

Artikel

WAAROM DE MENS GEEN ALGORITME IS. OVER DE BAVIAAN VAN HARARI

Pim Franssen

In zijn boek *Homo Deus*¹ gaat Yuval Noah Harari ervan uit dat organismen biochemische algoritmen zijn. Vanuit dat idee krijgt zijn “geschiedenis van de toekomst” vorm, waarin hij beschrijft hoe de mens steeds meer zal evolueren tot een nieuwe soort, die streeft naar onsterfelijkheid, geluk en goddelijkheid. Deze soort zal zich onder invloed van technische ontwikkelingen steeds minder gelegen laten liggen aan zijn fysieke en geestelijke beperkingen en zich ontwikkelen tot een supermens met vermogens die wij ons niet voor kunnen stellen.

Harari geeft in zijn boek zelf een korte en alledaagse definitie van wat hij verstaat onder een algoritme: “Een algoritme is een methodische reeks stappen die gebruikt kan worden om berekeningen te maken, problemen op te lossen en tot beslissingen te komen.”² Dit artikel gaat echter uit van de volgende, meer precieze en formele definitie, zoals Roger Penrose die hanteert in zijn boek *Shadows of the mind*³: een algoritme is een berekening van een mathematisch geïdealiseerde computer (*Turing machine*). Deze maakt geen fouten, loopt zo lang als nodig is en heeft een ongelimiteerde geheugencapaciteit.

Als organismen inderdaad algoritmen zijn volgens deze definitie, dan betekent dat dat we, als we het algoritme en alle input eenmaal kennen, met (een grote mate van) zekerheid kunnen zeggen wat de uitkomst zal zijn, waarmee we alle gedrag zouden kunnen voorspellen. Dat is de toekomst die Harari voor zich ziet als de huidige lijn van technische ontwikkelingen in Artificiële Intelligentie (AI) en biotechnologie wordt voortgezet: wij leveren door middel van allerlei apparaten die ons gedrag registreren (van zoekmachines op internet tot draagbare apparaten die onze bewegingen, hartslag, temperatuur, enz. meten) steeds meer data aan bij grote bedrijven zoals Google. Deze data worden vervolgens gebruikt om algoritmen te trainen, zodat Google ons uiteindelijk beter kent dan wij onszelf kennen en we terecht komen in een puur door data gereguleerde wereld. De biotechnologie zal zich aan de hand van die data zo ontwikkelen dat wij onszelf kunnen verbeteren: beschadigd of ongewenst DNA kan worden vervangen, geheugencapaciteit en verwerkingssnelheid van het brein kan worden vergroot, ledematen en organen kunnen steeds opnieuw worden vervangen door kunstmatige exemplaren die sterker zijn en efficiënter werken dan de onderdelen van ons huidige lichaam. Langzaam wordt homo sapiens geüpdatet tot een nieuwe, sterkere, slimmere, efficiëntere en misschien zelfs onsterfelijke soort:

homo deus.

Dit is een toekomstbeeld dat veel mensen aanspreekt. De perfecte wereld, waarin we alles kunnen verklaren, waarin al onze problemen zijn opgelost en waarin we bovendien onsterfelijk worden, is nabij. Wie wil dat nou niet? Tegelijkertijd doet die wereld wat *unheimisch* aan. Het lijkt er kil, en het idee dat Google straks alles van ons weet vervult ons ook met het nodige wantrouwen. Maar bovenal, om onsterfelijk te worden moeten we, paradoxaal genoeg, eerst sterven. Als we in een andere soort veranderen verliezen we onszelf. Wij, als soort, sterven uit. De toekomst die Harari schetst is dus aantrekkelijk, maar ook bedreigend. Wij blijken manipuleerbare apparaten, en dankzij Harari weten we hoe het werkt, zonder ons daartegen te kunnen wapenen. Een belofte van genot waarvan je weet dat het slecht voor je is. En er staat iets op het spel, namelijk ons eigen bestaan. En dan zowel de ultieme vervulling daarvan, als de uiteindelijke ondergang. Dit verklaart wellicht het succes van het boek.

Maar heeft Harari gelijk? Klopt zijn uitgangspunt wel? Zijn mensen daadwerkelijk algoritmen? En zo ja, heeft die veronderstelling inderdaad de consequenties die Harari schetst? Of is er ook een andere kijk op de mens en de werkelijkheid mogelijk?

Bij de bespreking van deze vragen komen de volgende tegenwerpingen aan de orde. (1) Bij het gedrag van mensen en dieren spelen ook niet-algoritmische overwegingen een rol. (2) Als er al sprake zou zijn van een algoritme dan zou het niet werken omdat een dergelijk algoritme zowel praktisch onuitvoerbaar als logisch inconsistent zou zijn. (3) Er is geen reden om aan te nemen dat de functionaliteit die mogelijk is met bewustzijn ook mogelijk is zonder bewustzijn.

Vervolgens komt een alternatieve kijk op een algoritmisch mensbeeld aan bod, uitgaande van de fenomenologie van Maurice Merleau-Ponty en de psychoanalyse van Jacques Lacan. Hoewel Lacan in eerste instantie, met zijn opvatting dat het onbewuste gestructureerd is als een taal, nog wel in de buurt lijkt te komen van Harari, blijkt hij bij nader inzien toch op een compleet ander mensbeeld uit te komen. Hij stelt dat juist het onvervulde verlangen naar perfectie, dat waarvan Harari verwacht dat we het achter ons kunnen laten, de *condition humaine* is, en dat elke belofte dat verlangen te bevredigen, een ideologie is.

Harari's voorbeeld van de baviaan

Bij zijn beschrijving van organismen als biochemische algoritmen geeft Harari het voorbeeld van een baviaan die bananen in een boom ziet hangen. De baviaan ziet ook dat er een leeuw in de buurt is. De vraag die Harari stelt is: moet de baviaan zijn leven riskeren voor die bananen?⁴ Volgens Harari gaat het hier om een kwestie van kansberekening: de kans dat de baviaan de bananen ongeschonden te pakken krijgt, tegenover de kans dat de leeuw de baviaan te grazen neemt. De baviaan moet daarvoor een groot aantal factoren tegen elkaar

afwegen: Hoe ver ben ik van de bananen vandaan? Hoe ver is de leeuw? Hoe hard kan ik lopen? Hoe hard kan de leeuw lopen? Is de leeuw wakker, of slaapt hij? Ziet de leeuw er hongerig uit of niet? Hoeveel bananen zijn er? Zijn ze groot of klein? Groen of rijp? Naast deze externe factoren moet de baviaan rekening houden met informatie over zijn eigen lichamelijke gesteldheid. Hoeveel honger heeft hij? Als hij is uitgehongerd is het zinvoller zijn leven te wagen dan als hij net heeft gegeten en de bananen alleen als een extraatje beschouwt.

Nu maakt de baviaan in zijn hoofd natuurlijk niet letterlijk allerlei berekeningen om tot de juiste beslissing te komen, zegt Harari, nee, hij *voelt* honger, hij *voelt* angst voor de leeuw en begint te bibberen als hij de leeuw ziet en hij *voelt* het water in zijn mond lopen als hij de bananen ziet. Hij wordt overspoeld door sensaties, emoties en verlangens. Volgens Harari zijn deze sensaties, emoties en verlangens in feite algoritmen en is de baviaan in wezen een rekenmachine. De uitkomst van de berekening (wel of niet voor de bananen gaan) is uiteindelijk ook weer een gevoel: “de baviaan zal zich plotseling optimistisch gaan voelen, zijn haren gaan overeind staan, zijn spieren spannen zich, zijn borst zet uit, hij haalt diep adem en: ‘Voorwaarts! Ik kan het! Op naar de bananen!’”⁵ Of hij wordt juist overmand door angst en durft niet meer. Of de uitkomst van de berekening ligt ergens daar tussenin en hij blijft besluiteloos staan.

Niet-algoritmische overwegingen

Bij het voorbeeld van de baviaan noemt Harari een aantal overwegingen die een rol spelen bij de beslissing om wel of niet te proberen de bananen te pakken te krijgen. Als die beslissing afhangt van een algoritme, moeten al die overwegingen gekwantificeerd worden om het algoritme uit te voeren. Maar wat als er ook overwegingen zijn die wel een rol spelen bij de beslissing van de baviaan, maar die niet gekwantificeerd kunnen worden? Als dergelijke overwegingen bestaan, kan de uiteindelijke beslissing niet geheel afhankelijk zijn van een algoritme. Harari gaat er simpelweg vanuit dat alle factoren die een rol spelen voor de beslissing van de baviaan op de juiste manier worden meegenomen en gewogen bij de bepaling van die beslissing, maar hij geeft elders in zijn boek zelf aan dat niet alle beslissingen kunnen worden genomen aan de hand van een rationeel-wetenschappelijke afweging. Bij kwesties van morele en sociale aard schiet de wetenschap tekort. Vandaar dat we volgens hem ook altijd behoefte zullen hebben aan religie.

Hoewel Harari ze niet noemt kunnen we ook in het geval van de baviaan dergelijke morele en sociale overwegingen onderscheiden: stel dat we te maken hebben met een vrouwtje en dat zij de bananen niet alleen nodig heeft voor zichzelf, maar ook voor haar kinderen, broertjes en zusjes, moeder en vader. Misschien is haar moeder oud en zwak en heeft zij dringend behoefte aan eten. Haar jongste kind daarentegen is voor overleven afhankelijk van haar melk. Wordt

zij gegrepen door de leeuw dan is het de vraag of haar kind het overleeft. Nu wordt de kwestie een ingewikkelde morele afweging: hoeveel risico is het redden van haar moeder waard ten opzichte van haar kind?

Een andere overweging die wellicht een rol speelt is haar plaats in de groep. Wat zullen de anderen van haar denken als zij onverrichter zake naar huis terugkeert? Als zij nu geen bijdrage levert aan de voedselvoorziening van de groep, wat betekent dat dan op langere termijn voor haar en haar familie? Misschien is het haar gisteren ook al niet gelukt voedsel te bemachtigen en wil ze daarom nu meer risico nemen. Of misschien neemt ze vandaag juist minder risico, omdat ze altijd al zoveel doet voor de groep. Wie zal het zeggen? En hoe kunnen deze overwegingen meegenomen worden in een algoritme?

Hoewel hij het belang van religie benadrukt wanneer het gaat om kwesties van morele en sociale aard, lijkt Harari er in het geval van de baviaan toch een reductionistisch-natuurwetenschappelijk wereldbeeld op na te houden, waarin die religie niet van belang is. Sensaties, gevoelens en emoties zijn uiteindelijk algoritmen, zo stelt hij. Een algoritme dat, in dit geval, gerealiseerd is in de biologische opmaak van de baviaan, al maakt het volgens Harari voor het algoritme niet uit op welke manier het gerealiseerd is. Oftewel, uiteindelijk zouden, volgens die visie, ook de morele en sociale overwegingen van de baviaan terug te voeren moeten zijn op vurende neuronen in het brein, en in laatste instantie op de natuurwetten die gelden voor de allerkleinste anorganische deeltjes.

Naast dat hij zichzelf hier lijkt tegen te spreken is het maar de vraag of het voor een algoritme niet uitmaakt in welk soort materiaal het gerealiseerd wordt. In theorie niet, want een algoritme is een puur functionele beschrijving. Maar in de praktijk maakt het voor de te volgen stappen wel degelijk uit. Sterker, *als* de baviaan al een algoritme zou zijn, dan is het veel voor de hand liggender om te veronderstellen dat dat algoritme juist *alleen maar* gerealiseerd kan worden op de manier waarop het in de praktijk gerealiseerd is. De eigenschappen van materiaal zijn wel degelijk van belang voor de functionaliteit. Probeer maar eens een computer te bouwen van alleen maar appels.⁶ Dit punt komt hieronder nog uitgebreider terug. Maar laten we er voor nu van uitgaan dat het inderdaad mogelijk is om een kunstmatig algoritme te bouwen met dezelfde functionaliteit als een echte baviaan. Dan nog ontstaat er een probleem, namelijk dat het idee van een organisme als biochemisch algoritme vanwege de te verwerken hoeveelheid data niet alleen praktisch onuitvoerbaar is, maar ook logisch inconsistent.

Niet alles is te berekenen

Laten we nog eens nader kijken naar een van de overwegingen die volgens Harari een rol speelt in het geval van de baviaan. Een van die overwegingen is hoeveel

honger de baviaan heeft. Maar hoe kwantificeer je honger op een manier zodat die een rol kan spelen in een algoritme? In het natuurwetenschappelijke wereldbeeld dat Harari voorstaat hangt het hongergevoel af van allerlei interne indicatoren: bloedsuikerwaarden, peristaltische bewegingen van de darmen, inhoud van de maag, etc. Al deze zaken zijn te kwantificeren en leveren uiteindelijk, volgens een nader te bepalen formule, een bepaalde mate van het hongergevoel op dat weer te geven zou zijn met een cijfer (bijvoorbeeld: op een schaal van 1 tot 10 is het hongergevoel van de baviaan, op basis van de gemeten interne waarden op dit moment 7). Maar we moeten nog een stapje verder gaan, een stapje dieper het lichaam in. Want waar wordt de bloedsuikerspiegel door bepaald? Wat is de oorzaak van de peristaltische bewegingen in de darmen? Uiteindelijk komen we dan uit op celniveau en de processen die daar gaande zijn. Maar ook daar kun je vragen: wat veroorzaakt precies het gedrag van die cellen? En dan beland je uiteindelijk op het grondniveau van de natuurwetenschappelijke werkelijkheid, dat bepaald wordt door het gedrag van atomen, Higgsdeeltjes en bosons. Oftewel: wil je aan het hongergevoel van de baviaan een werkelijk correcte waarde meegeven, dan zul je van elk kleinste deeltje waaruit de baviaan bestaat het gedrag moeten kennen en voorspellen. En dan hebben we het alleen nog maar over het hongergevoel. Wil je het gedrag van de baviaan als geheel voorspellen, rekening houdend met *alle* overwegingen, dan moeten we zelfs nog verder kijken dan alleen de deeltjes waaruit de baviaan zelf bestaat. Het gedrag wordt namelijk niet alleen bepaald door inwerkingen van de deeltjes waaruit de baviaan zelf bestaat op elkaar, maar ook door deeltjes die vanuit de omgeving inwerken op die van de baviaan. Door de deeltjes die de lucht uitmaken waarin de baviaan zich beweegt, maar ook bijvoorbeeld, via die lucht, door de deeltjes waaruit de leeuw is opgebouwd. En ook het gedrag van de leeuw is afhankelijk van deeltjes zowel binnen als buiten hemzelf (hiervoor gelden grotendeels dezelfde overwegingen als bij de baviaan). Bovendien wordt het gedrag van de baviaan ook bepaald door, zoals hierboven al geschetst, bijvoorbeeld hoeveel familie zij heeft en hoeveel honger die hebben, en door haar plaats in de sociale hiërarchie in de troep waarvan zij deel uitmaakt. Ook die invloeden zijn in het natuurwetenschappelijke wereldbeeld dat Harari lijkt aan te hangen, uiteindelijk terug te voeren op bewegende nanodeeltjes. Gezamenlijk maken al die op elkaar inwerkende deeltjes de toestand van de wereld op dat moment uit en spelen ze, hoe miniem ook, een rol in de beslissing die de baviaan uiteindelijk neemt (de vraag is overigens ook of je dan nog wel kunt spreken van een ‘beslissing’. Het gaat eerder om een gebeurtenis die al of niet plaatsvindt).

Kortom, wil je het gedrag van de baviaan op dat ene moment werkelijk begrijpen volgens de geldende wetten van de natuurwetenschap dan is een schier oneindige reeks data nodig. Strikt genomen zou je van elk bestaand deeltje in het universum de plaats, snelheid en richting moeten weten, plus de manier waarop al die deeltjes op elkaar inwerken om te bepalen wat er het volgende moment gaat gebeuren. Volgens dit wereldbeeld is het universum een soort supercomputer, die

voortdurend zijn eigen toestand berekent en daarmee zichzelf voortbrengt. De toestand van de wereld op elk moment is de uitkomst van een algoritme dat 'wereld' heet.

Hoewel het gedrag van de bavian op deze manier dus *in principe* te berekenen zou zijn, is dat in de praktijk onmogelijk. De hoeveelheid en de precisie van de data zullen nooit groot genoeg zijn om met enige acceptabele mate van accuraatheid het gedrag van een individuele bavian op een specifiek moment te voorspellen. Hoogstens zou je in algemene termen iets kunnen zeggen over het gedrag van bavianen in dergelijke situaties.

Maar dit geldt voor de klassieke natuurwetenschap, volgens welke de fysieke werkelijkheid deterministisch en algoritmisch is. Nu lijkt Harari dat wereldbeeld niet helemaal aan te hangen, want in zijn bespreking van de vrije wil stelt hij dat er op subatomair niveau (hij geeft het voorbeeld van het verval van uranium) ook sprake is van toeval.⁷ Hoewel hij de term niet noemt lijkt hij hierbij te verwijzen naar de kwantummechanica. Daarin is ruimte voor random data, terwijl fysieke toestanden toch nog steeds met een zeer grote accuraatheid berekend kunnen worden. De kwantummechanica is niet deterministisch, maar wel algoritmisch. In dat geval zou er inderdaad, zoals Harari stelt, sprake zijn van een kansberekening.

Toch is het niet aannemelijk dat het gedrag van de bavian bepaald wordt door een algoritme. Met een tweetal voorbeelden, ontleend aan Roger Penrose, is inzichtelijk te maken dat bewust gedrag anders werkt dan een berekening. Neem de vraag of, als Abraham Lincolns linkervoet in Washington is, zijn rechtervoet daar dan (zeer waarschijnlijk) ook is, en het probleem om te achterhalen of er een knoop in een stuk touw zit. Voor mensen is het meteen inzichtelijk dat het zeer ongebruikelijk zou zijn als de linkervoet van Lincoln in Washington zou zijn en de rechter niet, en de vraag of er een knoop in een stuk touw zit is heel simpel op te lossen door even te kijken en te voelen. Weliswaar zijn deze problemen ook door computerprogramma's op te lossen, maar dat vergt behoorlijk gecompliceerde en inefficiënte algoritmen en het werkt alleen omdat degenen die dat algoritmen schrijven *begrijpen* wat het probleem is. Maar dat een probleem op een algoritmische manier opgelost kan worden wil natuurlijk helemaal niet zeggen dat dat ook de manier is waarop mensen dat doen. Sterker nog, op basis van de onvolledigheidsstelling van Gödel is aan te tonen dat wij mensen dingen kunnen inzien die niet algoritmisch bewijsbaar zijn. Wij kunnen dus iets dat computers niet kunnen. Het is bijvoorbeeld onmogelijk om de eigenschappen van natuurlijke getallen (0, 1, 2, 3, etc.) volledig te beschrijven in termen van een algoritme, terwijl deze eigenschappen voor een kind vanzelfsprekend zijn vanaf het moment dat het kan rekenen. Aan de hand van enkele voorbeelden (bijvoorbeeld drie bananen, drie sinaasappels) begrijpt een kind dat het concept 'drie' geabstraheerd kan worden en toepasbaar is op een oneindig aantal andere concepten (drie dagen, drie stoelen, drie mensen, etc.) en deel uitmaakt van een eindeloze opeenvolging van vergelijkbare concepten ('vier', 'vijf', 'zes', etc.).

Voor dat inzicht is dus iets anders nodig dan berekening. Penrose noemt dat ‘awareness’.⁸

Het is niet helemaal duidelijk of Harari nou wel of niet denkt dat er bij de baviaan sprake is van een dergelijke awareness. Enerzijds zegt hij dat de baviaan sensaties, emoties en verlangens *heeft*, hetgeen suggereert dat er wel sprake is van awareness. Anderzijds zegt hij dat die sensaties, emoties en verlangens algoritmen *zijn*. Maar hij stelt ook dat die sensaties, emoties en verlangens *het gevolg zijn* van algoritmen, hetgeen suggereert dat ze toch weer iets anders zijn dan die algoritmen zelf. Vermoedelijk bedoelt hij dat de baviaan wel awareness heeft, maar dan *in de vorm van* algoritmen (de ‘niets-anders-dan-bewering’, zie de paragraaf hieronder). Welnu, dat is een situatie die volgens Penrose onmogelijk is, want zijn punt is dus nou net dat awareness iets anders is dan een algoritme. Het is volgens deze redenering weliswaar niet uitgesloten dat we een algoritme kunnen maken dat een soortgelijk gedrag vertoont als de baviaan, maar het is niet aannemelijk dat de baviaan datzelfde gedrag vertoont op basis van een dergelijk algoritme.

Een ander belangrijk punt is dat het, afgezien van het feit dat het praktisch onmogelijk zou zijn om alle data van het algoritme ‘wereld’ te kennen, logisch onmogelijk is dat zo’n algoritme bestaat. De berekening van een bepaalde toestand is immers altijd een berekening van een toestand zonder die berekening zelf, en dus nooit van alles. De berekening zelf kan in die berekening niet meegenomen worden. Berekening vergt een afstand tot dat wat berekend wordt. Feitelijk is dit een variant van het argument van Markus Gabriel dat er geen perspectief op alles kan bestaan, omdat dat perspectief zelf dan buiten de werkelijkheid zou staan.⁹ Dit betekent dat de wereld, en dus ook de hongerige baviaan, geen algoritme kan zijn.

De niets-anders-dan-bewering

Met het gelijkstellen van organismen aan biochemische algoritmen doet Harari in feite niets anders dan wat de aanhangers van het natuurwetenschappelijk wereldbeeld doen in de discussie aangaande het bestaan en de aard van bewustzijn in de *philosophy of mind*: in plaats van sensaties, emoties en gevoelens te nemen voor wat ze zijn, stelt hij dat ze ‘niets anders zijn dan’ algoritmen. In het wereldbeeld waarvan hijzelf in zijn boek het ontstaan beschrijft is geen plaats meer voor andere dan natuurwetenschappelijke concepten (terwijl hij het notabene zelf ook over het belang van verhalen heeft). In dat wereldbeeld kunnen sensaties, emoties en gevoelens niet als zodanig bestaan, maar moeten ze gereduceerd worden tot iets dat past in het begrippenkader van de bètawetenschappen. Alles wat niet met dat begrippenkader verklaard kan worden bestaat niet.

Het vergelijken van de mens met de meest geavanceerde machine van het

moment past, mede onder invloed van Descartes, in een historische traditie.¹⁰ Waar het vroeger ging om klokken, en daarna om stoommachines, gaat het nu om computers, informatie en algoritmen. Wat vaak vergeten wordt is dat de gemaakte vergelijking in eerste instantie bedoeld is als metafoor. Gaandeweg heeft het succes van die metafoor in de moderne tijd echter geleid tot het denkbeeld dat de mens daadwerkelijk een computer *is*. De reden hiervoor is dat er, zoals Harari schetst, onwaarschijnlijke successen zijn geboekt door de wereld op die manier te benaderen. Maar het is maar de vraag of we daarmee dichter bij een begrip van de wereld zijn gekomen. Het feit dat we aan de hand van big data steeds meer en beter gedrag kunnen voorspellen wil nog niet zeggen dat we snappen hoe en waarom dat gedrag tot stand komt. Elk kind kan voorspellen dat een appel omlaag valt als het die loslaat, maar dat wil nog niet zeggen dat het weet wat zwaartekracht is. Voorspellen is een min of meer accurate beschrijving van gedrag, maar daarmee nog geen ontologische verklaring. We moeten niet vergeten dat beschrijvingen van de wereld modellen zijn om te proberen die wereld beter te begrijpen, maar dat het daarbij gaat om, inderdaad, modellen, nabootsing van de werkelijkheid en dus niet de werkelijkheid zelf (al behoort een model zelf natuurlijk wel ook tot de werkelijkheid, het is alleen, per definitie, niet de werkelijkheid waarvan het een afspiegeling is).

Het reduceren van mentale toestanden (sensaties, emoties en gevoelens) tot materiële toestanden (biochemische algoritmen) lijkt een zwaktebod. Omdat we niet kunnen rekenen met mentale toestanden redeneren we ze kunstmatig weg en beschouwen we ze als een soort *shortcut*, een handige manier van praten over wat er *eigenlijk* aan de hand is. Maar we kunnen de zaak ook omdraaien: omdat we niet snappen wat er eigenlijk aan de hand is zoeken we naar een oplossing die in ons wereldbeeld past. Je zou het bestaan van mentale toestanden en de onmogelijkheid om ze te reduceren tot materiële toestanden net zo goed zou kunnen beschouwen als een bewijs dat het natuurwetenschappelijke wereldbeeld niet klopt.

Harari gaat er heel makkelijk vanuit dat het voor een algoritme niet uitmaakt of het gerealiseerd wordt in organisch of anorganisch materiaal. Maar waarom zou dat zo zijn? Waarom zou het niet zo zijn dat het precies het verschil in materiaal is dat het mogelijk maakt dat sommige algoritmen wel en andere niet uitgevoerd kunnen worden? Zoals gezegd, dat je bepaalde organische processen tot op zekere hoogte kunt beschrijven met behulp van algoritmen wil niet zeggen dat die processen dus ook algoritmen *zijn*.

Hoe kan het dat geen enkel anorganisch ding mentale toestanden of bewustzijn heeft? Het antwoord van technologen is dan vaak: omdat de structuur van het ding (nog) niet complex genoeg is. Als we een computer maar complex genoeg maken en dezelfde functionaliteit geven als een mens, zal hij ook het bewustzijn hebben van een mens. Of in ieder geval is dat het dan niet meer vanaf de buitenkant uit te maken of het zo is (zoals dat strikt genomen ook voor andere mensen geldt: ook bij andere mensen kan ik alleen maar uit het gedrag afleiden

dat ze bewustzijn hebben. Bewijzen kan ik het nooit).

Maar is de meer logische en simpele verklaring niet gewoon dat je organisch materiaal nodig hebt om die specifieke complexiteit te verkrijgen? Als bewustzijn blijkbaar alleen een eigenschap is van organismen, wat is er dan logischer om te veronderstellen dat je een organisme nodig hebt om bewustzijn te hebben? Waarom zou bewustzijn ook een eigenschap moeten (kunnen) zijn van anorganisch materiaal?

Harari stapt bij zijn betoog wel heel gemakkelijk heen over allerlei technische problemen die een rol spelen bij het ontstaan van homo deus. Hij gaat ervan uit dat we die wel zullen oplossen. Ook dit is een klassieke valkuil. Er wordt bij dit soort technologische problemen altijd verwezen naar de successen uit het verleden en die worden geprojecteerd op de toekomst: kijk eens wat we in het verleden in een korte tijd voor elkaar gekregen hebben. En dus zal dit probleem ook wel opgelost worden. Maar je kunt de redenering net zo goed omdraaien: in al die tijd dat we proberen om het bewustzijnsprobleem op te lossen en ondanks alle successen van de laatste decennia in de AI en biotechnologie, zijn we geen stap dichterbij gekomen. Ook Harari geeft toe dat we van bewustzijn in feite zo goed als niks weten.¹¹ David Chalmers noemt het daarom ‘*the hard problem*’.¹² Op basis van de ontwikkelingen in het verleden is er dus juist geen enkele reden om aan te nemen dat we op deze manier in de toekomst wel meer te weten gaan komen over bewustzijn.

Zoals gebruikelijk in het natuurwetenschappelijke wereldbeeld brengt Harari een scheiding aan tussen intelligentie en bewustzijn. Hij stelt dat alle functionaliteit die organismen eigen is ook mogelijk is zonder bewustzijn. Hoewel hij dus constateert dat we eigenlijk niks weten van bewustzijn, meent hij wel te kunnen beweren dat het voor cognitieve, intelligente processen geen functie heeft. En dat terwijl alle organismen die dat soort functionaliteit hebben, óók bewustzijn hebben. Het lijkt misschien alsof kunstmatig intelligente systemen dezelfde functionaliteit hebben, maar dat is maar ten dele waar. Dergelijke systemen zijn intelligent op een specifiek gebied. Geen enkele van die systemen vertoont *overall* intelligent gedrag. Je kunt de menselijke wereldkampioen GO, de Koreaan Lee Sedol, bijvoorbeeld ook vragen waar de dichtstbijzijnde wc is, of waar de zon is als die niet schijnt. Of om je het kopje thee aan te reiken dat aan de andere kant van de tafel staat. Hij zal je zonder enige moeite kunnen helpen. Vraag hetzelfde aan AlphaGo (het computerprogramma dat Lee Sedol in 2016 voor het eerst versloeg) en je zult geen antwoord krijgen. Zou het niet veel logischer zijn te veronderstellen dat voor het complexe gedrag dat organismen vertonen nu net wel een vorm van bewustzijn nodig is? Juist de mate van complexiteit die het vereist om zonder problemen door de wereld te bewegen en (cognitieve) problemen op te lossen wordt alleen maar vertoond door organismen met (een bepaalde mate van) bewustzijn. Waarom zou het niet zo kunnen zijn dat bewustzijn nou net de reden is dat wij intelligente systemen kunnen maken die een spelletje GO kunnen winnen?

Organismen als belichaamde wezens

Het wereldbeeld waarin het logisch lijkt om organismen als algoritmen te beschouwen is in de hand gewerkt door het cartesiaans dualisme. Volgens Descartes is het lichaam een machine, bestuurd door een geest. Dat het lichaam in deze visie gemakkelijk gezien kan worden als een algoritme is evident. Minder vanzelfsprekend is dat een dergelijke visie ook de vergelijking van de geest met een algoritme mogelijk heeft gemaakt. Bij Descartes was de geest nog een geheel andere substantie dan het lichaam, maar onder invloed van de Franse arts en filosoof Julien Offray de La Mettrie, die, voortbordurend op Descartes, de mens als geheel gelijkstelde met een machine, en als gevolg van de opkomst en het succes van de natuurwetenschappen, is dat idee langzaam uit beeld verdwenen (want volgens de natuurwetenschappen bestaat er niets anders dan materie). Het achterliggende idee van een centrale instantie die het lichaam bestuurt is echter in stand gebleven. Alleen is deze instantie nu niet meer een geest, maar het brein. En wat kan dit brein anders zijn dan een, weliswaar zeer ingewikkelde, maar toch, machine? En dan is het een kleine stap om dat brein als machine gelijk te stellen aan een algoritme, dat via het lichaam informatie binnenkrijgt, die verwerkt, en een output geeft. Maar zoals we hierboven hebben gezien, is het idee van een organisme als biochemisch algoritme niet houdbaar.

Wat Harari lijkt te vergeten als hij organismen gelijkstelt met algoritmen is dat organismen belichaamde wezens zijn. Het feit dat een organisme een lichaam is, is essentieel voor wat het betekent om een organisme te zijn. Omdat volgens het natuurwetenschappelijke wereldbeeld de werkelijkheid bestaat uit anorganisch materiaal, moet het organisme ook uit anorganisch materiaal bestaan. En hoewel lichamen uiteindelijk ook zijn opgebouwd uit anorganisch materiaal (laten we, in ieder geval voorlopig, tot zover meegaan met het natuurwetenschappelijke wereldbeeld) is er met lichamen iets anders aan de hand dan bijvoorbeeld met stenen. In lichamen heeft bepaald anorganisch materiaal een zodanige vorm aangenomen dat het bepaalde eigenschappen krijgt. Er is *leven* in gekomen, het is *organisch* materiaal geworden. En hoewel je, net als bij bewustzijn, een discussie kunt voeren over de mate van complexiteit waarbij dat leven precies begint, kun je stellen dat organismen over het algemeen bestaan uit levend, organisch materiaal.

Een van de eigenschappen van dat materiaal is dat het onderscheid kan maken tussen zelf en wereld. Wij weten, uitzonderingen met fantoomledematen en dergelijke daargelaten, waar ons lichaam ophoudt en de wereld begint. Daar waar ons lichaam ophoudt is niet meer 'ik'. Waar ik normaal gesproken vrijwillig allerlei zaken (armen, benen, hoofd, etc.) kan bewegen in de gewenste richting, op een manier die ik zelf kies, geldt dat niet voorbij een bepaald punt, namelijk daar waar mijn lichaam ophoudt en de wereld begint. Vanaf dat punt heb ik geen volledige controle meer (ook over je eigen lichaam heb je overigens geen volledige controle, denk aan allerlei automatische processen zoals ademhaling,

hartslag, de werking van de organen. Desondanks weet je dat al die processen *van jou* zijn op een manier waarop andere dingen in de wereld niet van jou zijn). Hoewel het onderscheid tussen zelf en wereld een afstand suggereert van onszelf tot die wereld is er bij het kennen van die wereld geen sprake van een cognitieve, intelligente act van een abstracte geest zoals die in een cartesiaanse visie wordt verondersteld. Juist omdat ons lichaam uiteindelijk uit hetzelfde materiaal bestaat als zijn omgeving, staan wij in direct contact met de (anorganische) wereld. We maken er deel van uit. Maar anders dan bijvoorbeeld een steen zijn wij niet zomaar een object tussen de objecten. Omdat wij organische wezens, organismen zijn, zijn wij naast een object, tegelijkertijd óók een subject. Wij nemen de wereld waar. In plaats van een cartesiaans subject dat als een soort drone los en onafhankelijk boven en door de wereld zweeft en de binnengekomen informatie verwerkt en interpreteert en vervolgens een output geeft, zijn wij, om met Merleau-Ponty te spreken, een *être-au-monde*, een ter-wereld-zijn.¹³ Ons kennen is een zich voortdurend voltrekkende gebeurtenis van interacties tussen lichaam en wereld. Merleau-Ponty spreekt van een prereflexief bewustzijn; een waarnemen van de wereld voorafgaand aan ons denken over die wereld.

Dat is een fundamenteel andere manier om naar organismen en de wereld te kijken dan de manier waarop Harari dat doet. De kracht van organismen is nu juist dat zij *niet* hoeven te rekenen om te weten wat ze moeten doen. Organismen ‘voelen’ waar ze zijn, hoe ze zijn, staan in rechtstreeks contact met de wereld. Dat is een onbemiddeld kennen. Een kennen dat tot stand is gekomen door te bewegen, door het langzaam leren kennen van het lichaam en wat dat met de wereld doet, en het leren kennen hoe de wereld invloed heeft op wat het lichaam doet. Het kennen van een organisme is een doorleefde ervaring. Om met Thomas Nagel te spreken: het voelt als iets om ons te zijn.¹⁴ Terwijl het niet als iets voelt om een algoritme te zijn.

Neem, bij wijze van voorbeeld, een sporter. In zijn bewegingen is hij één met zijn omgeving. Schaatser Sven Kramer is één met het ijs, voetballer Lionel Messi is één met de bal. Zij zijn tijdens het beoefenen van hun sport niet bezig met allerlei razendsnelle berekeningen, met hoe ze hun voet moeten neerzetten, hoe hun houding is, etc. Tijdens de training letten ze daar weliswaar bewust op, maar het doel van die training is nu juist om die berekeningen niet meer te hoeven maken, om alle bewegingen zo in te slijten dat het lichaam ze vanzelf maakt, zonder tussenkomst van een berekenend ik/zelf. Het ideaal is dat dat zelf in de actie verdwijnt. Tijdens de actie wordt niet meer nagedacht, er wordt gedaan. Zo gauw er teveel wordt nagedacht gaat het mis, dan is de vanzelfsprekendheid weg; Kramer maakt een misslag, Messi verliest de controle over de bal. Staat in het cartesiaanse/natuurwetenschappelijke wereldbeeld het bewustzijn tegenover de wereld, als iets dat zich bewust is *van* die wereld, in de fenomenologie van Merleau-Ponty staat het bewustzijn *middenin* de wereld, maakt het er deel van uit, neemt het er deel aan. Het is een uitstaan naar de wereld, belichaamd, en niet-reflexief.

Ook in de psychoanalyse van Lacan speelt het opgaan in de beweging van het lichaam een belangrijke rol. Voor Lacan is de mens een fundamenteel verlangend wezen. Dat verlangen behelst het opheffen van de vervreemding van de Reële Orde die in de kindertijd is ontstaan: vanaf een bepaald moment, in het spiegelstadium, ervaart het kind voor het eerst een discrepantie tussen wat het lichamelijk ervaart en wat het in de spiegel ziet. Zowel het innerlijke ervaringsbeeld als het zintuiglijk waargenomen beeld is het (kind) zelf, maar deze kunnen onmogelijk volledig samenvallen; de sensaties, emoties en gevoelens die opborrelen in zijn lichaam kan hij niet van buitenaf, via de spiegel waarnemen, net zoals anderen dat niet kunnen. Vanaf dat moment wordt het kind opgenomen in de Symbolische Orde van de taal, waarin die sensaties, emoties en gevoelens een talig label krijgen dat nooit volledig samenvalt met die sensaties, emoties en gevoelens zelf. Hoe immers te beschrijven hoe je je voelt? Hoe de geur van een appel te beschrijven? Omdat er geen accurate manier is om te beschrijven wat er omgaat in zijn lichaam ervaart het kind een vervreemding van zichzelf, die het wil opheffen. Het wil terug naar de Reële Orde van het lichaam. Die terugkeer is, vanwege de beperking van de taal, echter fundamenteel onmogelijk. De enige manier om de discrepantie, bij momenten, op te heffen is in het genot (de *jouissance*, in termen van Lacan).¹⁵ Die momenten van genot zijn precies momenten van zelfverlies, momenten waarin het bewuste zelf opgaat in de beweging van het lichaam, zoals bij Kramer en Messi in volle actie. Dat zijn momenten waarin juist geen sprake is van berekening.

Je zou de opvatting van Harari nog kunnen verenigen met de psychoanalyse door het onbewuste te beschouwen als een met het blote oog onzichtbare onderlaag van de werkelijkheid, analoog aan het genoom in de biologie en het kwantumniveau in de natuurkunde, waarin fenomenen alleen aantoonbaar zijn met behulp van speciale, hoogtechnologische experimenten.¹⁶ Desondanks is die onderlaag bepalend voor wat er ‘aan de oppervlakte’ gebeurt – uiterlijke kenmerken en gedrag van organismen in de biologie, direct observeerbare wetten in de klassieke fysica. In de psychologie zouden bewuste sensaties, emoties en verlangens dan epifenomenen zijn van onbewuste processen die zich afspelen op een dieper liggend fysisch niveau (in feite de niets-anders-dan-bewering zoals hierboven geschetst). Dit lijkt in eerste instantie aan te sluiten bij de opvatting van Freud, die, opgeleid als arts, ook zelf op zoek was naar een biologische grondslag voor het onbewuste. Lacan interpreteert Freud echter op een andere manier. Ook Lacan maakt wel eens vergelijkingen tussen het psychisch onbewuste en somatische processen op celniveau, waar het immers ook gaat om het uitwisselen van signalen en informatie. Hier is sprake van een afstemming tussen het organisme en de omgeving, gebaseerd op de evolutionair bepaalde genetische opmaak, die ervoor zorgt dat het organisme kan overleven (het instinct, de driften, het id bij Freud). Dat is het niveau waarop dieren functioneren, en in die zin zou Lacan het voor wat betreft de baviaan misschien nog wel eens kunnen zijn met Harari. Maar in tegenstelling tot dieren wordt de mens vanaf zijn geboorte

opgenomen in de Symbolische Orde van de taal, met als gevolg dat er een ‘dispariteit’ ontstaat tussen wat we zoeken in de wereld en wat we vinden. Het verlangen om die dispariteit ongedaan te maken, het psychisch onbewuste, is volgens Lacan gestructureerd als een taal, en het spreekt tot ons via dromen en versprekingen. Maar juist omdat dat onbewuste een taal is, is het niet in staat de werkelijkheid van de Reële Orde te vatten. Het is die discrepantie tussen de Reële Orde en de Symbolische Orde die Harari (en het cartesiaanse /natuurwetenschappelijke wereldbeeld) parten speelt. Het is die discrepantie ook die de oorzaak is van het *unheimische* gevoel waarvan sprake was in de inleiding. Het algoritme als symbolische beschrijving van de werkelijkheid schiet tekort. En dat is niet omdat het algoritme (nog) niet goed genoeg is, dat is omdat *geen enkele* symbolische beschrijving daartoe in staat is.

Harari merkt heel terecht op dat zelfs winnaars van de Nobelprijs voor de Economie maar een fractie van hun berekeningen doen met pen, papier en rekenmachine: “99% van onze berekeningen – waaronder onze belangrijkste levenskeuzes op het gebied van partner, carrière en woonplaats – wordt gemaakt door de uiterst verfijnde algoritmen die we sensaties, emoties en verlangens noemen,” stelt hij.¹⁷ Hij heeft helemaal gelijk dat die sensaties, emoties en verlangens die levenskeuzes bepalen. Het zijn alleen geen algoritmen.

Noten:

1 Harari, Y.N., *Homo Deus. Een kleine geschiedenis van de toekomst*, Amsterdam, Thomas Rap, 2017. Vertaling: Inge Pieters.

2 Harari, Y.N., *Homo Deus*, p. 95.

3 Penrose, R., *Shadows of the mind. A search for the missing science of consciousness*, Oxford, Oxford University Press, 1994.

4 Harari, Y.N., *Homo Deus*, p. 97.

5 Harari, Y.N., *Homo Deus*, p. 97/98.

6 Dit is een variant van het argument dat Ned Block gebruikt in zijn artikel tegen het functionalisme: Block, N., ‘Troubles with functionalism’, in: *Savage, W.* (red.), *Minnesota Studies in the Philosophy of Science* 9 (1978), p. 261-325.

7 Harari, Y.N., *Homo Deus*, p. 292/293.

8 Penrose, R., *Shadows of the mind*, p.54. Penrose geeft in dit boek ook een uitgebreide beschrijving van de onvolledigheidsstelling van Gödel, en een uitleg waarom volgens hem zowel de klassieke natuurkunde als de kwantummechanica in de huidige vorm tekort schieten als het gaat om het verklaren van bewustzijn.

9 Gabriel, M., *Waarom de wereld niet bestaat*. Amsterdam, Boom, 2014. Vertaling: Huub Stegeman.

10 Draaisma, D., ‘De machine achter het gordijn: Drie eeuwen debat over bewustzijn en mechanisme’, in: Hagoort, P. en Maessen, R. (red.), *Geest, Computer, Kunst*, Utrecht, Grafiet, 1986.

11 Harari, Y.N., *Homo Deus*, p. 119.

12 Chalmers, D., ‘Facing up to the problem of consciousness’, in: *Journal of Consciousness Studies* 2 (1995) 3, p. 200-19.

13 Merleau-Ponty, M., *De wereld waarnemen*, Amsterdam, Boom, 2003. Vertaling: Jenny Slatman.

- 14 Nagel, T., 'What is it like to be a bat?', in: *The Philosophical Review* 83 (1974) 4, p. 435–450.
- 15 De Kesel, M., *Het münchhausenparadigma. Waarom Freud en Lacan ertoe doen*, Nijmegen, Vantilt, 2019.
- 16 Zwart, H.A.E., 'The genome as the biological unconscious – and the unconscious as the psychic 'genome': a psychoanalytical rereading of molecular genetics', in: *Cosmos and History: The Journal of Natural and Social Philosophy* 9 (2013) 2, p. 198–222.
- 17 Harari, Y.N., *Homo Deus*, p. 99.