

## Excerpt

**Douglas R. Hofstadter, The Turing Test: A Coffeehouse Conversation**

G.J.E. Rutten

Dit artikel betreft een dialoog tussen Chris (natuurkunde student), Pat (biologie student) en Sandy (filosofie student) over de vraag of computers kunnen denken. Aanleiding van het gesprek is het artikel ‘Computing Machinery and Intelligence’ van Alan Turing. Chris en Pat staan allebei enigszins sceptisch tegenover artificiële intelligentie terwijl Sandy kunstmatige intelligentie juist verdedigt vanuit een functionalistische positie. In de besprekking van deze dialoog zal ik mij beperken tot het aanduiden van enkele hoofdmomenten.

Sandy vertelt dat de Turing test gebaseerd is op een imitatie spel waarbij een ondervrager via bijvoorbeeld een telex een man en een vrouw ondervraagt die zich elk in een andere kamer bevinden. De ondervrager moet achterhalen wie van beide gesprekspartners de vrouw is. De man probeert hierbij de ondervrager ervan te overtuigen dat hij de vrouw is. Hetzelfde geldt voor de vrouw. De Turing test is een variant op dit spel waarbij zich een machine in de ene en een mens in de andere kamer bevindt. De ondervrager moet nu achterhalen wie de mens is.

Pat merkt op dat er een groot verschil bestaat tussen beide testen. Wanneer de man de ondervrager ervan weet te overtuigen dat hij de vrouw is dan bewijst dit niet dat hij ook echt een vrouw is. Waarom zouden we dan concluderen dat een machine daadwerkelijk kan denken wanneer deze de ondervrager ervan weet te overtuigen dat hij kan denken? Chris stelt eveneens dat de Turing test niet als een test voor ‘denkende machines’ beschouwd kan worden omdat het imitatie spel ook niet als een test voor ‘vrouw zijn’ beschouwd kan worden. Het enige dat de Turing test volgens Pat lijkt te toetsen is of een machine in staat is denken te *simuleren*. Chris sluit zich hierbij aan. Niemand zal toch simulaties in een computer van bijvoorbeeld vloeistoffen, gassen of sterrenstelsels willen verwarringen met deze zaken zelf?

Het antwoord van Sandy op deze tegenwerping is typisch functionalistisch. Hoewel ze toegeeft dat het best raar is om te beweren dat bijvoorbeeld een computergesimuleerde orkaan daadwerkelijk een orkaan is, is het volgens haar toch niet zo raar als het lijkt. De patronen ofwel structuren van beiden komen immers overeen. Dit is zeker het geval bij een denkbeeldige *diepe* simulatie waarbij iedere atoom van de orkaan gesimuleerd wordt. Iemand die met de juiste ‘decodeer’ bril naar een diepe computersimulatie van een orkaan kijkt kan inzien dat deze simulatie qua structuur isomorf is met wat wij in de regel een orkaan noemen. De claim dat een voldoende diepe simulatie géén orkaan is wordt zo ‘ad absurdum’ gevoerd. Sandy vervolgt haar functionalistische betoog door op te merken dat juist bij een fenomeen als ‘denken’ de abstracte structuur en niet het fysieke medium bepalend is. Iedere geslaagde simulatie van denken (met Turing’s test als criterium) is dus niets anders dan denken zelf.

Een tegenwerping van Chris dat de interne structuur van een computer welke de Turing test doorstaat niet gelijk hoeft te zijn aan de interne structuur van onze geest wordt door Sandy weerlegd met de opmerking dat de functie van de Turing test analoog is aan de functie van

een deeltjesversneller voor fysici. Door het afvuren van vragen op een gesprekspartner kan de interne denkstructuur achterhaald worden net zoals een deeltjesversneller ons helpt de interne structuur van atomen te achterhalen door op deze atomen elementaire deeltjes af te vuren.

Pat stelt dat het al dan niet aanwezig zijn van een diepe gemeenschappelijke biologische oorsprong van cruciaal belang is voor het beantwoorden van de vraag of een bepaalde entiteit wel of niet kan denken. De Turing test gaat geheel voorbij aan dit volgens hem belangrijke aspect van het hebben van dezelfde organische oorsprong als de mens. Sandy is het hiermee niet eens. Het bezitten van DNA en een menselijke vorm is slechts indirect bewijs. Bovendien is volgens haar alleén de interne structuur bepalend voor de vraag of iets denkt.

Volgens Sandy is het onmogelijk om denken en emoties van elkaar te scheiden. Denken zonder emoties is volgens Sandy onmogelijk omdat denkende entiteiten gedreven worden door onderliggende intenties, motivaties, neigingen, interesses en wensen. Bovendien moeten denkende entiteiten in staat zijn om in iedere situatie aan bepaalde dingen meer belang toe te kennen dan aan andere dingen. Het testen van emotionele reacties is voor Sandy dan ook een belangrijk onderdeel van de Turing test. Chris neemt hier een geheel andere positie in. Hij ziet juist geen tegenspraak in een denkende machine zonder gevoelens.

Sandy bespreekt ook enkele volgens haar wezenlijke kenmerken van bewustzijn. Bewustzijn ontstaat vanuit een specifiek organisatiepatroon. Bewustzijn vereist een structuur welke in staat is om de externe wereld intern te representeren en vervolgens op basis van deze interne representatie adequaat op de buitenwereld te reageren. Een bewuste entiteit moet bovendien beschikken over een rijk en flexibel model van zichzelf. Geen van de huidige machines beschikt volgens haar over deze eigenschappen. Zo heeft bijvoorbeeld een schaakprogramma een veel te beperkt zelfmodel omdat het programma niet weet waarom het schaak speelt, dat het een schaakprogramma is, dat het zich in een computer bevindt en een menselijke opponent heeft. Het heeft ook geen gevoel van trots of schaamte bij winst of verlies.

Het vermogen om te denken, het vermogen om te voelen en het hebben van bewustzijn zijn voor Sandy uiteindelijk allemaal verschillende facetten van hetzelfde fenomeen. Geen van hen kan aanwezig zijn zonder de ander. Emoties (en dus ook denken en bewustzijn) ontstaan vanzelf op het hoogste niveau wanneer we op het laagste niveau voldoende gevoelloze calculaties op een gecoördineerde wijze bij elkaar voegen. Zodra de interne structuur van een machine in de juiste richting voldoende complex wordt zal de machine vanzelf een intentioneel gezichtspunt met betrekking tot zichzelf innemen. We kunnen (en moeten) de machine dan zien als een denkend wezen en ons taalgebruik daarop aanpassen. Dit zien we nu al héél beperkt bij schaakprogramma's waarover we soms ook praten in antropomorfe termen.

Sandy geeft toe aan Chris dat het slagen voor de Turing test geen noodzakelijke voorwaarde kan zijn voor bewustzijn en denken. Bepaalde zoogdieren zullen deze test immers niet doorstaan terwijl ze volgens Sandy wel een bewustzijn hebben. Het slagen voor de Turing test is echter wel een voldoende voorwaarde voor het hebben van bewustzijn en kunnen denken.

Pat vindt de gedachte van computers met bewustzijn moeilijk aanvaardbaar. Voor hem is

bewustzijn verbonden met zachte en warmbloedige lichamen. Mensen beschikken in tegenstelling tot computers over iets van een onvoorspelbare creatieve ‘vlam’. Volgens Sandy wordt dit vooroordeel veroorzaakt door het feit dat we een mechanistisch en materialistisch beeld hebben van machines welke is overgeleverd vanuit de industriële revolutie. Het idee van een vlam is echter volgens haar ook van toepassing op computers. Op het beeldscherm van computers is bijvoorbeeld sprake van chaotisch flikkerende en wild heen en weer schietende patronen van lichtflitsen. We kunnen computers in onze verbeelding dus net zo goed voorstellen als onstuimig patronen van pulserende lichtflitsen in plaats van als gigantische mechanische stoommachines.

Volgens Sandy kunnen de huidige computers niet denken en ook niet geprogrammeerd worden om te denken. Wel zullen computerarchitecturen zich sterk blijven ontwikkelen zodat op een gegeven moment denkende computers binnen handbereik komen. In de toekomst zullen er geavanceerde computers worden gebouwd waarvan de architectuur radicaal zal verschillen van wat wij nu computers noemen. Er zullen gespecialiseerde computers worden ontwikkeld die specifiek bedoeld zijn om te denken ofwel intelligent gedrag te vertonen. Zij zullen zich fysiek kunnen voortbewegen en bovendien over op zintuigen lijkende systemen beschikken. Desalniettemin zullen ze op het laagste niveau uit niets anders dan uit mechanische materiële bouwstenen bestaan.

Sandy geeft aan dat computers (net zoals mensen) fouten kunnen maken. Zij doelt hierbij niet op hardware fouten maar op zogenaamde functionele fouten. Een computer kan bijvoorbeeld het weer van morgen fout voorspellen. Naarmate computers intelligenter worden zullen ze vaker problemen tegenkomen waarvoor geen voldoende nauwkeurige modellen bestaan. Intelligentere computers zullen dan net zoals mensen noodzakelijk met vage en onvolledige modellen moeten gaan werken ofwel iets ‘vaags en glibberigs’ moeten bezitten. Gevolg is dat intelligentere computers juist vaker functionele fouten zullen maken en ook vaker onvoorspelbaar gedrag zullen vertonen. Volgens Sandy kan het ‘vage en glibberige’ bij intelligente computers in de software zelf zitten.

Samengevat zal daadwerkelijke artificiële intelligentie wanneer het zich aandient volgens Sandy zowel mechanisch (op het laagste niveau) als organisch (op het hoogste niveau) zijn. Mensen zijn volgens haar dan ook niets anders dan extreem complexe machines. Niet voor niets bestaan we volgens haar op het laagste niveau restloos uit mechanisch werkend DNA.

Volgens Sandy vinden mensen de gedachte dat zij een machine zijn moeilijk te accepteren omdat ze niet graag geconfronteerd willen worden met hun eigen sterfelijkheid. Naast deze emotionele weerstand is er ook een meer epistemologische weerstand. Het valt voor een mens namelijk niet in detail te begrijpen hoe de hogere niveaus van bewustzijn, emoties en denken fysisch veroorzaakt worden door het louter mechanische laagste niveau. Het is voor een mens niet mogelijk om intuïtief ineens de stap te maken van het laagste mechanische niveau naar het hoogste intentionele niveau omdat er allerlei tussenniveaus zijn die als een schild werken.

Aan het eind van het gesprek noemt Sandy een programma genaamd ‘Parry’ welke volgens de bedenkers een elementaire versie van de Turing test doorstaan zou hebben. Dit programma selecteert voorgebakken zinnen uit een database in plaats van zelf zinnen te maken. Deze

aanpak werkt volgens Sandy echter niet bij minder eenvoudige versies van de Turing test.

Tot slot vraagt Chris aan Sandy welke vraag zij zou willen stellen als zij tijdens een Turing test maar één vraag mag stellen. Pat suggereert om in dat geval de vraag van Chris te stellen!

### **Discussievragen**

1. Sandy stelt in de dialoog dat emoties, denken en bewustzijn niets anders zijn dan complexe interne structuren ofwel patronen welke voortkomen uit het op het laagste niveau gecoördineerd combineren van zeer grote aantallen elementaire mechanische bouwstenen. De mens is dus niets meer of minder dan een enorm complexe machine.
  - Is met deze opvatting het qualia probleem voldoende geadresseerd?
  - Zo ja, is deze opvatting een bevredigend antwoord op het qualia probleem?
2. Sandy neemt een functionalistische positie in omdat zij fenomenen als emoties, denken en bewustzijn wil begrijpen als complexe structuren ofwel patronen (zoals ‘software’). Het medium waarin deze fenomenen gerealiseerd zijn is voor een functionalist niet van belang. Volgens het functionalisme is er dus geen wezenlijk verschil tussen een voldoende diepe simulatie van bijvoorbeeld een orkaan en de orkaan zelf. Beiden bezitten immers dezelfde structuur. Sandy beweert dat het ‘gesimuleerde’ denken door een machine welke de Turing test doorstaat niet verschilt van ons denken. Volgens het functionalisme betekent dit dat het gedeelte van de machine dat verantwoordelijk is voor het denken dezelfde interne structuur moet hebben als ons denken.
  - Geeft Sandy voor deze structuurovereenkomst een argument?
  - Zo ja, welk? Is haar argument overtuigend?
3. In de beschouwing aan het eind van de dialoog wordt terecht opgemerkt dat er een verschil lijkt te bestaan tussen het representeren ofwel simuleren van materiële producten zoals melk (we kunnen de representatie niet drinken) en het representeren ofwel simuleren van abstracte artefacten zoals wiskundige bewijzen (de representatie van een bewijs is dat bewijs). Het functionalisme lijkt dus haar grenzen te hebben.
  - In welke categorie zal het representeren van de geest vallen volgens Sandy?
  - Is hiervoor een argument te geven (anders dan een herhaling van Sandy’s positie)?

