6.1 Gödel en de de Universal Truth Machine

In de filosofische discussies heeft de onvolledigheidsstelling soms betrekking op machines en soms op formele systemen. Deze twee noties zijn equivalent in zeker opzicht omdat er voor elk formeel systeem er een programma te schrijven is dat systematisch alle stellingen van het systeem genereert. Andersom is er voor elk programma die zinnen genereert in een formele taal een corresponderend formeel systeem dat deze zinnen als stellingen heeft.

Eén van de grootste misvattingen over de eerste onvolledigheidsstelling is dat Gödel's bewijs ervan, als het toegepast wordt op een consistent systeem, laat zien dat de Gödel zin van het systeem, die onbewijsbaar is in het systeem zelf, waar is. Een voorbeeld van deze misvatting is de redenatie van Rudy Rucker:

1. *We presenteren aan een UTM, een apparaat dat een universele waarheidsmachine hoort voor te stellen, de volgende Gödelzin G: "De machine die geconstrueerd is op basis van het programma P(UTM) zal nooit zeggen dat deze zin waar is". G is equivalent aan "UTM zal nooit zeggen dat G waar is".*

2. *Vervolgens vragen we aan de UTM of de zin G waar of onwaar is.*

3. *Als de UTM zegt dat G waar is, dan is "UTM zal nooit zeggen dat G waar is" onwaar. Als "UTM zal nooit zeggen dat G waar is" onwaar is, dan is G onwaar. Dus als UTM zegt dat G waar is, dan is G in feite onwaar, en heeft de UTM een onware uitspraak gedaan. Aangezien de UTM alleen ware uitspraken doet, zal deze dus nooit zeggen dat G waar is.*

4. *Hiermee hebben we vastgesteld dat de UTM nooit zal zeggen dat G waar is. Dus is "UTM zal nooit zeggen dat G waar is" in feite een ware uitspraak. Dus is G waar.*

5. *Hiermee hebben we een waarheid die de UTM nooit zal uitspreken, maar waarvan wij wel de waarheid inzien. Om deze reden is de UTM is dus niet echt universeel.*

Hiermee wordt echter niet bewezen dat er inderdaad een waarheid is die de UTM niet kan uitspreken. Want het enige dat we hebben is de implicatie dat "als de UTM altijd de waarheid vertelt, dan is G waar". Deze implicatie kan ook door de UTM uitgesproken worden. We hebben geen kennis van of de UTM inderdaad altijd de waarheid spreekt. Dat de UTM een universele waarheidsmachine hoort voor te stellen draagt niet bij aan het weten dat de UTM altijd de waarheid spreekt.

Het argument van Lucas, om te laten zien dat er voor elk consistent formeel systeem S, er een ware stelling is die wij kunnen bewijzen maar S niet, is om dezelfde reden ongeldig. Lucas gaat er onterecht van uit dat de Gödels stelling stelt dat in elk consistent formeel systeem dat krachtig genoeg is om simpele rekenkunde te produceren er formules bestaan die niet bewezen kunnen worden in het systeem, maar waarvan wij de waarheid wel kunnen inzien. Dit wordt echter helemaal niet gesteld door de onvolledigheidsstelling. Op basis van het bewijs van Gödel's onvolledigheid weten we niet dat de Gödel-zin G voor een theorie S waar is, maar dat slechts de implicatie “als S consistent is, dan is G waar” waar is. Aangezien deze implicatie ook bewijsbaar is in S, is er niets in Gödel's bewijs dat stelt dat wij meer weten dan dat bewezen kan worden in S (tot zover het rekenkunde aangaat). Pas als we weten of een theorie S consistent is, kennen we ook de waarheid van de stelling die niet bewijsbaar is in S. Maar zolang we dat niet weten, kennen we de waarheid van de Gödel-zin G voor S ook niet.

De zwakkere claim dat er geen enkel formeel systeem kan bestaan dat de menselijke geest exact representeert als het om de vaardigheid van het bewijzen van rekenkundige theorema's gaat, kan op dezelfde gronden ook weerlegd worden. Stel we nemen een theorie T, bijvoorbeeld ZFC met het axioma van oneindigheid toegevoegd. Dan wordt er nergens in de onvolledigheidsstelling uitgesloten dat elk rekenkundige uitspraak bewijsbaar in T ook bewezen kan worden door de menselijke geest. Net zoals dat er ook nergens de mogelijkheid uitgesloten wordt dat elke rekenkundige uitspraak bewijsbaar door de menselijke geest bewijsbaar is in T.

6.2 Penrose's “Tweede Argument”

Penrose's tweede argument heeft net zoals Lucas' argument tot doelt te laten zien, dat geen enkel machine de vaardigheid van de menselijke geest om rekenkundige theorema's te bewijzen exact kan representeren. De argumentatie gaat als volgt:

Gegeven een formeel systeem F, maakt Penrose gebruik van een corresponderend wiskundige uitspraak IAMF zodanig dat:

1. *Als IAMF dan F + IAMF is consistent.*
2. *Ik kan bewijzen: als IAMF dan F + IAMF is consistent.*
3. *Als IAMF dan voor elke A, als ik A kan bewijzen dan bewijst F, A.*

Uit 2 en 3 volgt dat als IAMF dan F bewijst “Als IAMF dan F + IAMF is consistent”. Maar dan bewijst F + IAMF dat F + IAMF consistent is, dus is F + IAMF inconsistent. Dit is onverenigbaar met IAMF wegens 1. Dus als IAMF waar is, dan is het onwaar, dus is IAMF onwaar. Niet-IAMF volgt uit 1-3 voor elke F waarop de onvolledigheidsstelling van toepassing is. Maar voor het argument is het alleen zinvol als we een IAMF kunnen vinden zodanig dat we kunnen stellen:

4. *Als ik F ben, in de zin dat F alle menselijk toegankelijke methoden van wiskundige bewijsvoeringen bevat, dan IAMF.*

De vraag is nu of er een IAMF gevonden kan worden waarvoor 1-4 geldt. Penrose stelt dat het inderdaad mogelijk is, echter wordt er in het boek gesteld dat dit een optimistische zienswijze is.

Penrose faalt in het specificeren dat er een IAMF is waarvoor 1-4 geldt. Echter is voor Penrose het meest overtuigende argument ons onvermogen om onze rekenkundige kennis volledig te specificeren.

6.3 Onuitputbaarheid opnieuw beschouwd

Als we een formeel systeem S als een correcte formalisatie van een deel van onze wiskundige kennis accepteren, dan doe we dat op dezelfde manier met dezelfde rechtvaardiging, voor een uitbreiding van dat systeem, die we verkrijgen door een nieuw axioma "S is consistent" toe te voegen. Aangezien het resulterende systeem logisch gezien sterken is dan S, kunnen we concluderen dat we een formeel systeem S kunnen specificeren dat onze wiskundige kennis uitput. Penrose formuleert een equivalente observatie die bekend staat als "Conclusie G" namelijk: "Menselijke wiskundigen gebruiken geen kenbaar correct algoritme om wiskundige waarheden na te gaan". Conclusie G stelt dus dat de menselijke geest non algoritmisch is op een bepaalde manier. De vraag is of het laat zien dat het in feite onmogelijk is voor ons om een computerprogramma te schrijven dat onze vaardigheid van het bewijzen van rekenkundige stellingen representeert.

Op het eerste gezicht lijkt het duidelijk, aangezien conclusie G precies stelt dat we geen formeel systeem kunnen specificeren dat onze wiskundige kennis kan uitputten. Maar er is helemaal geen reden om aan te nemen dat het programmeren van een computer met al onze vaardigheid om wiskundige theorema's te bewijzen, in zo'n formeel systeem zou moeten resulteren. Om de menselijke wiskundige te emuleren, moet de computer ook precies dezelfde redeneervermogens toepassen, waarin wij afleiden dat het accepteren van een formeel systeem T als wiskundig correct leidt tot het op dezelfde manier accepteren van de correctheid van een sterker systeem.

Penrose stelt echter dat het niet uitmaakt met welke redeneerregels we de computer programmeren, de totaliteit van uitspraken die bewijsbaar is door de computer zullen nog steeds theorema's zijn van een formeel systeem dat wij als correct herkennen. Conclusie G geldt dan nog steeds: de computer is niet geprogrammeerd met de totale som van al onze rekenkundige kennis en theorema-bewijscapaciteit.

Dit argument is echter niet overtuigend, omdat het soort redeneren wat ons in staat stelt om te accepteren dat er een sterker systeem is dat wiskundig correct is, gebruik maakt van een breed scala aan formele en informele principes, waarvan sommige evident en niet problematisch zijn, terwijl andere dit veel minder zijn. Als we proberen om sterkere principes te formuleren voor het uitbreiden van een correcte theorie, worden we namelijk geconfronteerd met vragen over wat wel of niet evident is. Vragen waar velen over zullen zeggen dat er geen definitief antwoord op is. Als het ons dan lukt om al onze redeneerprincipes in een computerprogramma te emuleren, hebben we nog steeds geen gronden waarop we kunnen stellen dat wij als menselijke wiskundigen iets kunnen bewijzen wat de computer niet kan.

6.4 De eigen geest begrijpen.

Vaak wordt er gesteld, dat volgens Gödel's onvolledigheidsstelling, het begrijpen van onze geest onmogelijk is. Echter is dit net zoals veel verwijzingen naar de onvolledigheidsstellingen buiten de logica en wiskunde simpelweg onjuist. Gödel's stelling stelt nergens of impliceert dat het begrijpen van de menselijke geest onmogelijk is. Vaak worden uitspraken zoals hierboven geïnspireerd door de onvolledigheidsstelling in plaats van dat uitspraken hieruit afgeleid worden. Vaak worden er in verwijzingen gerefereerd naar de tweede onvolledigheidsstelling. De onmogelijkheid van een formeel systeem S om zijn eigen consistentie te bewijzen wordt geïnterpreteerd als een onmogelijkheid van S om zichzelf voldoende te analyseren en rechtvaardigen. Het systeem begrijpt "zichzelf niet".

We kunnen hier tegen inbrengen dat de metafoor "de moeilijkheid voor een systeem om zijn eigen consistentie te bewijzen" afzwakt. Een systeem kan namelijk los van vragen over analyseren en rechtvaardigen niet echt zijn eigen consistentie postuleren, hoewel andere systemen wel echt de consistentie van dat systeem kunnen postuleren. (Dit is vergelijkbaar met de mens die nooit echt kan zeggen dat hij nooit over zichzelf praat, omdat het uitspreken een uitspraak "ik praat nooit over mezelf" het falsifieert.).

Echter als we de menselijke geest proberen te vergelijken met een formeel systeem als PA en ZFC, komen we tot de conclusie dat als de menselijke geest enigszins lijkt op deze formele systemen, het uitstekend in staat zal zijn om zichzelf te begrijpen.

Voor elke eindige deelverzameling van de axioma's van PA, bewijst PA de consistentie van die deelverzameling, door de logica van zo'n eindig deel van zichzelf te analyseren. Verder bewijst PA

> *Voor elk eindige deelverzameling van mijn axioma's, kan ik de consistentie van die deel verzameling bewijzen.*
>
> *Als elk eindige deelverzameling van mijn axioma's consistent is, dan ben ik consistent.*

Om paradoxen te voorkomen, kan PA echter niet stellen "Als ik de consistentie van een eindig deelverzameling van mijn axioma's kan bewijzen, dan is die deelverzameling consistent". PA kan wel dit feit over zijn eigen onvermogen om consistent dit principe te bevestigen. Dat wil zeggen PA bewijst

Als ik kan bewijzen "Voor elk eindige deelverzameling M van mijn axioma's, als ik kan bewijzen dat M consistent is dan is M consistent" dan ben ik consistent.

Uit deze indrukwekkende vaardigheid van PA om zichzelf te begrijpen, kunnen we concluderen, dat als de menselijke geest enigszins de zelfanalyse kracht van PA of ZFC bevat, dan kunnen we verwachten dat de menselijke geest uitstekend in staat zal zijn om zichzelf te begrijpen.