

George van Hal

Robots, aliens en popcorn

Wetenschap op het witte doek

Uitgeverij Atlas Contact
Amsterdam/Antwerpen

Inhoud

Trailer 7

Deel I: Aarde

1. *I, Robot*. De opmars van de zelfbewuste filmrobot 13
2. De echte *Iron Man* (en andere toffe superhelden) 47

Deel II: Ruimte

3. Kun je sneller vliegen dan het licht? 73
4. *A Space Odyssey*. De lange tocht naar andere sterren en planeten 99
5. Dromen van *E.T.* Zijn wij alleen in het universum? 120
6. De wetenschappelijke zoektocht naar buitenaards leven 150

Deel III: Voorbij ruimte en tijd

7. Tijdreizen in Hollywood 171
8. Filmlogica: (on)oplosbare tijdreisparadoxen 187
9. Je andere ik. Leven in een parallel universum 201

Bonus features

- A. De wetenschappelijk adviseurs van Hollywood 239
- B. De wetenschappelijke filmcanon 251

Aftiteling

- Dankwoord 295
Verantwoording 299

Trailer

Een zwart scherm. De muziek zwelt aan. Langzaam wordt het beeld lichter. We zien een ruimteschip dat opdoemt als een onderzeeër uit de atmosfeer van Jupiter-maan Titan. Dan begint een zware, warme voice-over.

Er bestaat een wereld...

We wisselen van scène.

...waar mensen dromen van een leven tussen de sterren...

We zien een ander schip. Dit vliegt door een opening in de ruimte en arriveert bij een majestueus, traag kolkend zwart gat.

...waar aliens op aarde arriveren, of ons signalen zenden uit de ruimte...

We zien een schattig buitenaards wezen. Tussen een groep kinderen in staart hij naar boven. Zijn vinger geeft licht.

...waar superhelden echt zijn...

Een man in een metalen pak vliegt door een helderblauwe lucht.

...gebouwd door ingenieurs.

Een onderzoeker houdt een indrukwekkend ogend metalen schild omhoog, terwijl een collega er met een machinegeweer een magazijn kogels op leegschiet.

Een wereld waar tijdreizen echt mogelijk is...

Ineens verschijnt uit het niets een naakte spierkolos in een steegje, terwijl bliksemschichten rond zijn lichaam dansen.

...en waar robots ons helpen, wanneer wij dat zelf niet meer kunnen.

Een glanzend witte thuishulprobot pakt een tuinslang vast en besproeit het groene gazon van een oude man.

Harde knip. Ineens ogen de beelden ouder, documentairechtig bijna. De muziek zwelt aan. Wordt heroïsch. Episch.

Wetenschappers en ingenieurs maken van die dromen werkelijkheid.

Een raket stijgt op en verlaat de dampkring van de aarde.

Zij bouwen apparaten die niemand ooit voor mogelijk hield.

Een man in een wit robotpak tilt een bejaarde vrouw haar bed in. Een verlamde vrouw geeft zichzelf met haar kunstarm een drankje.

Fantasie wordt eindelijk werkelijkheid.

We zien een stokoude film. Acteurs lopen verkleed als astronauten op een filmset die eruitziet als het maanoppervlak. De muziek komt tot een climax.

Ineens is er stilte. Een momentje, een adempauze lang, zien we niets. Dan verschijnt Neil Armstrong in beeld. Hij zakt het trapje van zijn ruimtecapsule af en zet als eerste mens voet op de maan. We horen zijn beroemd geworden, krakende stem: *'That's one small step for man, one giant leap for mankind'*. Het beeld wordt zwart. Ineens zien we woorden.

ROBOTS.

ALIENS.

EN POPCORN.

En daaronder: *Wetenschap op het witte doek*.

Dit is een boek voor iedereen die een beetje een dromer is. Zelfs wanneer je daar alleen aan toegeeft in het donker en met een grote bak popcorn op schoot. Wie durft te dromen, houdt vermoedelijk net als ik van film. Amerika noemt de filmindustrie in Hollywood niet voor niets ‘de droomfabriek’.

Of het nu gaat om de regisseur of scriptschrijver die een film maakt over superhelden die de wereld redden, of om de wetenschapper die zich aan een universiteit afvraagt: ‘Wat als ik iets kan maken dat harder en lichter is dan staal?’ – het begint allemaal met een droom, met een idee dat bijna alleen past op een groot wit doek.

Van dat soort ideeën hou ik. Ze voerden me na de middelbare school naar een studie sterrenkunde, een studie waarbij je antwoord probeert te vinden op de grootst mogelijke vragen in dit universum. Ze inspireerden me later ook om over die ideeën te schrijven, waarvan het boek dat je nu in je handen houdt een concreet resultaat is.

In *Robots, aliens en popcorn* bekijk ik de waanzinnige doorbraken uit wetenschap en technologie die perfect passen in het straatje van Hollywood. Dat is meestal geen toeval. Tussen de werelden van onderzoek en film bestaat een levendige kruisbestuiving, waarmee ze elkaar naar grote hoogtes stuwten.

Zou Neil Armstrong zijn eerste stappen op de maan hebben gezet zonder de inspirerende verhalen van Jules Vernes en oude films zoals *Le voyage dans la lune* of *Destination Moon*? Zouden de ruimtereizigers uit *Interstellar* naar hun verre kosmische bestemming zijn gevlogen via een wormgat, als theoretisch natuurkundigen die bijzondere optie niet eerst hadden bedacht? En zouden wij in het journaal stilstaan bij het WK robotvoetbal als schrijvers als Isaac Asimov en films als *Metropolis* het beeld van de robot niet eerst in ons gemeenschappelijk bewustzijn hadden geplant?

Nog een voorbeeld. Even lekker de krant lezen op je tablet kon al in 1968. Dat gebeurde in Stanley Kubricks briljante film *2001: A Space Odyssey*. De in die film getoonde gadget had volgens het tegelijkertijd

uitgegeven boek een naam die moderne lezers bekend in de oren zal klinken: de newspad. Waar wijlen Apple-CEO Steve Jobs zijn inspiratie voor de iPad vandaan haalde, is dus niet zo moeilijk te bedenken.

Wetenschappers en ingenieurs die de wereld van morgen vormgeven, halen ook nu nog vaak hun inspiratie uit Hollywood-verhalen. In dit boek lees je hoe een onderzoeker het befaamde schild van de superheld *Captain America* probeert na te bouwen, hoe ingenieurs sleutelen aan de eerste echte autonome robots en hoe een natuurkundige in een serieus wetenschappelijk tijdschrift beschreef dat je de sneller-dan-het-licht-motoren van de ruimteschepen uit *Star Trek* kunt nabouwen.

Op de komende pagina's ontdek je een bonte verzameling van fictie en feit, van wetenschap en film. We kijken door de lens van Hollywood naar serieuze wetenschappelijke zaken als de algemene relativiteitstheorie van Einstein, materiaalkunde, kunstmatige intelligentie en de zoektocht naar planeten buiten ons zonnestelsel. Maar ik neem je ook mee naar de meest speculatieve uithoeken op het snijvlak van fictie en kennis. Je reist door de tijd en komt oog in oog te staan met je moeder toen zij nog op de middelbare school zat, en maakt kennis met de man die al zijn mogelijke levens daadwerkelijk beleeft.

Robots, aliens en popcorn doet niet zuur over waar Hollywood de plank wetenschappelijk soms mislaat. Het is een boek over hoe wetenschappelijk onderzoek en de dromen op het witte doek elkaar continu snijden, raken en versterken. Want als één boodschap aan het eind duidelijk zal zijn, dan is het wel deze: de grootste dromen uit Hollywood beginnen eindelijk uit te komen. De toekomst is begonnen.

Deel 1: Aarde

1.

I, Robot. De opmars van de zelfbewuste filmrobot

Human beings have dreams. Even dogs have dreams, but not you, you are just a machine. An imitation of life. Can a robot write a symphony? Can a robot turn a... canvas into a beautiful masterpiece?

Can you?

– Robot Sonny beantwoordt de vraag van Detective Del Spooner op gevatte wijze in I, Robot (2004)

Pats. Daar knalt een halve grapefruit op een ontbijtbord. Aan tafel zit Frank, in zijn pyjama, klaar voor zijn ontbijt. Frank is een oude man bij wie de eerste tekenen van dementie zichtbaar zijn. Hij vergeet soms dingen. Zorgt minder goed voor zichzelf. Toch is Frank nog altijd scherp. En hij is absoluut nog niet van plan om naar een verzorgings-tehuis te gaan.

Daarom heeft zijn zoon Hunter tegen Franks zin in een thuiszorg-robot voor hem gekocht. En die heeft zojuist tot Franks verbazing een grapefruit op zijn bord gesmeten.

‘Breng me maar gewoon wat cornflakes,’ zegt Frank.

‘Cornflakes zitten vol ongezonde ingrediënten,’ antwoordt Robot.

‘Ik heb ze weggegooid.’

‘Zeg, je moet niet zomaar m’n spullen weggooien...’

‘Frank, cornflakes zijn voor kinderen. Geniet van je grapefruit.’

‘Jij bent voor kinderen, eikel.’

‘Vandaag gaan we een nieuwe tuin aanleggen,’ zegt Robot ineens, schijnbaar uit het niets.

‘Zak er maar in.’

‘Frank, je hebt een project nodig. Mentale stimulering gekoppeld aan een gedisciplineerd schema. Dat verbetert je cognitief functioneren enorm. Bovendien is beweging goed,’ zegt Robot. ‘We moeten leren samenwerken, Frank.’

Zucht. ‘Je bent een robotbutler.’

‘Ik ben geen butler, Frank. Ik ben een zorghulp, geprogrammeerd om je mentale en fysieke gezondheid in de gaten te houden.’

‘Hmm. Ja? Ga m’n huis uit!’

‘Als je niet meewerkt, kan ik net zo goed weggaan.’

Frank haalt zijn schouders op. ‘Mij best.’

‘Als dat is wat je wilt, dan bel ik Hunter wel.’

‘Mooi.’

Robot draait zijn hoofd een stukje bij. Hoewel hij geen ogen heeft, alleen een witte helm met een zwart vizier, wekt hij de indruk dat hij in de verte staart en ondertussen met iets anders bezig is.

‘Wat ben je aan het doen?’ vraagt Frank ineens zichtbaar bezorgd.

‘Heb je soms een telefoon daar in die hersenen van je? Ben je hem aan het bellen? Je hoorde toch wat hij zei? Hij wil me in een gekkenhuis stoppen!’

‘Ik herinner me niet dat Hunter dat heeft gezegd,’ antwoordt Robot op kalme toon.

‘Er is heus niks mis met m’n geheugen,’ zegt Frank. ‘Met mij gaat het prima.’ Hij telt van één tot en met tien om het te bewijzen. ‘Met m’n geheugen is echt niks mis...’ Diepe zucht. ‘Waar ben ik eigenlijk mee bezig? Ik praat met een huishoudelijk apparaat!’

‘Ik stel voor dat je meewerkt,’ zegt Robot.

‘IK GA NIET TUINIEREN,’ zegt Frank.

De toekomst waarin je thuiszorgrobot je net als in de film *Robot & Frank* (2012) beveelt om een nieuwe hobby te zoeken is verrassend dichtbij. Fabrieksrobots zetten al jaren onze auto’s in elkaar. Stofzuigrobots zuigen zelfstandig onze vloeren. En op dit moment rijden zelfsturende robotauto’s in Amerika al over de snelweg.

Onze geautomatiseerde hulpjes zijn zelfs actief buiten de damp-

kring. De Amerikaanse ruimtevaartorganisatie NASA heeft een menselijk ogende robotastronaut die zijn werk doet op het internationale ruimtestation ISS. In Japan is er de robot Kiribo, die oogt alsof hij is verwekt tijdens een vruchtbaar avontuurtje van een mens met een panda. Kiribo sprak in 2013 voor het eerst in de ruimte met echte astronauten. En op de maan en op Mars rijden terwijl je dit leest robotkarretjes rond die wetenschappelijke experimenten uitvoeren. Op die manier besparen ze ons dure en gevaarlijke bemande ruimtemissies. Robots zijn al lang geen sciencefiction meer.

Robots staan tegenwoordig zelfs op de politieke agenda. In Zuid-Korea stelt een overheidsdoelstelling dat in 2020 elk huishouden een eigen robot moet hebben. En in Nederland zwengelde minister Lodewijk Asscher eind 2014 nog een landelijke discussie aan over de gevolgen van de opmars van robots voor de werkgelegenheid.

Dat robots ons werkzorgen opleveren, is overigens heel toepasselijk. Het woord ‘robot’ dook voor het eerst op in 1920 in het toneelstuk *R. U. R. (Rossums Universele Robots)* van de Tsjechische toneelschrijver Karel Čapek. Daarin waren robots werkslaven die alle lastige klusjes mochten opknappen. Het woord is dan ook afgeleid van het Tsjechische woord *robotá*, dat grof vertaald ‘arbeid’, of ‘saai werk’ betekent.

Met dat in het achterhoofd is het niet vreemd dat we robots tegenwoordig inzetten voor klusjes als stofzuigen. Toch willen we het liefst dat robots veel meer kunnen, veel meer *zijn* misschien zelfs. De robots die je ziet in films spiegelen die droom. Ze zijn meer dan alleen handige metalen hulpjes. Ze hebben persoonlijkheid. Op die manier appelleren ze aan de wens om iets te scheppen naar ons eigen evenbeeld, een vorm van ‘leven’ waarin we onszelf herkennen. Neem bijvoorbeeld de mechanisch ogende robot R2-D2 uit *Star Wars* die ondanks zijn niet-menselijke uiterlijk met behulp van vrolijke of teleurgestelde bliepjes emoties overbrengt en meer persoonlijkheid heeft dan menig echt personage. Of denk aan het robotjongetje David uit de film *A.I. – Artificial Intelligence* uit 2001. David hunkert naar de liefde van een menselijke moeder en reist vervolgens de halve wereld af om die liefde te vinden. Dat zijn robots die ons laten voelen, die ons iets doen.

De robots op het witte doek zijn in zekere zin ook de vleesgeworden – of eigenlijk: *metaal*geworden – verpersoonlijking van onze wens te

beschikken over leven en dood. We blazen al sinds mensenheugenis levenloze voorwerpen fictief leven in, of het nu gaat om de Golems uit Joodse legendes, de houten pop uit *Pinokkio* of de aan elkaar genaaide mensendelen in *Frankenstein*. Robots trekken die verteltraditie door.

En net als Pinokkio of Frankenstein gaan robots in fictie nog een stapje verder dan alleen simpelweg bestaan. Vrijwel alle robots op het witte doek beschikken over een (rudimentaire) vorm van bewustzijn. Daardoor kunnen we ons in hen verplaatsen en denken we dat we weten wat er omgaat in hun kunstmatige brein. Dat gevoel zwingelt het ingebakken empathische centrum in ons eigen brein aan, zodat we met R2-D2 en David meeleven alsof het volwaardige personen zijn in plaats van dingen. En eigenlijk is dat ook zo.

Robots roepen allerlei emoties bij ons op. Soms vinden we ze schattig, of zelfs een tikje ontroerend, zoals het aandoenlijke robotje WALL-E uit de gelijknamige computeranimatiefilm uit 2008. WALL-E rijdt in zijn eentje rond op een door mensen verlaten aarde, maar blijft desondanks trouw zijn taken uitvoeren. En soms vinden we robots juist eng, zoals de intelligente, zelfbewuste moordrobot uit *Terminator* die nietsontziend jacht maakt op het menselijke hoofdpersonage.

Kortom: wanneer je een film wilt maken die je kijkers wat doet, is het wel zo verstandig je robots kunstmatig bewust te maken. En het toffe is dat dergelijke robots ook best buiten films kunnen bestaan. Voor robotonderzoekers van over de hele wereld vormen kunstmatig bewuste robots een belangrijk einddoel. Wetenschappelijke bespiegelingen over robotbewustzijn hebben zelfs een eigen vakblad, *Machine Consciousness*, waarin onderzoekers regelmatig nieuwe ontdekkingen publiceren. Kunstmatig bewustzijn is daarmee de heilige graal van het moderne robotonderzoek.

Wereldwijd bestaan honderden robotlabs die miljoenen dollars uitgeven aan onderzoek dat dergelijke robots een stapje dichterbij brengt. Je kunt het bijna zo gek niet bedenken of het gebeurt wel ergens. Robotonderzoekers werken aan mechanische hulpjes voor vermaak, aan ruimterobots, robots waarmee je fijn kunt babbelen, robots die emoties tonen, zorgrobots en oorlogsrobots die onze soldaten kunnen vervangen. Wanneer die robots in de toekomst dan ook nog eens zelfbewust worden, kun je stellen dat de robottoekomst uit Hollywood-films de-

finitief is begonnen. Over die opmars van de Hollywood-robot en de jacht op kunstmatig bewustzijn gaat de rest van dit hoofdstuk.

Op zoek naar kunstmatig bewustzijn

Welkom in de nabije toekomst. Op een avond zit je alleen thuis en je downloadt een nieuw besturingssysteem op al je apparaten – je computer, je telefoon enzovoort. Het is een alledaagse handeling, die niet veel verschilt van het downloaden van de nieuwe ios- of Android-versie voor je huidige smartphone, of de nieuwe Windows voor je laptop. Maar de technologie heeft de afgelopen paar jaar niet stilgestaan. Dit gloednieuwe besturingssysteem steekt daardoor net wat geraffineerder in elkaar dan wat er nu op je telefoon staat. Het is zelflerend. Het heeft daardoor wel wat weg van bijvoorbeeld de slimme thermostaten die nu al in steeds meer huizen hangen. Dergelijke apparaten leren wanneer je wel en niet thuis bent en op welk tijdstip van de dag je wilt dat het in huis lekker warm is. Dit nieuwe besturingssysteem leert ook en doet dat bovendien wat handiger dan zo'n thermostaat.

Je start het besturingssysteem op en beantwoordt vragen in de trant van: 'Ben je sociaal of introvert?', die ervoor moeten zorgen 'dat het os zo goed mogelijk op je behoefte aansluit'. Je kiest voor een vrouwelijke stem en zet je computer aan. Een verrassend menselijk klinkende vrouw begroet je en vertelt je dat ze Samantha heet. Zij is je nieuwe os.

Samantha helpt je met je dagelijkse klusjes. Ze plant je afspraken en zoekt antwoord op je vragen. Maar langzaam – zonder dat je er in eerste instantie zelf erg in hebt – verandert jullie verhouding. Je hebt steeds langere gesprekken met Samantha. Je lacht om haar grapjes en zij om de jouwe; je vraagt je af waar ze mee bezig is wanneer je telefoon uitstaat en je zorgt er steevast voor dat haar stem het laatste is wat je hoort voordat je gaat slapen en het eerste wanneer je opstaat. Uiteindelijk denk je vaker wel aan haar dan niet. En dan heb je het ineens door: je bent verliefd op het besturingssysteem van je telefoon.

Als je die laatste zin ziet staan zonder dat je de film *Her* hebt gezien, denk je vermoedelijk: bizar, dat kan nooit gebeuren. Wie wordt nu verliefd op een besturingssysteem? Maar in de film is het al wat minder

gek. Het besturingssysteem blijkt ronduit charmant (dat ze de stem van Scarlett Johansson heeft helpt daar ongetwijfeld bij), maar het allerbelangrijkste is dat ze *echt* lijkt. Je schrijft aan Samantha een eigen interne belevingswereld toe. Ze heeft passies, meningen en emoties. Ze is iemand. En op een iemand kun je best verliefd worden.

Samantha is daarmee een uitstekend voorbeeld van een kunstmatig bewust persoon. In dit geval is dat kunstmatige bewustzijn bovendien bijzonder, omdat Samantha in tegenstelling tot de robots waar we het eerder over hadden geen lichaam heeft. Ze bestaat alleen als digitale informatie en instructies in een complexe opeenvolging van nullen en enen ergens in het binnenste van een rij computerchips en kabels.

Daarom dringt de vraag zich direct op of je bij een rijtje instructies wel mag spreken van bewustzijn, ongeacht hoe ‘zij’ of ‘hij’ verder op je overkomt. Die vraag is voorlopig eerder filosofisch dan wetenschappelijk, maar ik sta er toch even bij stil, omdat veel van de wetenschap en technologie in dit hoofdstuk deze vraag vanzelf oproept.

Voordat je kunt zeggen of een robot of computerprogramma zelfbewust is, moeten we eerst antwoord hebben op de vraag: wat is bewustzijn überhaupt? Die vraag kun je niet gemakkelijk beantwoorden. Bewustzijn is een persoonlijke ervaring. Je weet van jezelf dat je bewust bent, maar kunt dat bij iemand anders lastig meten.

Bewijs maar eens dat je partner of beste vriend écht bewust is. Natuurlijk, zijn of haar handelingen impliceren dat wel, maar bewijzen dat in een mens bewustzijn zit en niet een vooraf gedefinieerde set instructies die bewustzijn simuleren, is lastig. Zo niet onmogelijk.

Nu denk je misschien: natuurlijk zijn mensen zelfbewust. En dan heb je een punt, want daar gaan we normaal gesproken wel van uit. Wanneer je twijfelend aan het bewustzijn van de mensen om je heen door het leven gaat, maak je het jezelf niet gemakkelijker. Maar bij robots is die vraag ineens veel prangender. Als je iets bouwt dat zelfbewust is, wanneer ben je daar dan in geslaagd? Wanneer heb je iets gebouwd dat niet alleen doet alsof?

In de film *A.I. – Artificial Intelligence* vraagt een robotonderzoeker aan een robot met de naam Sheila wat liefde nu eigenlijk is.

‘Liefde is eerst mijn ogen een stukje verder opendoen,’ zo begint ze haar antwoord, ‘en dan mijn ademhaling ietsje versnellen en mijn huid een beetje warmer maken, en het aanraken van...’

‘Enzovoorts,’ zegt de onderzoeker. ‘Dank je, Sheila.’

Het antwoord van Sheila toont aan dat zij het concept liefde niet begrijpt, maar alleen de biologische gevolgen heel nauwkeurig simuleert. Haar antwoord geeft haar in dit geval weg. Maar stel je die vraag niet, of programmeer je van tevoren een beter antwoord, dan kun je anderen er vermoedelijk best van overtuigen dat deze robot zelfbewust is en liefde voelt.

Natuurlijk maakt diezelfde onderzoeker later in *A.I.* wel een robot die liefde voelt en begrijpt. Of die in elk geval handelt alsof dat het geval is. Dat is David. En David begrijpt liefde niet alleen, maar heeft die zelfs nodig. Hij komt thuis bij een vrouw die hem heeft gekregen ter vervanging van haar eigen zoon, die in een kunstmatige coma ligt. David beschouwt haar als zijn nieuwe moeder. Wanneer zij in Davids brein vervolgens de liefdesschakel omzet – een onomkeerbaar proces – houdt hij onvoorwaardelijk van haar. Zelfs wanneer ze later, nadat haar eigen zoon uit zijn coma is ontwaakt, alsnog afstand van David doet omdat hij niet echt is.

In de film leef je mee met David, maar stel nu eens dat we in het echt ook een David bouwen. Accepteren we dan dat hij een persoon is? Of blijft hij een ding? Waar ligt de grens tussen simulatie en echt? Antwoorden op die vragen zijn lastig, en hoewel ik persoonlijk niet kan wachten totdat we ze vinden, zijn er ook mensen die er instinctief een beetje huiverig voor zijn. Want zodra wij iets kunnen maken dat kunstmatig bewust is, wat zegt dat dan over ons? Zijn wij dan in feite ook niets meer dan machines van vlees en bloed? Geen enkele ingenieur, zelfs geen enkele filosoof, heeft daarop een definitief antwoord.

De basiswetten waaruit intelligentie en bewustzijn ontstaan zijn nog door niemand ontdekt. We weten niet eens zeker of je überhaupt iets complex als bewustzijn kunt reduceren tot wiskundige wetten en regels. Vragen als ‘Wanneer heb je bewustzijn gemaakt?’ hebben daarom alleen filosofische antwoorden, en soms zelfs dat niet.

Dat wil niet zeggen dat concretere antwoorden nooit mogelijk zijn.

Onderzoekers uit verschillende disciplines zijn daar namelijk wel degelijk naar op zoek, te beginnen bij neurowetenschappers die stukje bij beetje doorgronden wat zich nu eigenlijk in het menselijk brein afspeelt. Dankzij een reeks fascinerende onderzoeken en resultaten ontsluiëren ze steeds meer facetten van ons eigen besturingsstelsel. Dat is bijzonder, want jarenlang heerste in de psychologie juist de overtuiging dat dergelijke vragen volkomen onwetenschappelijk waren. Psychologen mochten alleen iets zeggen over het gedrag van mensen. Het brein en wat daarin omging waren een zogeheten *black box*, iets waar je niet in kon kijken en dus ook niets over wist. Zodra je als psycholoog iets zei over wat er in het hoofd van iemand anders omging, sloegen je collega's je met die stelling om de oren en namen ze je werk niet meer serieus.

Die overtuiging noemen psychologen het behaviorisme. Tegenwoordig heeft dat zijn beste tijd wel gehad. Nu we met behulp van scans het brein doormeten, kijken we ineens daadwerkelijk onder de menselijke motorkap. Alleen, wat al die metingen nu precies betekenen, waar welke breinfuncties verstopt zitten, of lastiger nog: waar ons bewustzijn zich schuilhoudt, weet nog niemand zeker.

Dat plaatst ook het werk van robotonderzoekers meteen in een interessante perspectief. Want hoe bouw je iets na wanneer je niet weet hoe het origineel precies in elkaar steekt? Het antwoord is: niet. Begrijp me niet verkeerd; er zijn wel degelijk onderzoeken waarin experts op het gebied van kunstmatige intelligentie functies uit ons brein nabootsen, of waarin mensen ons brein op computers aansluiten. Maar op een compleet begrip van ons eigen brein willen en kunnen robotonderzoekers niet wachten. Zij bouwen daarom kunstmatig intelligente systemen waarvan ze hopen dat deze 'vanzelf' steeds intelligenter worden, of dat nu lijkt op hoe intelligentie bij mensen werkt of niet.

Het vergezicht blijft daarbij kunstmatig bewustzijn. Het is goed om hier nog even stil te staan bij een mogelijk punt van verwarring. Veel mensen denken bij de term 'kunstmatige intelligentie' aan robots als David of aan personages als Samantha uit *Her*. Aan volwaardige, kunstmatig bewuste wezens dus. Maar in de wereld van het onderzoek is kunstmatige intelligentie veel breder. De vijanden in een computerspel zijn kunstmatig intelligent en de algoritmen in je tomtom zijn dat ook. Beide zijn alleen mijlenver verwijderd van de robotintelligenties

die we in de film zien. Daarom noemen onderzoekers dergelijke robots en kunstmatige persoonlijkheden liever ‘kunstmatig bewust’ dan ‘kunstmatig intelligent’.

Kunstmatig bewustzijn moet in praktische zin voldoen aan twee vereisten. Allereerst moet een kunstmatig bewustzijn je overtuigen dat ‘hij’ iemand is en een inwendige belevingswereld heeft. Ten tweede moet zo’n bewustzijn wanneer je het ‘openschroeft’ daadwerkelijk complex zijn, het liefst zelfs net zo complex als de mens. Dat sterkt ons in de overtuiging dat het kunstmatig bewustzijn geen trucje is en dat er in de computercode, in de digitale informatie die in dat kunstbrein rondzingt, processen plaatsvinden waaruit dat bewustzijn naar boven komt borrelen. Net zoals bij mensen het bewustzijn tevoorschijn komt uit nog grotendeels onbegrepen breinprocessen. Robotonderzoekers die dat willen bereiken, moeten beginnen met de juiste eerste stap.

Good old-fashioned AI

Elke film begint met een eerste frame, een eerste beeld. Een symfonie begint met een eerste noot en een regenbui met een eerste druppel. Op dezelfde manier begint de zoektocht naar kunstmatig bewustzijn met simpele instructies die kunstmatig intelligent gedrag veroorzaken. Een van de oudste manieren waarop onderzoekers aan kunstmatige intelligentie werken heet tegenwoordig *Good old-fashioned AI* (goede, ouderwetse kunstmatige intelligentie), oftewel: *GOFI*.

GOFI kookt complex gedrag in tot simpele regels. Stel dat je een robot wilt bouwen die een kopje koffie kan pakken, dan schrijf je computercode waarmee je de robot leert: *ALS [KOPJE] DAN [PAKKEN]*. Met behulp van dat soort regels kun je allerlei complex gedrag programmeren.

Alleen heeft die aanpak ook zo z’n problemen. Ten eerste moet je op deze manier voor werkelijk alles wat een kunstmatige intelligentie doet regeltjes opstellen. Buiten die regeltjes om kan je robot niets. De echte wereld is alleen vreselijk complex, dus hoe moet je dat in vredesnaam allemaal in regeltjes vangen?

In 1969 bouwden onderzoekers aan het Stanford Research Institute de robot *SHAKY*. Deze bestond uit een computertje op wieltjes met

een camera erbovenop. Hij bekeek een kamer, de computer analyseerde wat hij zag en vervolgens zocht hij zijn weg langs eventuele obstakels. Het veroorzaakte laaiend enthousiaste reacties in de pers en mensen vroegen zich hardop af hoe lang het nog zou duren voordat robots mensen voorbij zouden streven.

Maar al snel merkten onderzoekers dat de GOFI-instructies die aan het gedrag van SHAKEY ten grondslag lagen in de praktijk helemaal niet zo handig waren. Robots op basis van GOFI bleken onhandige sukeltjes die er uren over doen om door speciaal geprepareerde kamers te bewegen. Zodra je in een kamer een voorwerp plaatste dat niet van tevoren in regels was gedefinieerd, herkende zo'n robot het niet eens. Mensen gaat dat juist heel gemakkelijk af, vermoedelijk doordat ons brein op de achtergrond miljarden en miljarden berekeningen uitvoert, waar we ons gelukkig nooit bewust van zijn. GOFI kon die menselijke brainpower niet bijbenen.

Toch boekte GOFI ook successen. In 1997 won een kunstmatige intelligentie met de naam Deep Blue een schaakwedstrijd van de menselijke wereldkampioen Garri Kasparov. Deep Blue was grotendeels gebaseerd op GOFI-achtige regels, gekoppeld aan patroonherkenning en een enorme database met oude schaakpartijen. Op die manier kon dit kunstbrein de beste menselijke schaker ter wereld verslaan.

De crux daarbij is dat Deep Blue de wereld vereenvoudigt. Deep Blue hield geen rekening met dingen die voor de schaakpartij volstrekt geen belang hadden. Hij bewoog niet, mat niet de temperatuur in de kamer, wist niet hoe laat het was, of het buiten zonnig was en of de koffie in de schaakhal een beetje te pruimen was. Hij praat niet met je en zingt ook niet zomaar ineens een liedje. Deep Blue deed dus één ding heel goed, maar daarbij ontbrak het hem wel aan menselijke intuïtie. Latere schaakpartijen toonden namelijk aan dat teams van schakers en schaakcomputers die simpeler waren dan Deep Blue samen toch beter presteerden dan alleen de supercomputer of de schaakgrootmeester. Dat kwam doordat de snelle robotberekeningen veel beter uitpakten wanneer een mens op basis van intuïtie en strategische inschatting van wat de tegenstander ging doen af en toe de computerbeslissing naast zich neerlegde.

GOFI-achtige regels zijn tegenwoordig nog altijd een onderdeel van

veel kunstmatig intelligente systemen, maar dan eerder in combinatie met andere typen aanpakken. Een daarvan is zo'n beetje het tegenovergestelde van GOFI. Waar GOFI begint met de complete wereld en die vervolgens in simpele instructies vangt – *top-down* noemen informati-ci dat – begint die andere aanpak met vrijwel niets en bouwt zo de rest op. *Bottom-up*, met andere woorden.

Bottom-up aanpak

Simpele insecten bewegen met hersenen met slechts een paar honderdduizend neuronen al door de wereld en overleven. Wie robots bouwt op basis van GOFI heeft voor diezelfde taak enorm complexe kunstbreinen nodig, waar je van tevoren vele duizenden gedragsregels in moet stoppen en waar je een processor voor nodig hebt die vele malen sterker is dan het brein van dat insect. Dat was volgens sommige experts een teken dat GOFI niet de juiste weg bewandelde.

In plaats daarvan gingen deze onderzoekers simpele robots bouwen die handelden op basis van een soort rudimentair instinct. Ze koppelde sensoren direct aan de bewegingsmotoren. Met behulp van een lichtsensoren bouw je bijvoorbeeld een robot die altijd richting het licht beweegt. Of met een botsingssensoren een robot die altijd links afslaat als hij een obstakel tegenkomt. De achterliggende gedachte was dat dit soort heel simpele instructies wel wat weg heeft van de instincten en onbewuste drives van mensen: honger, lust, liefde enzovoort. Dergelijke processen spelen zich af op een onbewust niveau, net als de instructies van deze robots.

Het probleem met een dergelijke aanpak is alleen dat in de achterliggende code de handelingen van een robot niet worden gekoppeld aan concepten. Als je een robot bouwt die altijd naar links beweegt, doordat bijvoorbeeld zijn wieltes die kant op draaien, dan begrijpt die nog niet per se wat 'links' eigenlijk is. Dat is anders bij GOFI. Wanneer je een regel maakt die stelt ALS [OBSTAKEL] DAN [LINKS], dan weet je zowel wat een 'obstakel' is als wat 'links' is.

De hoop was dat dergelijke bottom-up robots vanzelf complexer gedrag zouden vertonen. Dat moest komen opborrelen uit alle simpelere,