De onvolledigheidsstelling van Gödel

Wouter Zomervrucht, s0713317

26 maart 2009

Artikel voor het vak LPC
Onderwerp: de eerste onvolledigheidsstelling van Gödel

Inleiding

In het begin van de twintigste eeuw is door veel wiskundigen geprobeerd de wiskunde te formaliseren. Dit wilden ze bereiken door middel van zogenaamde formele systemen. Het beroemdste voorbeeld hiervan zijn de Principia Mathematica van Alfred North Whitehead en Bertrand Russell, waarin grote gebieden van de wiskunde direct vanuit axioma's worden afgeleid. De kracht van deze systemen was echter onbekend. Onder andere wist men niet of het mogelijk was om bepaalde wiskundige gebieden volledig met een formeel systeem te beschrijven. Pas in 1931 gaf Gödel voor de getaltheorie het schokkende antwoord in zijn artikel 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I' (zie [1]): dit is onmogelijk.

In de komende paragrafen zullen we dit resultaat afleiden. Paragraaf 1 is een korte inleiding in formele systemen. In paragraaf 2 zullen we zien hoe Gödel met behulp van zogenaamde Gödelgetallen uitspraken van het formele systeem om wist te zetten in getallen. Het doel hiervan is om binnen het systeem over deze uitspraken te kunnen praten (het systeem kan immers over getallen redeneren). Vervolgens, in paragraaf 3, beschouwen we het begrip uitdrukbaarheid, dat aangeeft wat voor uitspraken in het systeem kunnen worden gedaan. In paragraaf 4 construeren we een uitspraak die in feite zegt: 'Deze uitspraak is onbewijsbaar.' Tenslotte laten we zien dat noch deze uitspraak noch zijn ontkenning kan worden bewezen, waarmee de stelling van Gödel is bewezen.

1 Formele systemen

Het middel waarmee de wiskunde wordt geformaliseerd zijn formele systemen. Een formeel systeem is in feite een methode om wiskundige uitspraken te schrijven in (betekenisloze) symbolen, en om redeneren te vervangen door het manipuleren van deze symbolen.

Definitie. Een *formeel systeem* is een viertupel $(\mathcal{S}, \mathcal{F}, \mathcal{A}, \mathcal{D})$, met:
- $\mathcal{S}$ een aftelbare collectie *symbolen*; een eindig rijtje symbolen noemen we een *formule*
- $\mathcal{F}$ een deelcollectie *welgevormde formules* binnen de collectie formules
- $\mathcal{A}$ een deelcollectie *axioma's* binnen de collectie welgevormde formules
- $\mathcal{D}$ een collectie *afleidingsregels* $C \to \phi$ waarbij C een eindige collectie welgevormde formules is en ϕ een welgevormde formule.

In de appendix staat een uitgewerkt voorbeeld van een formeel systeem, dat we P zullen noemen. We zullen nog meerdere malen naar dit systeem verwijzen.

De collectie welgevormde formules is niets anders dan het precies maken van het intuïtieve begrip 'zinnige uitspraak'. Zo is in P de formule $0 = 0$ welgevormd, maar $\forall\neg)s$ niet. Redeneren over uitspraken gebeurt nu precies zoals te verwachten is op grond van de naamgeving: we accepteren de axioma's zonder bewijs, en leiden hieruit met de afleidingsregels nieuwe uitspraken af. Dit wordt precies gemaakt met de begrippen 'bewijs' en 'stelling'.

Definitie. Een *bewijs* in een formeel systeem is een eindig rijtje welgevormde formules, zodanig dat voor iedere formule ϕ in dat rijtje geldt dat ϕ een axioma is, of dat er een afleidingsregel $C \to \phi$ is waarvoor C bestaat uit eerdere formules in het rijtje.

Definitie. Een *stelling* in een formeel systeem is een welgevormde formule ϕ waarvoor er een bewijs is met als laatste formule ϕ.

Nu we in het formele systeem kunnen redeneren, zijn we ook in staat om enkele eigenschappen van dit redeneren te bekijken. In een bruikbaar formeel systeem willen we bijvoorbeeld geen tegenspraak kunnen afleiden. Het is echter niet algemeen duidelijk wat tegenspraak precies is, en daarom zullen we ons beperken tot een bepaald type systemen, die we *type A-systemen* zullen noemen.[1] Dit zijn formele systemen die alle symbolen van P bevatten, waarin iedere welgevormde formule van P welgevormd

[1]Dit is geen officiële term, maar we voeren hem in ten behoeve van de overzichtelijkheid.

is, die alle afleidingsregels van P bevatten, en waarin de axioma's 1 tot en met 6 van P stellingen zijn. (In het bijzonder is dus P een type A-systeem, omdat ieder axioma een stelling is.) Binnen deze systemen kunnen we praten over tegenspraken.

> **Definitie.** Een type A-systeem is *consistent* als er geen welgevormde formule ϕ is waarvoor ϕ en $\neg\phi$ allebei stellingen zijn.

We zullen straks zien dat in een inconsistent systeem iedere welgevormde formule een stelling is. Deze systemen zijn dus niet bijzonder interessant, en het is niet meer redelijk om, wanneer we een uitspraak over een type A-systeem doen, aan te nemen dat dit systeem consistent is. In zijn stelling neemt Gödel zelfs nog iets sterkers aan, namelijk ω-consistentie.

> **Definitie.** Een type A-systeem is *ω-consistent* als er geen welgevormde formule $\phi(x_i)$ is (met x_i een variabele die ongebonden voorkomt in ϕ) waarvoor $\phi(s^n 0)$ een stelling is voor ieder natuurlijk getal[2] n, en ook $\neg(\forall x_i)\phi(x_i)$ een stelling is.

Dit komt erop neer dat als een uitspraak kan worden bewezen voor iedere n, dan niet ook kan worden bewezen dat niet voor iedere n de uitspraak geldt. Dit zien we immers eveneens als een soort tegenspraak. Het volgende lemma toont aan dat ω-consistentie inderdaad een minstens zo sterke aanname is als consistentie.

> **Lemma.** Als een type A-systeem ω-consistent is, dan is het ook consistent.

> **Bewijs.** We bewijzen dit uit het ongerijmde. Stel dat het systeem inconsistent is, dan is er een welgevormde formule ϕ waarvoor ϕ en $\neg\phi$ beide stellingen zijn. We geven een bewijs in het formele systeem dat nu iedere welgevormde formule ψ een stelling is.

1. ϕ	aanname
2. $\neg\phi$	aanname
3. $\neg\phi \Rightarrow (\neg\psi \Rightarrow \neg\phi)$	axioma 1
4. $\neg\psi \Rightarrow \neg\phi$	afleidingsregel 2 met regel 2 en 3
5. $(\neg\psi \Rightarrow \neg\phi) \Rightarrow (\phi \Rightarrow \psi)$	axioma 3
6. $\phi \Rightarrow \psi$	afleidingsregel 2 met regel 4 en 5
7. ψ	afleidingsregel 2 met regel 1 en 6

> Omdat we bewijzen van ϕ en $\neg\phi$ kennen, kunnen we een bewijs voor ψ construeren door deze bewijzen achter elkaar te plaatsen en bovenstaande bewijs (met de juiste formules voor ϕ en ψ ingevuld) daarachter te plaatsen. We concluderen dat iedere welgevormde formule een stelling is. In het bijzonder is $s^n 0 = s^n 0$ een stelling voor ieder natuurlijk getal n, en ook $\neg(\forall x_1)x_1 = x_1$ is een stelling. Maar dan is de formule $x_1 = x_1$ een tegenvoorbeeld voor ω-consistentie. We zien dat ieder inconsistent type A-systeem ook ω-inconsistent is, zodat iedere ω-consistent systeem ook consistent moet zijn. $\square$

We komen aan bij de vraag waar Gödels stelling een antwoord op geeft: hoeveel kan een (consistent) formeel systeem bewijzen? Als we, bijvoorbeeld in P, aan welgevormde formules de gebruikelijke interpretatie geven, dan is altijd ofwel ϕ ofwel $\neg\phi$ een formule die correspondeert met een ware uitspraak. De betreffende formule zouden we dan ook willen kunnen bewijzen. Formele systemen waarin dit altijd kan noemen we volledig.

> **Definitie.** Een type A-systeem is *volledig* als voor iedere welgevormde formule ϕ ofwel ϕ zelf ofwel $\neg\phi$ een stelling is.

Opvallend is de symmetrie met de definitie van consistentie: deze laatste eist dat ϕ en $\neg\phi$ niet beiden een stelling zijn, terwijl volledigheid eist dat wel één van de twee een stelling is. Graag zouden we dus een systeem hebben dat beide eigenschappen heeft. Kortgezegd beweert Gödels stelling nu dat dit niet kan als het systeem de getaltheorie bevat. Met een *type A-systeem dat de getaltheorie bevat*

[2] In dit artikel verstaan we onder natuurlijke getallen alle niet-negatieve gehele getallen.

bedoelen we een type A-systeem waarin alle axioma's van P stellingen zijn.[3] Hoewel het lijkt alsof we ons hiermee beperken tot formele systemen die sterk lijken op P, is dit niet het geval: het enige wat we eisen, is dat iedere geldige redenering in P ook in het betreffende formele systeem kan worden uitgevoerd. Voor de rest mag het systeem net zo uitgebreid en ingewikkeld zijn als gewenst.

2 Gödelgetallen

Het idee van Gödels bewijs is om binnen het formele systeem over het formele systeem zelf te praten. Daartoe maakte Gödel een codering van formules in getallen, bekend onder de naam 'Gödelgetallen'. Hiermee kan het systeem, dat immers de getaltheorie beschrijft, impliciet uitspraken doen over zichzelf terwijl het eigenlijk over getallen praat. Het toekennen van Gödelgetallen kan op veel manieren, en we zullen hier één van deze manieren bekijken.

Allereerst geven we elk symbool een eigen uniek positief Gödelgetal. Dit kan altijd omdat de collectie symbolen aftelbaar is. In het geval van P ligt de volgende nummering voor de hand:

symbool	$\neg$	$\Rightarrow$	$\forall$	(	)	0	s	$+$	$\times$	$=$	x_1	x_2	x_3	...
Gödelgetal	1	2	3	4	5	6	7	8	9	10	11	12	13	...

We schrijven voor het gemak de toekenning van Gödelgetallen als de functie g, bijvoorbeeld $g(\neg) = 1$.

Deze codering van symbolen gebruiken we vervolgens om iedere formule een unieke Gödelgetal te geven. Stel dat we de formule $\phi = \alpha_1\alpha_2 \ldots \alpha_n$ hebben (waarin de α_i dus symbolen zijn), dan krijgt deze het Gödelgetal

$$g(\phi) = \prod_{i=1}^{n} p_i^{g(\alpha_i)} = 2^{g(\alpha_1)} 3^{g(\alpha_2)} \cdots p_n^{g(\alpha_n)}$$

waarin p_i het i^{de} priemgetal in grootte is. De uniciteit van priemontbinding geeft direct dat op deze manier iedere formule een uniek Gödelgetal krijgt. Dezelfde constructie kunnen we ook gebruiken om aan eindige rijtjes formules Gödelgetallen toe te kennen (met als doel om bewijzen, die niets meer zijn dan rijtjes formules met bepaalde eigenschappen, in getallen te coderen): we geven het rijtje formules $\phi_1, \phi_2, \ldots, \phi_n$ het Gödelgetal

$$g(\phi_1, \phi_2, \ldots, \phi_n) = \prod_{i=1}^{n} p_i^{g(\phi_i)}.$$

Het moet worden opgemerkt dat (bijvoorbeeld) een symbool en een formule hetzelfde Gödelgetal kunnen hebben. Er zijn toekenningen van Gödelgetallen mogelijk waarbij dit niet het geval is, maar dit is niet nodig. Steeds als we over een Gödelgetal spreken zal vastliggen of dit het Gödelgetal van een symbool, een formule of een rijtje formules zou moeten zijn. Van belang is slechts dat geen twee symbolen hetzelfde Gödelgetal kunnen hebben, en hetzelfde voor formules en rijtjes formules.

Tot slot geven we een voorbeeld van het toekennen van Gödelgetallen. Om de formule $0 = 0$ een Gödelgetal te geven, bepalen we eerst de Gödelgetallen van de voorkomende symbolen, respectievelijk 6, 10 en 6. Het Gödelgetal van de formule is dus $2^6 3^{10} 5^6 = 59.049.000.000$. Andersom, als we willen weten welke formule Gödelgetal 59.049.000.000 heeft (en of er überhaubt zo'n formule bestaat), dan ontbinden we dit getal in priemfactoren zodat we vinden $59.049.000.000 = 2^6 3^{10} 5^6$, waaruit volgt dat de formule $0 = 0$ moet zijn. We zien dat Gödelgetallen snel heel groot worden, en het is niet prettig om hier expliciet mee te moeten rekenen. Gelukkig is het genoeg om te weten dat het mogelijk is.

3 Uitdrukbaarheid

We hebben gezien dat we formules en bewijzen in het formele systeem kunnen opvatten als getallen. Het doel is nu om binnen het formele systeem uitspraken over deze getallen te doen. Daarvoor is het

[3] De klasse systemen die de getaltheorie bevatten is eigenlijk groter, maar het voert hier te ver om dit precies te maken.

nodig om te weten wat voor uitspraken we eigenlijk kunnen uitdrukken in het systeem. Eerst definiëren we wat we precies bedoelen met uitdrukken.

Definitie. Een getaltheoretische relatie (een relatie op de natuurlijke getallen) $R(x_1, \ldots, x_k)$ is *uitdrukbaar* in een type A-systeem, als er een welgevormde formule $\phi(x_1, \ldots, x_k)$ is zodat voor alle natuurlijke getallen $n_1, \ldots, n_k$ het volgende geldt:
- als $R(n_1, \ldots, n_k)$ geldt, dan is $\phi(s^{n_1}0, \ldots, s^{n_k}0)$ een stelling
- als $R(n_1, \ldots, n_k)$ niet geldt, dan is $\neg\phi(s^{n_1}0, \ldots, s^{n_k}0)$ een stelling.

Merk op dat definitie afhankelijk is van het formele systeem waarin we werken. Als voorbeeld bekijken we de relatie $\text{deler}(x, y)$ die geldt als x een deler is van y. Deze relatie is uitdrukbaar in P door de welgevormde formule $(\exists x_3)x_1 \times x_3 = x_2$. Immers, als m een deler is van n, dan is $(\exists x_3)s^m0 \times x_3 = s^n0$ een stelling in P, en als m geen deler is van n, dan is $\neg(\exists x_3)s^m0 \times x_3 = s^n0$ een stelling in P (we zullen hier geen bewijs van deze claims geven).

Gödel heeft een grote klasse van relaties aangegeven, die in ieder type A-systeem dat de getaltheorie bevat uitdrukbaar zijn. Dit zijn de zogenaamde primitief recursieve relaties, die we definiëren aan de hand van primitief recursieve functies.

Definitie. Een getaltheoretische functie is *primitief recursief* als het één van de volgende basisfuncties is:
- de nulfunctie $z(x) = 0$
- de successorfunctie $s(x) = x + 1$
- de projectiefuncties $p_i^k(x_1, \ldots, x_k) = x_i$

of via eindig veel toepassingen van de volgende regels hieruit geconstrueerd kan worden:
- als $g(x_1, \ldots, x_k), h_1(x_1, \ldots, x_m), \ldots, h_k(x_1, \ldots, x_m)$ primitief recursieve functies zijn, dan is ook
$$f(x_1, \ldots, x_m) = g(h_1(x_1, \ldots, x_m), \ldots, h_k(x_1, \ldots, x_m))$$
primitief recursief
- als $g(x_1, \ldots, x_k)$ en $h(x_1, \ldots, x_{k+2})$ primitief recursief zijn, dan is ook
$$f(x_1, \ldots, x_k, x) = \begin{cases} g(x_1, \ldots, x_k) & \text{als } x = 0 \\ h(x_1, \ldots, x_k, x-1, f(x_1, \ldots, x_k, x-1)) & \text{anders} \end{cases}$$
primitief recursief.

Definitie. Een getaltheoretische relatie $R(x_1, \ldots, x_k)$ is *primitief recursief* als de karakteristieke functie
$$f_R(x_1, \ldots, x_k) = \begin{cases} 1 & \text{als } R(x_1, \ldots, x_k) \text{ geldt} \\ 0 & \text{anders} \end{cases}$$
primitief recursief is.

Hoewel de regels voor het construeren van primitief recursieve functies er ingewikkeld uit zien, valt het mee: de eerste regel zegt dat de samenstelling van primitief recursieve functies weer primitief recursief is, en de tweede regel zegt dat een functie die recursief (hier in de meer gebruikelijke betekenis) gedefiniëerd is aan de hand van primitief recursieve functies, zelf ook primitief recursief is.

Ontzettend veel relaties en functies zijn primitief recursief. In het bijzonder zal straks blijken dat deze sterk genoeg zijn om eigenschappen van Gödelgetallen te bepalen. Daarom is de volgende stelling van groot belang.

Stelling. In een type A-systeem dat de getaltheorie bevat is iedere primitief recursieve relatie uitdrukbaar.

Het bewijs hiervan is lang en technisch, en laten we daarom achterwege. In principe is het echter niet moeilijk. Het idee is om de definitie van uitdrukbaarheid uit te breiden naar getaltheoretische functies (dit is vrij eenvoudig, omdat een functie een speciaal soort relatie is). Dan bewijzen we eerst dat iedere basisfunctie uitdrukbaar is, en vervolgens dat een functie die geconstrueerd wordt met één van de twee regels ook uitdrukbaar is, als de functies waaruit deze geconstrueerd werd uitdrukbaar zijn. Dat ook primitief recursieve relaties uitdrukbaar zijn volgt dan vrij direct.

In het bewijs van Gödels stelling speelt de relatie

$$\text{bew}(x,y) \Leftrightarrow x \text{ is het Gödelgetal van een bewijs voor de welgevormde formule met Gödelgetal } y$$

een grote rol. Informeel gezegd geldt $\text{bew}(x,y)$ dus als x een bewijs is voor y. Verder is de functie

$\text{sub}(y,z,k)$ = het Gödelgetal van de formule die ontstaat door in de welgevormde formule met Gödelgetal y ieder ongebonden voorkomen van de variabele met Gödelgetal z te vervangen door s^k0 (als y niet het Gödelgetal van een welgevormde formule is, of z niet het Gödelgetal van een variabele, neemt de functie de waarde 0 aan)

van belang. Deze functie bepaalt dus, informeel gezegd, de formule die ontstaat door in de welgevormde formule y ieder ongebonden voorkomen van de variabele z te vervangen (substitueren) door s^k0.

We zouden graag willen dat $\text{bew}(x,y)$ en $\text{sub}(y,z,k)$ primitief recursief zijn, maar dat is niet in het algemeen het geval. Het probleem is dat we hiervoor bijvoorbeeld moeten kunnen bepalen of een welgevormde formule een axioma is, en dit kan in het algemeen heel ingewikkeld zijn. Om dit op te lossen eisen we dat de collectie Gödelgetallen van axioma's primitief recursief is (een collectie natuurlijke getallen is primitief recursief als de relatie $R(x)$, die geldt als x in de collectie zit, primitief recursief is), en nog enkele vergelijkbare eisen. Merk op dat we een afleidingsregel $\{\phi_1,\ldots,\phi_n\} \to \phi$ kunnen zien als het rijtje $\phi_1,\ldots,\phi_n,\phi$ en daarom een Gödelgetal kunnen geven. Met deze informatie zijn we in staat 'geschikte' systemen te selecteren: we noemen een formeel systeem *primitief recursief* als de collectie Gödelgetallen van welgevormde formules, de collectie Gödelgetallen van axioma's en de collectie Gödelgetallen van afleidingsregels allemaal primitief recursief zijn.[4] Het volgende lemma toont aan dat $\text{bew}(x,y)$ en $\text{sub}(y,z,k)$ voor deze systemen wel altijd primitief recursief zijn.

Lemma. Voor een primitief recursief type A-systeem dat de getaltheorie bevat zijn de relatie $\text{bew}(x,y)$ en de functie $\text{sub}(y,z,k)$ primitief recursief.

Het bewijs bestaat eruit de relatie en functie stap voor stap op te bouwen uit de primitief recursieve basisfuncties. Gödel heeft daarvoor een lijst met 45 primitief recursieve functies en relaties nodig, en we zullen het bewijs hier niet uitschrijven. We benadrukken nogmaals dat de eis dat het systeem primitief recursief is, niet veel meer zegt dan dat we moeten kunnen bepalen welke formules welgevormd zijn, welke welgevormde formules axioma's zijn, en wat de afleidingsregels zijn.

4 De onvolledigheidsstelling

We hebben nu eindelijk genoeg kennis om Gödels onvolledigheidsstelling te kunnen bewijzen. Het bewijs komt erop neer dat we binnen het formele systeem een uitspraak maken die zegt: 'Deze uitspraak is onbewijsbaar'. Vervolgens tonen we aan dat noch deze uitspraak noch zijn ontkenning een stelling is, wat betekent dat het systeem onvolledig is. We beginnen dus met het construeren van deze uitspraak.

Het uitgangspunt is de relatie $\text{bew}(x,y)$, die geldt als x een bewijs is voor y. We willen (ongeveer) een welgevormde formule maken die deze relatie uitdrukt, waarbij op de positie van y een uitdrukking komt te staan die precies de waarde aanneemt van het Gödelgetal van de gehele formule. Dan hebben we immers een formule die iets zegt over zijn eigen bewijsbaarheid. De truc die Gödel hiervoor gebruikt, is een speciaal geval van de functie $\text{sub}(y,z,k)$, namelijk het geval waarin $k = y$. De functie $\text{sub}(y,z,y)$ vervangt in feite in een formule y ieder ongebonden voorkomen van de variabele z door s^y0 (onthoud dat y en z eigenlijk getallen zijn, en dat de functie ook een getalswaarde aanneemt). Het is deze vorm

[4] Een formeel systeem waarin de welgevormde formules, axioma's en afleidingsregels worden gegeven door een eindig aantal regels, gebaseerd op de structuur van formules, zoals in P, is altijd primitief recursief.

van zelfreferentie die ons in staat stelt om een welgevormde formule indirect een uitspraak over zichzelf te laten doen. We beschouwen een nieuwe relatie:

$$\mathrm{w}(x, y) \Leftrightarrow \mathrm{bew}(x, \mathrm{sub}(y, 12, y)).$$

Het getal 12 is hier gekozen omdat het in P het Gödelgetal van de variabele x_2 is. In een ander systeem kan dit een ander (vastliggend) getal zijn, maar voor het gemak schrijven we hier 12.

Een belangrijke opmerking is, dat als $\mathrm{bew}(x, y)$ en $\mathrm{sub}(y, z, k)$ primitief recursief zijn, dan $\mathrm{w}(x, y)$ dat ook is. De reden daarvoor is dat de karakteristieke functie van $\mathrm{w}(x, y)$ gelijk is aan de samenstelling van de karakteristieke functie van $\mathrm{bew}(x, y)$ met $\mathrm{sub}(y, 12, y)$, en dus primitief recursief is. (Een constante functie die altijd gelijk aan 12 is maken we door herhaald de successorfunctie samen te stellen met de nulfunctie.) Nu weten we dat voor een primitief recursief type A-systeem dat de getaltheorie bevat $\mathrm{bew}(x, y)$ en $\mathrm{sub}(y, z, k)$ inderdaad primitief recursief zijn, en dus ook $\mathrm{w}(x, y)$. Er volgt ook dat $\mathrm{w}(x, y)$ in deze systemen is uit te drukken. De welgevormde formule die $\mathrm{w}(x, y)$ uitdrukt zullen we noteren met $\mathrm{W}(x_1, x_2)$, let op de hoofdletter om het verschil tussen de relatie en de formule aan te geven. We herhalen nog eens informeel wat $\mathrm{W}(x_1, x_2)$ zegt:

$\mathrm{W}(x_1, x_2)$: 'x_1 is een bewijs voor de formule x_2 waarin de variabele x_2 is vervangen door $s^{x_2}0$.'

(Merk op dat x_2 in deze zin op twee manieren wordt gebruikt: als getal en als symbool voor een variabele. Verder zullen we vanaf nu niet meer zeggen dat alleen ieder ongebonden voorkomen van de variabele wordt vervangen, omdat dit niet van belang zal zijn.) Beschouw nu deze welgevormde formule:

$$F\colon \quad (\forall x_1)\neg\mathrm{W}(x_1, x_2).$$

Hier staat in feite dat iedere x_1 geen bewijs is voor de genoemde formule, oftewel dat deze formule onbewijsbaar is. We hoeven alleen nog te zorgen dat de formule waarover we een uitspraak doen, gelijk is aan de formule die de uitspraak doet. Laat daartoe f het Gödelgetal van F zijn, en neem de volgende welgevormde formule:

$$G\colon \quad (\forall x_1)\neg\mathrm{W}(x_1, s^f0).$$

Deze formule kunnen we interpreteren als de uitspraak dat de formule met Gödelgetal f (dus F) waarin de variabele x_2 is vervangen door s^f0, onbewijsbaar is. Maar deze formule is precies G, dus we zien dat G zegt: 'G is onbewijsbaar'. Het Gödelgetal van G is bovendien $\mathrm{sub}(f, 12, f)$, eveneens omdat G de formule is die we krijgen door in de formule met Gödelgetal f (dus F) de variabele x_2 (die Gödelgetal 12 heeft) te vervangen door s^f0.

Nu we deze formule geconstrueerd hebben, kunnen we de stelling van Gödel bewijzen. We hoeven alleen nog maar na te gaan dat G en $\neg G$ allebei geen stellingen zijn.

Eerste onvolledigheidsstelling van Gödel. Ieder primitief recursief type A-systeem dat de getaltheorie bevat en ω-consistent is, is onvolledig.

Bewijs. We bewijzen eerst uit het ongerijmde dat G geen stelling is. Neem dus aan dat G wel een stelling is, dan is er een bewijs van G. We noemen het Gödelgetal van dit bewijs n. Omdat $\mathrm{sub}(f, 12, f)$ het Gödelgetal van G is, geldt $\mathrm{bew}(n, \mathrm{sub}(f, 12, f))$, en per definitie geldt dan ook $\mathrm{w}(n, f)$. De relatie $\mathrm{w}(x, y)$ wordt in het systeem uitgedrukt door de formule $\mathrm{W}(x_1, x_2)$. Omdat $\mathrm{w}(n, f)$ geldt, volgt hieruit dat $\mathrm{W}(s^n0, s^f0)$ een stelling is. Aan de andere kant kunnen we het volgende bewijs in het formele systeem uitvoeren, omdat G een stelling is:

1. $(\forall x_1)\neg\mathrm{W}(x_1, s^f0)$	aanname
2. $((\forall x_1)\neg\mathrm{W}(x_1, s^f0)) \Rightarrow \neg\mathrm{W}(s^n0, s^f0)$	axioma 5
3. $\neg\mathrm{W}(s^n0, s^f0)$	afleidingsregel 2 met regel 1 en 2

Als we dit laten voorafgaan door het bewijs voor G, hebben we een bewijs voor $\neg\mathrm{W}(s^n0, s^f0)$ gevonden. Maar nu zijn $\mathrm{W}(s^n0, s^f0)$ en $\neg\mathrm{W}(s^n0, s^f0)$ beide stellingen, en is het systeem

dus niet consistent. We hebben al bewezen dat het systeem dan ook niet ω-consistent is, in tegenspraak met de aanname. We concluderen dat G geen stelling is.
Aangezien G geen stelling is, heeft G geen bewijs. Voor ieder natuurlijk getal n geldt de relatie $\mathrm{bew}(n, \mathrm{sub}(f, 12, f))$ dus niet, en uiteraard $\mathrm{w}(n, f)$ dan ook niet. Omdat $\mathrm{W}(x_1, x_2)$ de relatie $\mathrm{w}(x, y)$ uitdrukt, betekent dit dat $\neg\mathrm{W}(s^n 0, s^f 0)$ een stelling is voor ieder natuurlijk getal n. Omdat het systeem ω-consistent is, kan dan niet ook de welgevormde formule

$$\neg(\forall x_1)\neg\mathrm{W}(x_1, s^f 0)$$

een stelling zijn. Maar deze formule is precies $\neg G$, dus $\neg G$ is geen stelling.
We hebben gezien dat noch G noch $\neg G$ een stelling is, dus het systeem is onvolledig. $\square$

We merken nog twee dingen op over de stelling. Ten eerste, de aanname van ω-consistentie kan worden verzwakt tot consistentie. Dit is in 1936 bewezen door J.B. Rosser (zie ook [2]). Daarvoor gebruikte hij een iets andere formule G. We hebben hier Gödels originele bewijs behandeld omdat Rossers aanpassing minder inzichtelijk is.

Ten tweede benadrukken we dat deze stelling zo krachtig is omdat ze betrekking heeft op élk geschikt systeem. Het zou bijvoorbeeld een oplossing kunnen lijken om G als axioma aan het systeem toe te voegen.[5] Echter, dan is de notie van een bewijs enigszins veranderd, en daarmee de formule $\mathrm{W}(x_1, x_2)$, zodat we een andere formule G' krijgen waarvoor G' en $\neg G'$ beide niet bewijsbaar zijn. De ware betekenis van Gödels onvolledigheidsstelling is dus dat eruit blijkt dat het formele systeem als methode niet sterk genoeg is om de gehele getaltheorie te kunnen beschrijven.

Referenties

[1] K. Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatshefte für Mathematik und Physik, vol. 38, pag. 173–198, 1931. Engelse vertaling: M. Hirzel, `www.research.ibm.com/people/h/hirzel/papers/canon00-goedel.pdf`.

[2] E. Nagel en J.R. Newman, Gödel's Proof (Revised Edition), New York University Press, 2001.

[3] A.G. Hamilton, Logic for Mathematicians, Cambridge University Press, 1978.

[5] We kunnen $\neg G$ niet als axioma toevoegen, omdat dit tot ω-inconsistentie zou leiden, zoals uit het bewijs van Gödels stelling blijkt. Dit is de reden waarom wordt gezegd dat G, hoewel onbewijsbaar, wel waar is.

Appendix: het formele systeem P

Het formele systeem dat we hier geven is gebaseerd op de aanpak in [3]. We definiëren P door achtereenvolgens de symbolen, welgevormde formules, axioma's en afleidingsregels aan te geven. Daartussen voeren we ook nog de begrippen 'variabele', 'term' en 'gebonden variabele' in, om de definitie wat overzichtelijker te maken. Merk op dat deze begrippen geen direct onderdeel van het formele systeem zijn.

- *symbolen*: $\neg$, $\Rightarrow$, $\forall$, (,), 0, s, $+$, $\times$, $=$, x_1, x_2, x_3, ... Opmerking: het symbool s zullen we gebruiken om de successorfunctie $s(x) = x + 1$ uit te drukken (maar heeft op dit moment nog geen betekenis).
- *variabelen*: de symbolen x_1, x_2, x_3, ...
- *termen*: de formules van lengte 1 met het symbool 0 of een variabele zijn termen; als t_1 en t_2 termen zijn, dan zijn ook st_1, $(t_1 + t_2)$ en $(t_1 \times t_2)$ termen.
- *welgevormde formules*: als t_1 en t_2 termen zijn, dan is $t_1 = t_2$ een welgevormde formule; als ϕ en ψ welgevormde formules zijn en x_i een variabele, dan zijn $(\neg\phi)$, $(\phi \Rightarrow \psi)$ en $(\forall x_i)\phi$ ook welgevormde formules.
- *gebonden variabele*: een voorkomen van de variabele x_i in een welgevormde formule ϕ heet gebonden als deze direct wordt voorafgegaan door het symbool $\forall$, of als deze bevat is in een welgevormde deelformule ψ van ϕ (dus ψ is een aaneengesloten deelrijtje van ϕ dat het voorkomen van x_i bevat, en ψ is een welgevormde formule) die direct wordt voorafgegaan door $(\forall x_i)$.
- *axioma's*: als ϕ, ψ en ξ welgevormde formules zijn, x_i een variabele, en t_1, t_2 en t_3 termen, dan zijn de volgende welgevormde formules axioma's:
  1. $(\phi \Rightarrow (\psi \Rightarrow \phi))$
  2. $((\phi \Rightarrow (\psi \Rightarrow \xi)) \Rightarrow ((\phi \Rightarrow \psi) \Rightarrow (\phi \Rightarrow \xi)))$
  3. $(((\neg\phi) \Rightarrow (\neg\psi)) \Rightarrow (\psi \Rightarrow \phi))$
  4. $((\forall x_i)\phi \Rightarrow \phi)$ als x_i niet ongebonden in ϕ voorkomt
  5. $((\forall x_i)\phi(x_i) \Rightarrow \phi(t_1))$ als geen van de ongebonden variabelen in t_1 gebonden is op de positie van x_i in ϕ (hier is $\phi(t_1)$ de formule ϕ met daarin ieder ongebonden voorkomen van x_i vervangen door t_1)
  6. $((\forall x_i)(\phi \Rightarrow \psi) \Rightarrow (\phi \Rightarrow (\forall x_i)\psi))$ als x_i niet ongebonden in ϕ voorkomt
  7. $x_i = x_i$
  8. $(t_1 = t_2 \Rightarrow st_1 = st_2)$
  9. $(t_1 = t_2 \Rightarrow (t_1 + t_3) = (t_2 + t_3))$
  10. $(t_1 = t_2 \Rightarrow (t_3 + t_1) = (t_3 + t_2))$
  11. $(t_1 = t_2 \Rightarrow (t_1 \times t_3) = (t_2 \times t_3))$
  12. $(t_1 = t_2 \Rightarrow (t_3 \times t_1) = (t_3 \times t_2))$
  13. $(t_1 = t_2 \Rightarrow (t_1 = t_3 \Rightarrow t_2 = t_3))$
  14. $(t_1 = t_2 \Rightarrow (t_3 = t_1 \Rightarrow t_3 = t_2))$
  15. $(\forall x_1)(\neg sx_1 = 0)$
  16. $(\forall x_1)(\forall x_2)(sx_1 = sx_2 \Rightarrow x_1 = x_2)$
  17. $(\forall x_1)(x_1 + 0) = x_1$

18. $(\forall x_1)(\forall x_2)(x_1 + sx_2) = s(x_1 + x_2)$

19. $(\forall x_1)(x_1 \times 0) = 0$

20. $(\forall x_1)(\forall x_2)(x_1 + sx_2) = ((x_1 \times x_2) + x_1)$

21. $(\phi(0) \Rightarrow ((\forall x_i)(\phi(x_i) \Rightarrow \phi(sx_i)) \Rightarrow (\forall x_i)\phi(x_i)))$ waar $\phi(t)$ steeds de formule ϕ is met daarin ieder ongebonden voorkomen van de variabele x_i vervangen door de term t

– *afleidingsregels*: als ϕ en ψ welgevormde formules zijn en x_i een variabele, dan zijn

  1. $\{\phi\} \rightarrow (\forall x_i)\phi$
  2. $\{\phi, (\phi \Rightarrow \psi)\} \rightarrow \psi$

  afleidingsregels.

Om te voorkomen dat we het overzicht verliezen, zullen we welgevormde formules in P slordig noteren waar duidelijk is wat bedoeld wordt. Dit gebeurt op twee manieren. Ten eerste gebruiken we afkortingen, bijvoorbeeld

$$\begin{array}{rll} (\phi \vee \psi) & \text{voor} & ((\neg\phi) \Rightarrow \psi), \\ (\phi \wedge \psi) & \text{voor} & (\neg(\phi \Rightarrow (\neg\psi))), \\ (\exists x_i)\phi & \text{voor} & (\neg(\forall x_i)(\neg\phi)), \\ s^n 0 & \text{voor} & \underbrace{ss\ldots s}_{n \text{ keer}}0. \end{array}$$

(Merk op dat de gebruikelijke interpretatie van $s^n 0$ dus het getal n is.) Ook zullen we haakjes weglaten als duidelijk is wat we bedoelen. Daarbij houden we ons aan de gebruikelijke conventies.