebben, dan worden ze overgeslagen (iets wat eenvoudig te realiseren is door de samengestelde getallen weg te selecteren). Het is nu veel informatiever te

²Men is er in geslaagd de onvolledigheid aan te tonen via een uitspraak die alleen over wiskundige begrippen gaat, zie Paris & Harrington [1977]. Ook deze onafhankelijke uitspraak blijkt binnen de rekenkunde equivalent te zijn met die van Gödel.

zeggen dat neuron 641 niet gaat vuren omdat het een priemgetal als nummer heeft. Dus een globaal idee, het elimineren van samengestelde getallen, heeft wel degelijk als oorzaak dat 641 als priemgetal overgeslagen wordt in het eliminatie proces. De oorzaak ligt in het idee om neuronen met een priemgetal als nummer niet te laten vuren; en dit idee is vormgegeven als schakelingspatroon van neuronen. Uiteraard hebben de wetten van de neuronen ook met dit alles te maken. Deze dienen echter om het onderhavige idee te ‘implementeren’. Dit is de reden dat Hofstadter pleit tegen het eenzijdige reductionisme, dat de hogere causale patronen niet denkt nodig te hebben.

Ik en anderen. Met de filosoof Parfit is Hofstadter het eens 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, dan volgt dat een andere persoon in iemands mentale wereld door kan dringen. Vaak gebeurt dat wederzijds. Daarom is het niet vreemd dat na overlijden iemand 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.

Bewustzijn en vrije wil. Hofstadter maakt een tweede hypothese: bewustzijn bestaat uit denken vanuit het symbool ‘ik’. Hier onderscheidt zijn visie zich met die van zijn voormalige promovendus David Chalmers. Deze laatste heeft als gedachte experiment twee universa: het Q- en het Z-universum (respectievelijk ‘qualia’ en ‘zombie’). Het Q-universum is zoals het onze met zijn bewuste wezens zoals wij. In het Z-universum zitten onze klonen, 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 *élan mental* (in navolging van het *élan vital* gepostuleerd toen men het mechanisme van het leven nog niet begreep). In een ander hoofdstuk wordt dit gekscherend het element ‘*feelium*’ genoemd. Ook wordt gesteld dat dit leidt tot dualisme (naast de fysische zaken bestaat er een onafhankelijke geest), met alle filosofische problemen van dien (welke materiele 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.

Commentaar

Hierbij een aantal aanvullingen op het boek vanuit de visie van de recensent.

Penrose. Het is opmerkelijk dat fysicus en auteur Penrose [1994] ook conclusies trekt uit de stelling van Gödel en wel de omgekeerde dat ons bewustzijn niet mechanisch kan zijn. Penrose’s argument is dat Gödel bewijst dat de uitspraak *G* waar is terwijl deze niet door een gegeven formeel systeem afgeleid kan worden. Als wij 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 Penrose³.”

Lambda calculus. 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 meer efficiënte codering van *strange loops* in de logische theorie genaamd lambda calculus en deze zijn verbazingwekkend klein, zie Barendregt [1997].

Reïficeren. Met de hoofdthese van Hofstadter kan ik het eens zijn. ‘De illusie van het ‘ik’ zou ik zo 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 hoogte doorgeven aan de naastgelegen waterdeeltjes. Maar er beweegt geen enkel water molecuul naar de kust toe. Wij hebben de gave en de neiging processen te *reïficeren*, ze tot ding te maken. Dat gebeurt bij de golf, bij ons lichaam en ook bij ons ‘ik’.

Meditatie. Hofstadter stelt een aantal keer 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, dan antwoordt de auteur: “Dat is gedoemd tot falen” (p. 295). In GEB stonden nog informatieve stukken over Zen meditatie met zijn Koans die het denken tegen gaan. In het onderhavige boek wil de auteur, naar hij zelf 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 men met het sturen van symbolen in de wereld staat is er nog een ander model, dat van het connectionisme. Dit model stelt dat ons gedrag bepaald wordt door talloze neuronen, waarvan we de betekenis niet altijd kunnen bepalen (noch wij als mens noch wetenschappelijk).

Dit aspect van het menselijke leven dat zich niet op het niveau van het denken, maar van het directe waarnemen bevindt, is ondervertegenwoordigd in het boek. Daar 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 vanuit zichzelf. Deze ‘angst voor de leegte’ is echter wel te overwinnen.

Mindfulness. Deze bewustzijnstoestand komt uit de inzichtsmeditatie en wordt tegenwoordig veel gebruikt in de psychotherapie, zie Williams et al.

³Naar mijn mening en die van andere logici (zie Feferman [1995]) vergist Penrose zich. 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 onbewijsbaar is en feitelijk binnen de rekenkunde equivalent is met G .

[2007]. Het is een vorm van bewustzijn, waarin de input tussen aanhalingstekens wordt geplaatst, een meta-bewustzijn. In Barendregt [1996] stel ik dat mindfulness de benodigde afbeelding is voor de *strange loop* van Hofstadter om ons ‘ik’-gevoel te verklaren.

Q-universum. Hier is een hypothese van de recensent. In eerste instantie zijn mensen in het Q-universum gelijkwaardig met die in het Z-universum. Wanneer men echter via meditatie of anderzinds de leegte tegen komt, dan kan de ‘Q’, de fundering van het bewustzijn, causaal werkzaam zijn om de angst voor de leegte te overwinnen.

Met de tweede these van Hofstadter ben ik het niet eens, al kan ik mijn standpunt, net zoals hij, niet met argumenten onderbouwen. Het gevoelsmatige argument van een paar eeuw geleden dat er een *élan vital* moet zijn om iets tot leven te brengen is onjuist. We begrijpen de cellulaire mechanismes. Het verschil tussen complexe levenloze chemische reactisch en reactisch met zelf-duplicatie is niet wezenlijk. Ik kan me niet voorstellen dat hersenprocessen, ook al bevatten ze strange loops, op den duur tot het ons bekende bewustzijn zal leiden. Maar misschien ben ik te gehecht aan ons bewustzijn en moet ik meer mediteren.

Vrije wil. 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 de zelfde 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 dat we niet willen. Hij geeft wel toe dat er een meta-motivatie voor kan zijn. Dat is de reden dat 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.

Tot slot. Hoewel ik het boek soms een beetje langdradig vond (ik kende de hypothese uit GEB goed) was het uiteindelijk toch de moeite waard om het gelezen te hebben. Het is een goede inleiding in de kracht van de ‘reflectie’ techniek van Gödel. Deze werd wel eens gezien als onnatuurlijk. In moderne wiskunde begint het echter een steeds grotere rol aan te nemen, zoals in het geheel mechanisch geverifieerde bewijs van de vier-kleurenstelling, Gonthier [2005].

Het boek geeft een bepaalde visie op de *theory of mind*, zonder veel andere visies te noemen. Een uitstekend beknopt overzicht is Blackmore [2005], bestaande uit interviews met filosofen en hersenonderzoekers over hun meningen aangaande het ongrijpbare bewustzijn.

Henk Barendregt
Radboud Universiteit Nijmegen

Literatuur

- Barendregt, H. 1996: *Mysticism and beyond, Buddhist phenomenology, part II, The Eastern Buddhist, New Series, vol XXIX, pp. 262-287, ook beschikbaar op* www.cs.ru.nl/~henk/BP/bp2.html.
- Barendregt, H. 1997: The impact of the lambda calculus in logic and computer science. *The bulletin of symbolic logic, Vol. 3 Iss. 2 pp. 181-215.*
- Barendregt, H. 2006: Bewustzijn: geloof en hypothese. In: *Wat we geloven maar niet kunnen bewijzen. De belangrijkste denkers van nu over wat zij geloven maar nog niet kunnen bewijzen*, Red. E. J. Richter, Spectrum, Utrecht, pp. 201-205. Ook beschikbaar op ftp://ftp.cs.ru.nl/pub/CompMath.Found/geloof_hypothese.pdf.
- Blackmore, S. 2005: *Conversations on Consciousness*, Oxford University Press.
- Feferman, S. 1995: Penrose's Gödelian Argument, in: *Psyche*, beschikbaar op psyche.cs.monash.edu.au/v2/psyche-2-07-feferman.html.
- Gonthier, G. 2005: A computer-checked proof of the Four Colour Theorem, beschikbaar op research.microsoft.com/~gonthier/4colproof.pdf.
- Penrose, R. 1994: *Shadows of the Mind*. Oxford University Press.
- Paris, J. & J. Harington 1977: A mathematical incompleteness in Peano arithmetic, *Handbook of Mathematical Logic* (J. Barwise, ed.), North-Holland, Amsterdam, pp. 1133-1142.
- Williams, M., J. D. Teasdale, Z. V. Segal & J. Kabat-Zinn 2007: *Mindfulness en bevrijding van depressie, voorbij chronische ongelukkigheid*. Uitgeverij Nieuwezijds, Amsterdam.