GÖDELS DISJUNCTIE

door L. HORSTEN (Leuven)

1. INLEIDING

In 1951 gaf Kurt Gödel de vijfentwintigste Joshuah Willard Gibbs-lezing. Gödel was de eerste en tot nu toe de enige logicus aan wie de eer te beurt viel om deze prestigieuze lezing te geven. Zijn lezing droeg als titel *Some Basic Theorems on the Foundations of Mathematics and Their Implications.* Hoewel het publiek vooral bestond uit wiskundigen en natuurkundigen was de inhoud van zijn lezing erg filosofisch.

Het was de bedoeling dat deze lezing in het *Bulletin of the American Mathematical Society* gepubliceerd zou worden, maar dat is er niet van gekomen. Dit is niet echt verwonderlijk, want Gödel was erg terughoudend met het publiek maken van zijn filosofische overtuigingen, ofschoon hij zelf aan zijn filosofische positie groot belang hechtte. Toch circuleren er in kringen van de filosofie van de wiskunde al geruime tijd versies van deze lezing. De inhoud ervan is dan ook vrij goed gekend. In 1995 werd Gödels lezing uiteindelijk officieel gepubliceerd in het derde volume van Gödels verzamelde werken, voorzien van een inleiding geschreven door George Boolos[1].

Leon HORSTEN (1966) is postdoctoraal onderzoeker van het F.W.O. Vlaanderen en deeltijds docent aan het Hoger Instituut voor Wijsbegeerte te Leuven.

1. Bij verwijzingen naar Gödels artikels baseer ik me steeds op de paginanummers in de verschillende volumes van Gödels *Collected Works.*

Gödels Gibbs-lezing bestaat uit twee delen. In het eerste deel tracht Gödel uit een logische stelling die hij in het begin van de jaren dertig had bewezen (de zogenaamde tweede onvolledigheidsstelling) een disjunctieve filosofische stelling af te leiden. Deze filosofische stelling zegt dat *ofwel* de menselijke geest niet gerepresenteerd kan worden als een digitale computer, *ofwel* er absoluut onbeslisbare wiskundige uitspraken bestaan. Het tweede deel van zijn lezing wijdt Gödel aan de metafysische implicaties van zijn disjunctieve stelling. Hij tracht hierin aan te tonen dat zijn disjunctieve stelling 'anti-materialistische' gevolgen heeft, en dat ze een platonistische positie in de filosofie van de wiskunde plausibel maakt. Op de argumenten en beschouwingen uit het tweede deel van de Gibbs-lezing zal ik niet ingaan. Ik zal me vooral concentreren op Gödels argument voor zijn disjunctieve stelling en op de vraag naar de plausibiliteit van elk van de disjuncten van deze stelling.

De structuur van dit artikel is als volgt. Eerst wordt een beschrijving gegeven van Gödels argument voor zijn disjunctieve stelling. Gödel vertrekt hierbij van twee verschillende fenomenen die beide aantonen dat de klasse van ware wiskundige uitspraken niet exhaustief gevat kan worden in een axiomasysteem en dat de wiskunde in die zin dus onuitputtelijk is. Deze fenomenen worden toegelicht in de tweede paragraaf van dit artikel. Vervolgens gebruikt Gödel het tweede van deze fenomenen om zijn disjunctieve stelling af te leiden. Gödels argument voor zijn disjunctieve filosofische stelling wordt beschreven in sectie 3. Vervolgens wordt in sectie 4 Gödels argument kritisch op zijn geldigheid onderzocht. Ik zal tot het besluit komen dat, indien aan de filosofische stelling en aan Gödels argument een aantal kwalificaties mogen worden toegekend, er aan de disjunctieve stelling uiteindelijk niet te ontkomen valt. In sectie 5 worden de twee disjuncten van Gödels stelling op zich bekeken. Het zal dan blijken dat het eerste disjunct op dit moment nog volledig onbeslist is, terwijl er goede redenen zijn om te denken dat het tweede disjunct van Gödels filosofische stelling waar is.

2. De onuitputtelijkheid van de wiskunde

Gödel vertrekt van twee argumenten voor de stelling dat de klasse van wiskundige waarheden in zekere zin onuitputtelijk is.

Het eerste argument heeft betrekking op de verzamelingenleer. De gangbare axioma's van de verzamelingenleer bouwen het verzamelingentheoretisch universum stap voor stap op. Men kan zich dit proces ongeveer als volgt voorstellen. We beginnen met de verzameling N van de natuurlijke getallen, die we met behulp van verzamelingentheoretische axioma's kunnen beschrijven. Vervolgens nemen we alle deelverzamelingen van de verzameling van natuurlijke getallen en vatten die in een verzameling. De resulterende verzameling P(N) van alle verzamelingen van natuurlijke getallen kunnen we weer met axioma's beschrijven. Dan nemen we de verzameling van alle deelverzamelingen van de natuurlijke getallen, P(P(N)), beschrijven die met axioma's, en gaan zo door. We passen steeds opnieuw de machtsverzameling toe, zoveel keer als er natuurlijke getallen zijn. Het resultaat hiervan vatten we weer in een verzameling, die we $P^{\omega}(N)$ noemen en weer beschrijven door axioma's. $P^{\omega}(N)$ bevat veel meer elementen dan N. Nu gaan we de machtsverzameling-operatie itereren op $P^{\omega}(N)$, zoveel keer als er elementen in $P^{\omega}(N)$ zijn, en het resultaat weer beschrijven met axioma's. Enzovoort. Het is duidelijk dat dit proces nooit tot een einde komt, en dat corresponderend daarmee het postuleren van nieuwe axioma's nooit tot een einde komt. Dit is een eerste uitdrukking van de onuitputtelijkheid van de wiskunde.

Het tweede argument heeft enkel betrekking op de natuurlijke getallen en de basisbewerkingen van optelling en vermenigvuldiging die op deze structuur gedefinieerd zijn. Als men een stel axioma's heeft die deze structuur tenminste tot op zekere hoogte correct beschrijven, dan zullen er *altijd* ware rekenkundige uitspraken zijn die op basis van die axioma's niet bewijsbaar zijn. In het bijzonder is het zo dat er voor elk axiomasysteem een rekenkundige uitdrukking bestaat die 'uitdrukt' dat het axiomasysteem consistent is. De uitspraak dat een axiomasysteem consistent is, is immers een vrij elementaire combinatorische propositie. Ze kan als volgt worden uitgedrukt: Er is geen eindige reeks van zinnen, zodat elke zin ofwel een axioma van het systeem is, ofwel uit eerdere zinnen in de reeks volgt door modus ponens en waarvan de laatste zin van de vorm $A \wedge \neg A$ is. Door de symbolen, uitdrukkingen en bewijzen van het systeem te coderen als natuurlijke getallen kan deze eindige combinatorische vraag als een getallentheoretisch probleem uitgedrukt worden[2]. Voor elk consistent axiomasysteem dat de structuur van de natuurlijke getallen tenminste tot op zekere hoogte beschrijft, is de getallentheoretische uitspraak die de consistentie van het systeem codeert niet bewijsbaar op basis van de axioma's van het systeem[3]. Zelfs de structuur van de natuurlijke getallen is dus niet exhaustief vatbaar door axioma's, in de zin dat in elk axiomasysteem bepaalde getallentheoretische waarheden onbewijsbaar zullen zijn. Dit is een ruwe uitdrukking van de tweede onvolledigheidsstelling, die Gödel in 1931 bewees. Later werd duidelijk dat deze stelling een bijzonder geval was van een meer algemene stelling die zegt dat (onder erg algemeen geldige randvoorwaarden) een axiomatisch systeem zijn eigen *correctheid* niet kan bewijzen. Dit wil zeggen dat onder erg algemeen geldige voorwaarden een axiomatisch systeem T de (getallentheoretische codering van de) volgende schematische uitspraak niet kan bewijzen:

Als de zin „ ..." bewijsbaar is in theorie T, dan ...,

waarbij ... ingevuld mag worden door een willekeurige uitspraak uit de formele taal waarin de axiomatische theorie T geformuleerd is. Dit is zelfs zo wanneer deze schematische uitspraak *waar* is![4]

2. Het inzicht dat uitspraken zoals de consistentiebewering voor een axiomatisch systeem gecodeerd kunnen worden als getallentheoretische uitspraken maakt de kern uit van de bewijzen van de onvolledigheidsstellingen.

3. Toch is dit een uitspraak met een lage recursie-theoretische complexiteit. Het is een zogenaamde π_1^0-zin.

4. Hieruit blijkt dat er meer dan twee onvolledigheidsstellingen zijn. Een gedetailleerd overzicht van de veelheid van onvolledigheidsstellingen die tegenwoordig bekend zijn vindt men in KRIPKE, 1995 (zie literatuuropgave p. 105).

Deze twee argumenten tonen volgens Gödel aan dat de klasse van wiskundige waarheden niet te vatten is in een axiomatisch systeem en in die zin dus onuitputtelijk is.

3. Gödels disjunctie

Naast de notie van bewijsbaarheid in een axiomatisch systeem is er het intuïtieve, informele begrip van wiskundige bewijsbaarheid. Laten we dit het *absolute* begrip van wiskundige bewijsbaarheid noemen. De klasse van absoluut wiskundig bewijsbare zinnen bestaat uit alle zinnen die bewezen kunnen worden vanuit correcte bewijsprincipes (axioma's en afleidingsregels) die de wiskundige gemeenschap ooit zal bezitten. Hierbij wordt abstractie gemaakt van alle praktische beperkingen. Stel bijvoorbeeld dat een wiskundige uitspraak A bewijsbaar is uit de wiskundige axioma's en afleidingsregels waarvan de correctheid in het jaar 2110 door de wiskundige gemeenschap erkend wordt, maar dat het kortste bewijs van A meer symbolen bevat dan er elementaire deeltjes in het universum zijn. Dan zal het wellicht praktisch onmogelijk zijn om ooit een bewijs van A te produceren. Toch is A een absoluut bewijsbare wiskundige uitspraak.

Men kan zich dan afvragen of de klasse van absoluut bewijsbare wiskundige uitspraken samenvalt met de klasse van wiskundige waarheden en of de klasse van absoluut wiskundig bewijsbare uitspraken ook in zekere zin onuitputtelijk is. Gödel denkt dat hier een aantal technische resultaten weer relevant is, in het bijzonder zijn tweede onvolledigheidsstelling. Maar de filosofische conclusie zal ditmaal genuanceerder zijn.

Gödel formuleert zijn filosofische conclusie over de relatie tussen absolute bewijsbaarheid, bewijsbaarheid in een formeel systeem en waarheid van wiskundige uitspraken in termen van *Turing-machines.* Dit zijn abstracte machines, die genoemd zijn naar hun uitvinder, de Britse logicus Alan Turing. Hij gaf voor het eerst een goede abstracte analyse van het begrip algoritmische berekenbaarheid. Een algoritme is een eindige, deterministische regelstructuur, die door mensen gebruikt wordt om typen van concrete problemen op een systematische, routinematige manier op te lossen. Turing gaf een abstracte representatie van zulke regelstructuren in termen van abstracte wiskundige machines, Turing-machines genaamd. Een berekening met een algoritme wordt dan voorgesteld als het proces dat in gang gezet wordt wanneer de corresponderende Turing-machine op beginwaarden gestart wordt. Turing-machines zijn de archetypes van computerprogramma's: elk computer-programma kan worden beschouwd als een Turing-machine, en omgekeerd kan elke Turing-machine worden beschouwd als een computer-programma[5]. In modernere termen kunnen we dus zeggen dat algoritmes worden gerepresenteerd door computerprogramma's en algoritmische berekeningen worden voorgesteld door het laten lopen van een computerprogramma op beginwaarden. De stelling dat er aan elk algoritme voor het berekenen van een functie gedefinieerd op de natuurlijke getallen een Turing-machine beantwoordt die precies dezelfde functie berekent, staat bekend als de *Church-Turing thesis.* Het is een thesis die niet bewezen is — er is geen overeenstemming over de vraag of die stelling überhaupt wel *kan* worden bewezen — maar wel algemeen als waar beschouwd wordt.

Een axiomasysteem is in zekere zin ook een eindige regelstructuur. Het behoort tot de definitie van het begrip axiomasysteem dat het een routineaangelegenheid is om te beslissen of een zin een axioma is en of een bepaalde inferentie een instantie is van een inferentieregel van het axiomasysteem. Het is daarom ook een routineaangelegenheid om alle stellingen van een axiomasysteem één na één op te noemen: som exhaustief alle eindige rijen van zinnen uit de taal van het axiomasys-

5. Een computer kan worden beschouwd als een soort *universeel computerprogramma*: gegeven een beschrijving van een computerprogramma en beginwaarden, laat een computer het gegeven programma op de beginwaarden lopen en print de uitkomst. Analoog bestaat er ook een universele Turing-machine (dit was ook een ontdekking van Alan Turing). Gegeven een beschrijving van een Turing-machine en beginwaarden, laat de universele Turing-machine de gegeven machine op de beginwaarden lopen.

teem op (er bestaan algoritmes om dat te doen), en ga tegelijk voor elke opgesomde eindige rij zinnen na of elke zin ofwel een axioma is, ofwel volgt uit eerdere zinnen in de rij op basis van één van de inferentieregels van het axiomasysteem. Indien ja, dan is het een stelling, indien nee, niet. Maar dan bestaat er volgens de Church-Turing thesis een Turing-machine die diezelfde routine uitvoert en dus een exhaustieve opsomming geeft van alle stellingen van het axiomasysteem. Omgekeerd kan worden bewezen dat aan elke zinnen-opsommende Turing-machine een axiomasysteem correspondeert dat als stellingen precies de door de Turing-machine opgesomde zinnen heeft[6]. Er is dus een correspondentie tussen axiomasystemen en zinnen-opsommende Turing-machines.

Stel dat absolute bewijsbaarheid van wiskundige proposities extensioneel samenvalt met bewijsbaarheid in een bepaald axiomatisch systeem, met andere woorden dat er een axiomatisch systeem bestaat waarvan de stellingen precies de absoluut wiskundig bewijsbare uitspraken zijn. Dan zegt de tweede onvolledigheidsstelling dat er een ware rekenkundige uitspraak is die niet bewijsbaar is in het betreffende axiomatische systeem, en dus absoluut onbewijsbaar is. Maar vermits die uitspraak waar is, zal ook haar negatie natuurlijk absoluut onbewijsbaar zijn. Derhalve is die rekenkundige uitspraak absoluut onbeslisbaar. Samen-vattend: indien absolute wiskundige bewijsbaarheid samenvalt met bewijsbaarheid in een axiomasysteem, dan zijn er absoluut onbeslisbare rekenkundige uitspraken. Dit is equivalent met te zeggen dat *ofwel* absolute bewijsbaarheid van wiskundige uitspraken niet samenvalt met bewijsbaarheid in een formeel systeem, *ofwel* er absoluut onbeslisbare wiskundige uitspraken zijn. Dit is een eerste formulering van Gödels disjunctieve stelling. Maar Gödel verkiest een formulering van het eerste disjunct van deze stelling in termen van Turing-machines. Nu volgt uit de correspondentie tussen axiomatische systemen en zinnen-

6. Dat is de zogenaamde stelling van Craig.

opsommende Turing-machines dat het extensioneel samenvallen van absolute bewijsbaarheid met bewijsbaarheid in een bepaald axiomatisch systeem op hetzelfde neerkomt als het bestaan van een Turing-machine die alle absoluut bewijsbare wiskundige uitspraken genereert. En het bestaan van een Turing-machine die alle absoluut bewijsbare wiskundige uitspraken genereert, komt voor Gödel op hetzelfde neer als te zeggen dat de menselijke geest een Turing-machine is. Dit levert ons dan de uiteindelijke vorm van Gödels disjunctie: *ofwel* is de menselijke geest geen Turing-machine, *ofwel* zijn er absoluut onbeslisbare wiskundige uitspraken. Gödel beklemtoont dat deze disjunctie een wiskundig feit is[7], dat volgens hem een groot filosofisch belang heeft.

De onwaarheid van het eerste disjunct zou trouwens volgens Gödel een interessant gevolg hebben. Stel namelijk dat de menselijke geest wel een Turing-machine is, of met andere woorden dat absolute bewijsbaarheid samenvalt met bewijsbaarheid in een axiomatisch systeem. Het axiomatische systeem in kwestie kan dan, volgens de tweede onvolledigheidsstelling, zijn eigen consistentie niet bewijzen. Met andere woorden, indien absolute wiskundige bewijsbaarheid samenvalt met bewijsbaarheid in een bepaald axiomatisch systeem, dan kan de menselijke geest niet bewijzen dat zijn bewijsprincipes correct zijn[8]. Het is in dit scenario volgens Gödel niet uitgesloten dat de menselijke geest zijn bewijsprincipes effectief kan opsommen, er empirisch zo goed als zeker van kan zijn dat die opsomming alle bewijsprincipes bevat die hij ooit kan bezitten en met *inductieve* zekerheid kan stellen dat al zijn bewijsprincipes correct of tenminste consistent zijn. Maar wiskundige zekerheid zal de menselijke geest hiervan in dit scenario niet kunnen bezitten.

7. Hij spreekt van een „mathematically established fact" (GÖDEL, 1995 [1951], p. 310).

8. Gödel merkt hierbij op dat „this inability [of man] to [completely] understand himself would then wrongly appear to him as its [(the mind's)] boundlessness or inexhaustibility." (GÖDEL, 1995 [1951], p. 310).

Hierbij moet echter direct een kanttekening worden gemaakt. Het argument dat de consistentie van een voldoende sterk (consistent) axiomatisch systeem niet bewijsbaar is in het systeem zelf, is gevoelig voor de manier waarop het begrip van bewijsbaarheid in het axiomatische systeem gecodeerd wordt. Er zijn immers manieren om de notie van bewijsbaarheid in een axiomatisch systeem te coderen zodat de tweede onvolledigheidsstelling niet van toepassing is op dat systeem, en het systeem zijn eigen consistentie wel kan bewijzen[9]. Wel is het zo dat indien bewijsbaarheid op de 'standaard'-manier wordt geformaliseerd[10], de tweede onvolledigheidsstelling inderdaad van toepassing is. Ook zijn er algemene en voldoende voorwaarden voor coderingen van bewijsbaarheid in een axiomatisch systeem bekend opdat de tweede onvolledigheidsstelling van toepassing zou zijn[11]. Laten we daarom vanaf nu aannemen dat axiomatische bewijsbaarheidsnoties altijd op de standaardmanier gecodeerd worden.

4. Wiskundige bewijzen en filosofische bewijzen

Hoe kunnen we weten dat een axiomatisering van een wiskundige theorie consistent is? Laten we bijvoorbeeld de standaard-axiomatisering van de theorie van de natuurlijke getallen nemen[12]. We weten dat dat axiomatische systeem consistent is omdat we er een *bewijs* van hebben. Het bewijs gaat als volgt. Alle axioma's van het axiomatische systeem zijn waar in zijn geïntendeerd model, de structuur van de natuurlijke getallen. En de enige inferentieregel van het axiomatische systeem (modus ponens) is waarheidsbehoudend. Door een toepassing van het

9. Voor een uitgebreide bespreking van hoe de tweede onvolledigheidsstelling zo omzeild kan worden, zie Feferman, 1990.

10. Wat dit precies inhoudt, wordt uitgelegd in Boolos en Jeffrey, 1989, p. 183.

11. Deze voorwaarden zijn bekend als *Löbs afleidbaarheidsvoorwaarden*. Voor een afleiding van de tweede onvolledigheidsstelling op basis van deze afleidbaarheidsvoorwaarden, zie Boolos en Jeffrey, 1989, hoofdstuk 16.

12. Dit is de zogenaamde *Peano-rekenkunde*. Voor een beschrijving van dit axiomatisch systeem, zie bijvoorbeeld Boolos en Jeffrey, 1989, p. 182.

principe van mathematische inductie zien we dan in dat alle stellingen van het axiomatische systeem *correct* zijn voor het geïntendeerde model. Indien nu een uitspraak van de vorm $A \wedge \neg A$ een stelling zou zijn van het axiomatische systeem, dan zouden zowel A als $\neg A$ stellingen zijn en dus waar zijn in het geïntendeerde model. Maar dit is uitgesloten door de definitie van het begrip 'waarheid in een model'. Dus vermits het standaard-axiomatische systeem voor de structuur van de natuurlijke getallen correct is, is het *a fortiori* consistent.

Dit is een sluitend en overtuigend bewijs van de consistentie van de standaard-axiomatisering van de theorie van de natuurlijke getallen. Maar het bewijs doet essentieel een beroep op het waarheidsbegrip en dat is een filosofisch, en geen wiskundig begrip. Daarom is dit consistentiebewijs een (deels) filosofisch bewijs en geen zuiver wiskundig bewijs. Natuurlijk is dit bewijs ook *deels* wiskundig, want het maakt essentieel gebruik van het principe van mathematische inductie. Nu bestaat er ook een zuiver wiskundig bewijs van de consistentie van de standaard-axiomatisering van de rekenkunde. Dit werd geleverd door Gerhard Gentzen in 1936. De tweede onvolledigheidsstelling impliceert dat dit bewijs niet in de standaard-axiomatisering van de rekenkunde doorgevoerd kan worden. Dit consistentiebewijs zal dus van wiskundige bewijsprincipes gebruik maken die de bewijsprincipes van de standaard-axiomatisering van de getallentheorie te boven gaan[13]. Een gelijkaardige situatie hebben we voor de theorie van de reële getallen, de analyse. Er bestaat een filosofisch bewijs van de consistentie van de standaard-axiomatisering van de analyse. Dit bewijs verloopt helemaal analoog met het filosofische bewijs voor de consistentie van de standaard-axiomatisering van de rekenkunde. Er bestaat daarnaast ook een zuiver wiskundig bewijs voor de consistentie van de standaard-axioma-

13. Maar zo een consistentiebewijs hoeft niet noodzakelijk *alle* bewijsprincipes te gebruiken van de axiomatische theorie waarvan ze de consistentie bewijst. In Gentzens consistentiebewijs wordt bijvoorbeeld geen gebruik gemaakt van het principe van uitgesloten derde, dat wel in de standaard-axiomatisering van de rekenkunde vervat ligt.

tisering van de analyse. Voor de verzamelingentheorie ligt de zaak anders. De standaard-axiomatisering van de verzamelingentheorie is de *Zermelo-Fraenkel verzamelingentheorie*[14] (inclusief het keuze-axioma), afgekort als ZF. We hebben een filosofisch consistentiebewijs vo̍or de consistentie van de standaard-axiomatisering van de verzamelingentheorie. Maar we hebben voor ZF op dit ogenblik geen wiskundig consistentiebewijs. Veel wiskundigen vinden het filosofisch consistentiebewijs voor de standaard-axiomatisering van de verzamelingenleer weinig overtuigend. Onze wiskundige intuïtie van het geïntendeerde model (of de geïntendeerde modellen[15]) van ZF zou namelijk wel eens onvoldoende kunnen zijn om het evident te maken dat dit model (of deze modellen) de axioma's van ZF waar maakt. Voor meer eenvoudige structuren zoals de structuur van de natuurlijke getallen lijkt een dergelijke twijfel minder gegrond.

Filosofische bewijzen van wiskundige uitspraken hoeven niet minder overtuigend te zijn of minder zekerheid te bieden dan wiskundige bewijzen van wiskundige uitspraken. Het lijkt me bijvoorbeeld absoluut niet onredelijk om meer vertrouwen te hebben in het filosofische consistentiebewijs voor de rekenkunde dan in een wiskundig bewijs van de consistentie van de analyse. Ook hoeft de overtuigingskracht van een filosofisch bewijs van een wiskundige uitspraak niet afhankelijk te zijn van corresponderende wiskundige bewijzen van dezelfde uitspraak. Zo is de overtuigingskracht van ons filosofisch bewijs van de consistentie van het standaard-axiomasysteem van de rekenkunde bijvoorbeeld niet parasitair op de graad van zekerheid van het sterke inductieprincipe dat een essentiële rol speelt in Gentzens consistentiebewijs.

Indien we dit onderscheid maken tussen filosofische en wiskundige bewijzen van wiskundige uitspraken, is het argument vanuit de tweede

14. Voor een beschrijving van dit systeem, zie bijvoorbeeld ENDERTON, 1977.

15. Een aantal logici en wiskundigen denken dat er één geïntendeerd model van ZF is. Anderen denken dat er een hele klasse van geïntendeerde modellen van ZF is, waarvan het ene niet 'beter' is dan het andere.

onvolledigheidsstelling voor Gödels disjunctie dan nog wel geldig? Neem onze meest omvattende wiskundige theorie. Dat is de verzamelingentheorie, geaxiomatiseerd als ZF. Veronderstel dat ZF alle wiskundige bewijsprincipes bevat waarvan we de correctheid ooit zullen inzien. Volgt dan dat de consistentie van ZF een ware, maar absoluut onbeslisbare wiskundige uitspraak is? Neen. Inderdaad zal er in dat geval volgens de tweede volledigheidsstelling nooit een wiskundig consistentiebewijs voor ZF gegeven worden. Maar, zoals we gezien hebben, blijft de mogelijkheid open dat er een filosofisch bewijs is van de consistentie van ZF — bijvoorbeeld een argument naar het stramien van het filosofisch consistentiebewijs voor de standaard-axiomatisering van de getallentheorie. Indien men dit filosofisch consistentiebewijs overtuigend acht, dan is de consistentie van ZF wel beslisbaar.

Gödel maakt in zijn Gibbs-lezing geen expliciet onderscheid tussen wiskundige argumenten voor wiskundige uitspraken en filosofische argumenten voor wiskundige uitspraken. Verwijzend naar een elliptische vorm van het filosofisch consistentiebewijs voor de rekenkunde dat ik hierboven heb beschreven, spreekt Gödel van een *wiskundig* inzicht in de consistentie van de theorie in kwestie die door dit argument verkregen wordt:

> „It is *this* theorem [de tweede onvolledigheidsstelling] which makes the incompletability of mathematics particularly evident. For, *it makes it impossible that someone should set up a certain well-defined system of axioms and rules and consistently make the following assertion about it: All of these axioms and rules I perceive (with mathematical certitude) to be correct, and moreover I believe that they contain all of mathematics.* If someone makes such a statement he contradicts himself. For if he believes the axioms under consideration to be correct, he also believes (with the same certainty) that they are consistent. Hence he has a mathematical insight not derivable from his axioms." (GÖDEL, 1995 [1951], p. 309) [ik onderstreep].

Volgens mij is dit inzicht in de consistentie van de axiomatische theorie in kwestie een inzicht in de waarheid van een wiskundige uitspraak, maar toch een deels filosofisch en geen zuiver wiskundig inzicht. Een wiskundig inzicht in de waarheid van een wiskundige uitspraak vergt immers een zuiver wiskundig bewijs.

Toch laat Gödel in zijn Gibbs-lezing expliciet de mogelijkheid van conclusieve a priori-argumenten van niet-wiskundige aard open[16]. Uit een korte tekst van Gödel uit 1946 blijkt trouwens dat hij zich duidelijk bewust is van de mogelijkheid van filosofische bewijzen van wiskundige proposities. Gödel oppert daar zelfs het intrigerende vermoeden dat met elk filosofisch bewijs van een wiskundige propositie een zuiver wiskundig bewijs van dezelfde propositie correspondeert:

> „...the following could be true: Any proof for a set-theoretic theorem in the next higher system above set theory (i.e. any proof involving the concept of truth [...]) is replaceable by a proof from [...] an axiom of infinity." (GÖDEL, 1990 [1946], p. 151).

Hier zijn „proofs involving the concept of truth" natuurlijk filosofische bewijzen en zijn „axioms of infinity" verzamelingentheoretische, en dus wiskundige principes. Op Gödels conjectuur over de relatie tussen filosofische en wiskundige bewijzen wil ik in de afsluitende paragraaf nog even terugkomen. Voorlopig wil ik uit deze passage enkel onthouden dat Gödel zich van de mogelijkheid van filosofische bewijzen voor wiskundige proposities zeker bewust is.

Maar als dat zo is, moeten we dan niet, bij onze veronderstelling dat absolute bewijsbaarheid samenvalt met bewijsbaarheid in een axiomatisch systeem, ook de niet-wiskundige principes die eventueel een rol kunnen spelen in bewijzen van wiskundige uitspraken in dit axiomatisch systeem opnemen? De tekst van de Gibbs-lezing is volgens mij ambigu op dit punt. Neem bijvoorbeeld de volgende passage:

> „However, as to subjective mathematics, it is not precluded that there should exist a finite rule producing all its evident axioms. However, if such a rule exists, we with our human understanding could never know it to be such, that is, we could never know with mathematical certainty that all propositions it produces are correct; [...] The assertion, however, that they are all true could at most be known with empirical certainty..." (Gödel, 1995 [1951], p. 309)

16. Op p. 312, voetnoot 17, schrijft hij: „Mathematical propositions, if properly analyzed, turn out to assert nothing about the actualities of the space-time world. This is particularly clear in applied propositions such as: Either it has or it has not rained yesterday. *The existence of purely conceptual knowledge (besides mathematics) satisfying these requirements is not excluded by these remarks.*" [ik cursiveer]

Met „subjective mathematics" bedoelt Gödel hier de klasse van absoluut bewijsbare wiskundige uitspraken. Indien met „its evident axioms" enkel wiskundige axioma's bedoeld worden, dan volgt de conclusie niet. Zoals we gezien hebben, is dan een filosofisch correctheidsbewijs niet uitgesloten. Maar het kan ook zijn dat Gödel ook de niet-zuiver wiskundige bewijsprincipes tot die evidente axioma's rekent.

Er is, voor zover ik kan zien, geen tekstueel bewijsmateriaal in de Gibbs-lezing dat deze ambiguïteit helemaal oplost. Maar gegeven het feit dat Gödel zich van het bestaan van filosofische bewijzen voor wiskundige uitspraken bewust is, lijkt me deze laatste interpretatie eigenlijk wel een zekere plausibiliteit te bezitten. Laten we dus met enige welwillendheid stellen dat wanneer Gödel vertrekt van de aanname dat absolute bewijsbaarheid samenvalt met bewijsbaarheid in een axiomatisch systeem, in het betreffende axiomatische systeem ook alle niet-wiskundige bewijsprincipes moeten worden opgenomen die kunnen worden gebruikt bij het afleiden van wiskundige proposities — bijvoorbeeld axioma's betreffende de logische eigenschappen van het waarheidsbegrip. Dan zullen er, volgens de tweede onvolledigheidsstelling, inderdaad rekenkundige uitspraken zijn (van lage recursie-theoretische complexiteit) die niet bewezen kunnen worden in dat systeem. Met andere woorden, in dat geval zijn er absoluut onbeslisbare wiskundige uitspraken.

Toch wordt de situatie nu iets minder duidelijk. Sinds meer dan vijftig jaar is er binnen de wiskundige gemeenschap een ruime consensus over wat de wiskundige basisprincipes zijn van onze meest omvattende zuiver wiskundige theorie: deze worden uitgedrukt door de axioma's van ZF. Maar wat de basisprincipes zijn van een goede waarheidstheorie, daarover is er absoluut geen consensus. Het axioma-schema waarrond iedereen zich zou willen verenigen zegt:

„ ..." is waar dan en slechts dan wanneer ...,

waarbij... door een willekeurige zin kan worden ingevuld. Maar de leugenaarsparadox toont informeel aan dat indien de ingevulde zin zelf het waarheidspredikaat mag bevatten, dit schema problemen oplevert. En

Tarski heeft formeel aangetoond dat het inconsistent is om dit axiomaschema zelfs maar aan vrij zwakke wiskundige theorieën toe te voegen, laat staan aan ZF. Er bestaat absoluut geen overeenstemming over welke (zwakkere) axioma's we dan in de plaats van het paradoxale schema zouden moeten nemen. Bovendien zijn er nog andere begrippen, zoals het begrip 'absolute bewijsbaarheid', of 'absolute willekeurigheid' („absolute randomness"), die misschien gebruikt zouden kunnen worden om de consistentie van wiskundige theorieën te bewijzen. We hebben op dit moment zelfs geen goed criterium waarmee de begrippen die een dergelijke rol kunnen spelen geïsoleerd kunnen worden. Het is daarenboven nog nauwelijks onderzocht hoe dergelijke begrippen geaxiomatiseerd zouden moeten worden en wat de logische relaties tussen deze begrippen zijn. Uit dit alles volgt dat het geheel van wiskundige axioma's plus wiskundig relevante maar niet zuiver wiskundige axioma's waarvan de wiskundige gemeenschap inziet dat ze correct zijn, in ieder geval vandaag de dag een erg moeilijk identificeerbaar geheel is. Of dat aan het einde der tijden — want daar gaat het bij absolute bewijsbaarheid uiteindelijk om — nog zo zal zijn, valt natuurlijk af te wachten. Maar het blijft een feit dat de tweede onvolledigheidsstelling van toepassing is op elke scherpe manier om dit geheel als een axiomatisch systeem te identificeren. Uiteindelijk, en met de nodige kwalificaties, lijkt me dus Gödels argument voor zijn disjunctieve stelling onontkoombaar.

5. De disjuncten

Gödel zelf dacht dat het eerste disjunct van zijn filosofische stelling waar is, en dat het tweede disjunct onwaar is. Maar tegelijk was hij er zich van bewust dat hij voor deze verdere conclusies geen dwingende argumenten bezat.

Wat het eerste disjunct betreft, lijkt het Gödel dat de onwaarheid ervan impliceert dat de menselijke geest reduceerbaar is tot de werking van de hersenen[17], waarbij hij ervan uitgaat dat de hersenen een eindig mechanisme zijn, dat als een Turing-machine kan worden beschouwd. Gödel acht het onwaarschijnlijk dat een dergelijke reductie van het mentale tot het fysische werkelijk doorgevoerd kan worden. Later zal Gödel beklemtonen dat men ook rekening moet houden met de *evolutie* van de menselijke geest[18]. Door te reflecteren op de inhoud van wiskundige basisbegrippen (zoals het begrip verzameling, het begrip deelverzameling, ...) kan men komen tot nieuwe wiskundige axioma's. Misschien kunnen we dit reflectieproces uitvoeren op een manier die systematisch is, maar tegelijk zodanig dat dit proces in de oneindige limiet tot een verzameling axioma's leidt die niet door een Turing-machine kan worden opgesomd. Gödel realiseerde zich dat de geschiedenis van de wiskunde voor deze hypothese tot dan toe geen duidelijke steun gaf.

Een resultaat van Solomon Feferman uit de jaren zestig over reeksen van formele systemen is in deze context relevant[19]. Beschouw weer de standaard-axiomatisering van de rekenkunde. In een eerste stap voegen we de (getallentheoretische codering van) de consistentie-uitspraak voor die theorie als nieuw axioma toe aan de standaard-axiomatisering. Dit levert een nieuw axiomatisch systeem op, dat volgens de tweede onvolledigheidsstelling sterker is dan de standaard-axiomatisering. We noemen dat systeem T_1. In een tweede stadium voegen we aan T_1 de consistentie-uitspraak voor het systeem T_1 toe. Dit levert ons een systeem T_2 op, dat sterker is dan T_1. Zo gaan we door tot in het transfiniete[20], waarbij we de stadia steeds aanduiden door ordinaalgetallen. Dit proces moet ooit tot een einde komen, want er zijn meer ordinaalgetallen, en dus stadia, dan zinnen uit de taal van de rekenkunde. Feferman toonde aan dat dit proces stopt in het limietstadium met als index het

17. Ik moet toegeven dat ikzelf deze implicatie niet inzie.

18. Zie Gödel, (1990) [1972], p. 306.

19. Zie Feferman, 1962.

20. In limietstadia nemen we als nieuw systeem het systeem dat bestaat uit alle stellingen die bewezen kunnen worden op basis van axioma's van reeds bekomen systemen.

aftelbaar oneindige ordinaalgetal ε_0. Het systeem T_{ε_0} dat in dat stadium gecreëerd is, kan het begrip „bewijsbaar in T_{ε_0}" niet meer coderen en kan bijgevolg zijn eigen consistentie niet meer uitdrukken. Maar dit betekent dat T_{ε_0} geen axiomatisch systeem meer kan zijn, want we hebben gezien dat axiomasystemen hun bewijsbegrip wel kunnen uitdrukken. Dus T_{ε_0} is niet opsombaar door een Turing-machine. We hebben in sectie 4 een filosofisch bewijs gegeven van de correctheid en consistentie van de standaard-axiomatisering van de rekenkunde. Een gelijkaardig argument toont, gegeven de correctheid en consistentie van de standaard-axiomatisering van de rekenkunde, de correctheid en consistentie van T_1 aan. En gegeven *dat* feit, kunnen we de correctheid en consistentie van T_2 aantonen. Enzovoort. Met andere woorden, het proces van Feferman leidt tot steeds sterkere, aantoonbaar correcte en consistente systemen, en culmineert in een systeem T_{ε_0} dat zeker niet axiomatiseerbaar is. Indien de klasse van absoluut bewijsbare wiskundige uitspraken zou samenvallen met T_{ε_0}, dan zou Gödels eerste disjunct waar zijn.

Zouden hiërarchieën van axiomatische theorieën zoals die van Feferman dan niet gebruikt kunnen worden als argumenten voor de waarheid van Gödels eerste disjunct? Ik vrees dat deze vraag negatief moet worden beantwoord. Als we de redelijke aanname maken dat tussen het uitvoeren van het correctheidsbewijs van T_α en het uitvoeren van het correctheidsbewijs van $T_{\alpha+1}$ (voor $\alpha < \varepsilon_0$) telkens een tijdsinterval ligt dat groter is dan nul, dan is er een transfiniete hoeveelheid tijd nodig om Fefermans proces helemaal te doorlopen. Op dit moment hebben we echter geen goede wetenschappelijke of niet-wetenschappelijke argumenten om aan te nemen dat er nog zoveel tijd *is* in het universum. Als we de meest recente astronomische theorieën mogen geloven, dan zal het heelal voor altijd blijven uitdijen. Dit zou betekenen dat de tijdsdimensie in de toekomstrichting (wanneer die vanaf een willekeurig tijdsmoment wordt beschouwd) de structuur heeft van de verzameling van de positieve reële getallen. Hieruit zou dan volgen dat we slechts ω stadia van Fefermans proces kunnen doorlopen, wat een veel kleiner ordinaalgetal is dan ε_0[21]. Er ligt dan in de toekomstrichting van de tijdsdimensie geen lineaire ordening van type ε_0 zonder ophopingspunten vervat! Met andere woorden, om een argument voor Gödels eerste disjunct te verkrijgen zouden we een reflectieproces moeten vinden dat in hoogstens ω stappen vanuit een axiomatiseerbare theorie een niet-axiomatiseerbaar systeem oplevert. Voor zover ik weet is er op dit moment geen dergelijk proces bekend.

Varianten van het eerste disjunct zijn gedurende de laatste dertig jaar veel besproken. Dat cognitieve functies gemodelleerd kunnen worden door Turing-machines is een vooronderstelling van veel werk in cognitieve psychologie en Artificial Intelligence. Hierbij wordt veelal abstractie gemaakt van de manier waarop de menselijke geest evolueert (laat staan in de oneindige limiet), wat het een zwakkere thesis maakt dan Gödels eerste disjunct. Er bestaan tegenwoordig wel pogingen om aspecten van de kortetermijnevolutie van de mens (namelijk leergedrag) te modelleren met behulp van neurale netwerk-programma's uit de computerwetenschappen. Maar dit maakt voor ons weinig verschil, omdat elke functie die door een neuraal netwerk berekend kan worden ook door een ouderwetse Turing-machine kan worden berekend (en omgekeerd). Over het statuut van de gebruikte computermodellen is men het trouwens grondig oneens. Mogen ze slechts opgevat worden als *instrumenten* die gebruikt kunnen worden om een beknopte benaderende beschrijving te geven van geobserveerd cognitief gedrag, of is de menselijke geest *echt* gestructureerd zoals een computerprogramma?

Tegelijk is het een wijdverbreide stelling in de *philosophy of mind* dat het mentale reduceerbaar is tot het fysiologische. Dat is een uitdrukking van het fysicalisme en dit brengt ons terug naar Gödels bewering dat de hersenen een eindig mechanisme, een Turing-machine zijn. Dit laatste lijkt me absoluut geen uitgemaakte zaak. Hersenprocessen zijn,

21. ω is het kleinste aftelbaar oneindige ordinaalgetal dat de ordestructuur van de natuurlijke getallen weergeeft. Het is gemakkelijk in te zien dat T_ω een axiomatiseerbare theorie is.

voor zover we nu weten, continue en geen discrete processen. Dus als onze hersenen al computers zijn, dan zijn het *analoge* in plaats van digitale computers. De theoretische greep die we tot op heden hebben op analoge rekenprocessen is veel zwakker dan ons theoretisch inzicht in digitale rekenprocessen. We kennen bijvoorbeeld geen informatieve beschrijving van de klasse van functies die door analoge computers berekend kunnen worden, terwijl we zo een beschrijving wel hebben voor Turing-berekenbare functies. Hierbij komt dan nog de complicatie dat, naarmate onze hersenen evolueren door interactie met hun fysische omgeving, ze misschien een ruimere klasse van functies kunnen berekenen.

Er zijn pogingen ondernomen om Gödels tweede onvolledigheidsstelling te gebruiken om aan te tonen dat de menselijke geest geen Turing-machine kan zijn, zelfs wanneer we abstractie maken van de manier waarop de menselijke geest evolueert[22]. Men redeneert dan ongeveer als volgt. Stel dat mijn bewijsprincipes een axiomatisch systeem vormen. Dan zou, volgens de tweede onvolledigheidsstelling, de correctheid van die bewijsprincipes onbewijsbaar zijn in dat axiomatische systeem. Met andere woorden, ik zou de correctheid van mijn bewijsprincipes niet kunnen inzien. Maar ik kan de correctheid van mijn bewijsprincipes *wel* inzien. Dus vormen mijn bewijsprincipes geen axiomatisch systeem. De meeste logici denken tegenwoordig dat dit argument, dat men het *Lucas-argument* noemt, niet overtuigend is. Een gedetailleerde bespreking van dit argument zou ons te ver leiden, ik ga er hier dan ook niet verder op in[23].

Wat het tweede disjunct betreft, dacht Gödel, in navolging van Hilbert, dat er geen absoluut wiskundig onbeslisbare uitspraken bestaan. De rede zou immers onredelijk zijn indien ze vragen naar voren zou schuiven die met de redelijkheid zelf niet te beantwoorden

22. Zie bijvoorbeeld LUCAS, 1961, en PENROSE, 1989.
23. Voor een kritische bespreking van verschillende versies van Lucas' argument, zie VISSER, 1986.

zijn. Boolos oppert het vermoeden dat Gödel dit argument aan Kant heeft ontleend[24]. Kant schrijft namelijk in de Kritik der reinen Vernunft:

> „There are sciences the very nature of which requires that every question arising within their domain should be completely answerable in terms of what is known, inasmuch as the answer must issue from the same sources from which the question proceeds" (KANT, 1933 [1781, 1787], A 476/B 504).

Hierbij haalt Kant de wiskunde als voorbeeld van één van dergelijke wetenschappen aan.

Veel wiskundigen zijn tegenwoordig geneigd om als bijna vanzelfsprekend aan te nemen dat er absoluut onbewijsbare wiskundige uitspraken bestaan. Verschillende argumenten worden hiervoor aangegeven. Ten eerste zijn er uitspraken over de natuurlijke getallen die, hoewel recursie-theoretisch niet erg complex, toch zo oninzichtelijk zijn[25] dat het hoogst twijfelachtig is of we ooit de bewijsmiddelen zullen bezitten die nodig zijn om ze te beslissen. Indien we nauwelijks inzicht hebben in de inhoud van een rekenkundige uitspraak, hoe kunnen we dan gericht zoeken naar correcte axioma's die die uitspraak beslissen? Natuurlijk speelt in dit argument de tweede onvolledigheidsstelling, die zegt dat elk axiomatisch systeem oneindig veel van zulke uitspraken onbeslist moet laten, een cruciale rol. Ten tweede zijn er uitspraken die absoluut onbeslisbaar lijken omdat bepaalde begrippen die erin voorkomen ofwel *in se* onvoldoende bepaald zijn om de waarheidswaarde van de uitspraak in kwestie te determineren, of omdat ons inzicht in deze begrippen altijd onvoldoende zal blijven om de waarheidswaarde ervan ooit te achterhalen. Het canonieke voorbeeld van zo een wiskundige uitspraak is Cantors continuumhypothese uit de verzamelingentheorie. Het ondergedetermineerde of ondetermineerbare begrip dat hierin zou voorkomen is het begrip 'deelverzameling':

24. Zie BOOLOS, 1995, p. 294.

25. Bijvoorbeeld omdat er enorm grote getallen in voorkomen, of, wat eigenlijk op hetzelfde neerkomt, omdat er veel variabelen in voorkomen of omdat ze erg lang zijn. Een voorbeeld hiervan wordt gegeven door FEFERMAN en SOLOVAY, 1990, p. 292.

Bovendien zou het wel eens kunnen dat er absoluut onbeslisbare uitspraken zijn die het begrip absolute bewijsbaarheid essentieel bevatten. De uitspraak dat er absoluut onbeslisbare wiskundige uitspraken bestaan zou wel eens *zelf* absoluut onbeslisbaar kunnen zijn. (Hierbij merk ik op dat de argumenten die ik zo-even gaf voor de stelling dat er absoluut onbeslisbare wiskundige uitspraken bestaan, niet als filosofische of wiskundige *bewijzen* in de zin van dwingende a priori argumenten voor het bestaan van absoluut onbeslisbare wiskundige uitspraken beschouwd kunnen worden.)

Wat er ook van zij, de kantiaanse argumenten tegen het bestaan van absoluut onbeslisbare wiskundige uitspraken lijken velen vandaag de dag nog weinig overtuigend. Precies door het werk in de wiskundige logica waarvan Gödel aan de basis lag, lijkt de wiskundige ervaring nu in de andere richting te wijzen.

6. Afsluitende beschouwingen

Gödel dacht dat de these dat ofwel de menselijke geest een Turingmachine is, ofwel er absoluut onbeslisbare wiskundige uitspraken zijn, een filosofisch belangrijke stelling is, die met mathematische zekerheid volgt uit de tweede onvolledigheidsstelling. Hoewel Gödels argument volgens mij niet duidelijk genoeg rekening houdt met het bestaan van filosofische bewijzen van wiskundige uitspraken, is er een versie van Gödels argument die volgens mij duidelijk geldig is. Met enige welwillendheid kunnen we deze versie zelfs aan Gödel zelf toeschrijven.

De disjunctieve stelling op zich lijkt vanuit een hedendaags standpunt niet zo opzienbarend. Hoewel de vraag of absolute bewijsbaarheid samenvalt met bewijsbaarheid in een axiomatisch systeem volledig open is, zijn er — *ondanks* wat Gödel zelf daarover zegt, en *dankzij* Gödels werk in wiskundige logica — goede redenen om aan te nemen dat er absoluut onbeslisbare wiskundige uitspraken zijn. Maar daar staat tegenover dat er een (filosofisch) *bewijs* lijkt te zijn van de disjunctieve stelling, terwijl we geen bewijs hebben van het tweede disjunct. Hoe sterker de conclusie, hoe minder dwingend de argumenten die ervoor te geven zijn.

Laat ik tot slot nog even terugkomen op Gödels interessante vermoeden dat voor ieder deels filosofisch bewijs van een wiskundige uitspraak een zuiver wiskundig bewijs bestaat dat dezelfde uitspraak bewijst. Ik zie in de geschiedenis van de logica van deze eeuw eigenlijk weinig steun voor dit vermoeden, in het bijzonder voor wat de verzamelingentheorie betreft. Gödel hoopte dat er systematisch steeds sterkere verzamelingentheoretische axioma's ontdekt zouden worden, die als bewijsprincipes door de wiskundige gemeenschap zouden worden aanvaard en die in conjunctie met de axioma's van ZF zouden kunnen worden gebruikt om bijvoorbeeld de consistentie van zwakkere axiomatische systemen te bewijzen. Ondertussen zijn er een grote hoeveelheid van dergelijke principes voorgesteld, maar geen van hen is zelfs maar vrij algemeen als bewijsprincipe door de wiskundige gemeenschap aanvaard. Het zou volgens mij wel eens kunnen dat er wiskundige uitspraken zijn die enkel door een deels filosofisch bewijs bewezen kunnen worden. Misschien is de consistentie van ZF zo een uitspraak.

SUMMARY: *Gödels Disjunction*

In his Gibbs lecture, Gödel argued for the thesis that either the human mind is not a Turing machine, or there exist absolutely undecidable mathematical propositions. He believed that this disjunction can be deduced with mathematical certainty from certain results in mathematical logic. He thought that his disjunctive thesis is of great philosophical importance.

First, Gödel's argument for his disjunctive thesis is discussed. It is argued that this argument contains an ambiguity. But when it is made precise in a certain way, the argument is seen to be ultimately inescapable. Second, the disjuncts of the thesis are considered in their own right. It is argued that there are today no convincing arguments in favor of the first disjunct, nor are there at present convincing arguments against it. But there do exist today strong reasons to believe the second disjunct to be true, although these reasons do not establish the disjunct with anything like mathematical certainty.

Literatuur:

BOOLOS, G. en JEFFREY, R. (1989) *Computability and Logic.* Cambridge, Cambridge University Press, 1989³.

BOOLOS, G. (1995) Introductory Note to *1951. In: K. GÖDEL, *Collected Works,* Volume 3, edited by S. FEFERMAN et al., New York/Oxford, Oxford U.P., 1995, p. 290-304.

ENDERTON, H. (1977) *Elements of Set Theory.* San Diego, Academic Press, 1977.

FEFERMAN, S. (1962) Transfinite Recursive Progressions of Axiomatic Theories. *Journal of Symbolic Logic* 27 (1962), p. 259-316.

FEFERMAN, S. (1990) Remark 1 of the Introductory Note to 1972a. In: K. GÖDEL, *Collected Works,* Volume 2, edited by S. FEFERMAN et al., New York/Oxford, Oxford U.P., 1990, p. 282-287.

FEFERMAN, S. en SOLOVAY, R. (1990) Remark 2 of the Introductory Note to 1972a. In: K. GÖDEL, *Collected Works,* Volume 2, p. 287-292.

GENTZEN, G. (1936) Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen* 112 (1936), p. 493-565.

GÖDEL, K. (1990) [1946] Remarks before the Princeton Bicentennial Conference on Problems in Mathematics. In: K. GÖDEL, *Collected Works,* Volume 2, p. 150-153.

GÖDEL, K. (1990) [1972] Some Remarks on the Undecidability Results. In: K. GÖDEL, *Collected Works,* Volume 2, p. 305-306.

GÖDEL, K. (1995) [1951] Some Basic Theorems on the Foundations of Mathematics and Their Implications. In: K. GÖDEL, *Collected Works,* Volume 3, p. 304-323.

KANT, I. (1933) [1781, 1787] *Critique of Pure Reason.* English translation of Kant 1781 and 1787 by Norman KEMP SMITH. 2nd impression (with corrections), London, Macmillan, 1933.

KRIPKE, S. (1995) *Elementary Recursion Theory and its Application to Formal Systems.* Ongepubliceerd manuscript, 1995.

LUCAS, J.R. (1961) Minds, Machines and Gödel, *Philosophy* 36 (1961), p. 112-127.

PENROSE, R. (1989) *The Emperor's New Mind.* New York/Oxford, Oxford University Press, 1989.

VISSER, A. (1986) Kunnen wij elke machine verslaan? Beschouwen rondom Lucas' argument. *Logic Group Preprint Series* no. 11, Universiteit Utrecht, Departement Wijsbegeerte.