

2 Humanisme, humaniteit en Harari¹

Joachim Duyndam

Het traditionele humanisme staat vandaag voor grote uitdagingen. Ik noem er twee. In de eerste plaats dwingt de huidige ecologische crisis ons om kritisch na te denken over de centrale positie die het traditionele humanisme de mens toekent in het universum. De dominantie van de mens over de fauna en flora van de aarde – misschien is ‘dictatuur’ een passender term – heeft geleid tot uitwassen als de vee-industrie, de kap van oerwouden, de vergiftiging van bijen, vogels en andere dieren ten behoeve van de landbouw, enzovoort. De invloed van menselijke activiteit op de natuurlijke condities van de aarde (bodem, atmosfeer, klimaat) is zo groot geworden dat het geologische tijdvak waarin wij thans leven door velen wordt aangeduid als Antropoceen (Zalasiewicz e.a., 2010). Had het humanisme in het verleden weinig aandacht voor alles wat tot ‘de natuur’ werd gerekend, vandaag kan het er niet meer omheen de *condition humaine* kritisch te herbezielen in het licht van de *condition écologique*. Het humanisme, opgevat als intellectuele en artistieke beweging die de manifestaties van cultuur (wetenschap, politiek, economie, religie, enzovoort) kritisch onderzoekt vanuit de bronnen en uitgangspunten van de cultuur, moet als kritische beweging ook zelfkritisch zijn. Daarbij is humaniteit de inzet, het ‘omwille-waarvan’ van de humanistische kritiek en zelfkritiek. De inhoud en strekking van dat lastige begrip ‘humaniteit’ zijn uiteraard niet in steen gegoten, maar moeten steeds opnieuw en in de context van de tijd worden onderzocht. Precies daarin ligt de zelfkritische functie van het humanisme, dat overigens geen monopolie heeft op humaniteit. Zo doet Harry Kunneman, geïnspireerd door onder meer het werk van Frans de Waal, voorstellen om het bereik van humaniteit uit te breiden naar andere levensvormen op aarde, door Kunneman samengevat met de term ‘bioten’. Kunneman drukt de grondige zelfkritiek van het humanisme, alsook de urgentie daarvan, uit in de term ‘radicaal humanisme’ (Kunneman, 2017).

Een tweede uitdaging wordt ons voorgehouden door de Israëlische historicus Yuval Noah Harari. Over deze uitdaging gaat dit artikel. In zijn imposante bestseller *Homo Deus* betoogt Harari (2017a) dat de mensheid zich thans, na de agrarische revolutie, de

¹ Dit artikel verscheen eerder in: Waardenwerk (2019), 76, 26-35

wetenschappelijke, de technologische en industriële revoluties, midden in een digitale en robottechnologische revolutie bevindt. Hij voorspelt dat computer- en robottechnologieën de mens – of althans de meeste mensen – overbodig zullen maken. Naast de ecologische crisis dwingen ook deze ontwikkelingen ons, en in het bijzonder het humanisme, om kritisch na te denken over de betekenis van humaniteit nu en in de nabije toekomst. Harari's analyses en zijn voorspellingen van de overbodigheid van de mens lijken de door het humanisme verdedigde waarde van humaniteit in het hart te treffen. Ook al laat humaniteit zich moeilijk precies – laat staan volledig – omschrijven, wat in elk geval behoort tot humaniteit, alsook tot de kritische verdediging ervan door het humanisme (volgens onze opvatting van humanisme als cultuurkritiek, inclusief zelfkritiek), is het menselijk oordeelsvermogen: het vermogen om te oordelen over de waarheid, goedheid, schoonheid en geloofwaardigheid van gebeurtenissen, situaties, teksten, opvattingen, enzovoort. Uit het werk van Frans de Waal blijkt dat ook andere dieren – in zekere mate – in staat zijn tot dergelijke oordelen. En volgens Harari kunnen dergelijke oordelen ook, en zelfs veel beter, door computergestuurde robots worden gegeven. Als het waar is dat het vermogen tot kritisch oordelen onderdeel is van humaniteit, zoals ik stel, en als het in het humanisme om humaniteit draait, wat betekent het dan voor het humanisme dat humaniteit niet meer is voorbehouden aan de mens? In dit artikel hoop ik bij te dragen aan de discussie over deze vraag.

Ik wil daarom deze bijdrage over Harari's 'kleine geschiedenis van de toekomst', zoals de ondertitel van *Homo Deus* luidt, beginnen met een casus waarin de vraag naar het menselijk oordeelsvermogen centraal staat. De casus behelst de vraag: Kan een robot de scheidsrechter vervangen bij een voetbalwedstrijd? Vervangen is iets anders dan de scheidsrechter technisch bijstaan, bijvoorbeeld met doellijnteknologie en de zogeheten video-scheidsrechter – hulpmiddelen die in toenemende mate worden toegepast, met als doel menselijke fouten zoveel mogelijk uit te sluiten en zodoende het spel zo eerlijk mogelijk te laten verlopen (Duyndam, 2019). Het is goed voorstelbaar dat een robot (in feite een geavanceerde computer voorzien van een groot aantal camera's) feilloos kan registreren wat er in het spel gebeurt. Maar kan een robot ook situaties *beoordelen* als een scheidsrechter? Kan een robot bijvoorbeeld een buitenspel situatie in een voetbalwedstrijd beoordelen?

Uiteraard kan een robot registreren of een speler 'fysiek buitenspel' staat. Dat is het geval als een speler die wordt aangespeeld door een medespeler slechts één speler van de tegenpartij, bijvoorbeeld de keeper, vóór zich heeft in het veld. Dat is al lastig genoeg, want als de speler in kwestie en de tegenstander(s) zich, op het moment van aangespeeld worden, op één lijn in de breedte van het veld bevinden, is het géén buitenspel. Dit zou een robot echter in principe precies kunnen vaststellen. Maar volgens de spelregels kan een speler

alleen maar buitenspel staan als hij/zij ‘aan het spel deelneemt’. Iedereen die het heeft meegemaakt, herinnert zich vast nog wel de door Feyenoord gewonnen UEFA-cup-finale van 2002 in Rotterdam, enkele dagen na de moord op Pim Fortuyn. Bij een van de Feyenoord-doelpunten stond toenmalig sterspeler Pierre van Hooijdonk fysiek ‘meters buitenspel’, maar omdat hij rustig naar de middellijn terugwandelde terwijl een medespeler met de bal op het Duitse doel afstormde, werd Pierre niet afgefloten, en werd er reglementair gescoord voor Feyenoord. Van Hooijdonk nam op dat moment niet aan het spel deel.

Kan een robot *beoordelen* of een speler wel of niet aan het spel deelneemt? Want dit is een oordeel, geen simpele registratie. In het geval van Van Hooijdonk zou je nog kunnen zeggen: hij bevond zich niet in de buurt van de bal, hij liep overduidelijk ‘de verkeerde kant’ op, hij stelde zich niet aanspeelbaar op, enzovoort. In dit voorbeeld is het relatief gemakkelijk vast te stellen dat Van Hooijdonk niet aan het spel deelnam. Een meer gecompliceerde spelsituatie, echter, betreft een speler die fysiek buitenspel staat maar niet wordt aangespeeld (en dus niet wordt afgefloten), maar die even later tóch buitenspel staat (en alsnog wordt afgefloten) als hij bij toeval, bijvoorbeeld door een door het houtwerk of via de keeper teruggekaatste bal, in het bezit komt van de bal terwijl hij op dát moment fysiek niet meer buitenspel staat. Hij heeft dan toch, en alsnog, ‘aan het spel deelgenomen’, en staat – als het ware met terugwerkende kracht – buitenspel. Kan een robot in zo’n situatie beoordelen of een speler aan het spel deelneemt? In het zojuist gegeven voorbeeld van een speler die eerst niet maar daarna alsnog buitenspel staat omdat hij *in dezelfde spelsituatie* toch aan het spel heeft deelgenomen, hangt het oordeel van buitenspel af van wanneer precies een spelsituatie begint en wanneer die eindigt. Soms is dat duidelijk, bijvoorbeeld bij het fluitsignaal van de scheidsrechter bij een overtreding of een bal over de lijn. Maar op andere momenten gaan spelsituaties vloeiend in elkaar over, en wanneer begint er dan een nieuwe spelsituatie en wanneer eindigt die, anders dan door een fluitsignaal? Kan een robot dat beoordelen?

Men zou kunnen tegenwerpen dat een robot wel degelijk een buitenspelsituatie kan vaststellen, namelijk door de actuele situatie in het veld razendsnel te vergelijken met een zeer groot aantal in een database opgeslagen spelsituaties waarvan het oordeel, wel of geen buitenspel, reeds vooraf is gegeven en mede is opgeslagen. Maar hoe behulpzaam ook, dit is geen *beoordeling* door de computer zelf, in de zin van een toepassing van de buitenspelregel. De vraag blijft: Kan een robot zélf oordelen vellen?

1 Data en algoritmes

In *Homo Deus* – de mens als god – beantwoordt Yuval Noah Harari voorgaande vragen onomwonden positief. Weliswaar komt de gegeven casus niet letterlijk uit zijn

boek, maar hij geeft tal van vergelijkbare voorbeelden. Ja, een robot kan een menselijke scheidsrechter vervangen. Sterker: een robot doet zijn werk veel beter dan een menselijke scheidsrechter dat kan, omdat hij heel veel sneller is, nooit moe wordt, en geen fouten maakt. Over de snelheid en de onvermoeibaarheid kunnen we het snel eens zijn, maar of een robot geen fouten maakt, valt nog te bezien. Ik kom daar nog op terug.

Een kleine geschiedenis van de toekomst (in het Engels: *A Brief History of Tomorrow*), zo luidt enigszins paradoxaal de ondertitel van *Homo Deus*. De ondertitel geeft echter precies aan wat het boek doet: het beschrijft de toekomst alsof deze al is gebeurd, als geschiedenis dus. Niet vanuit een *science fiction*-perspectief, dat vanuit een toekomst terugkijkt op ons heden, maar vanuit het heden waar deze toekomst al aan het gebeuren is. Het heden zoals dat is voortgekomen uit de geschiedenis van de mensheid. De tijdlijn van deze geschiedenis loopt van de *homo sapiens*, waaraan Harari in zijn vorige boek *Sapiens* (2017b) al was begonnen, naar de *homo deus* van de toekomst. Het is de geschiedenis van de mensheid vanaf het moment dat zij zich aan haar louter dierlijke staat heeft onttrokken tot aan de tijd dat mensen als goden (zullen) zijn geworden. Weliswaar een ‘kleine’ of ‘korte’ (*brief*) geschiedenis, maar in zo’n vijfhonderd pagina’s niettemin zeer gedetailleerd verteld aan de hand van concrete gebeurtenissen en personages. Precies hierin is de overtuigingskracht van het boek gelegen. Intussen is, na *Homo Deus*, ook Harari’s *21 lessen voor de 21^{ste} eeuw* (2018) verschenen – het derde boek in een trilogie? – waarin hij zijn bevindingen uit *Sapiens* (over het verleden) en *Homo Deus* (over de toekomst) toespitst op het heden, in de vorm van een agenda van *21 things to do* voor de mensen van nu.

De belangrijkste aanname of vooronderstelling in *Homo Deus* – om maar meteen de plot van het boek prijs te geven – is dat er geen *principieel* verschil is tussen anorganische data-verwerkende systemen zoals computers en robots, en organische systemen zoals mensen en andere dieren. Beide werken volgens ‘algoritmes’, een sleutelbegrip in het werk van Harari. Een algoritme, zegt Harari, is een methodische reeks stappen die gebruikt kan worden om berekeningen te maken, problemen op te lossen en tot beslissingen te komen (2017a, 95). Een algoritme is niet zelf een berekening, oplossing of beslissing, het is de methode die gevolgd wordt bij de uitvoering daarvan. Een eenvoudig algoritme om het gemiddelde tussen twee getallen te berekenen, luidt bijvoorbeeld: (1) tel de twee getallen bij elkaar op; (2) deel de som van beide getallen door twee. Een ander voorbeeld is een kookrecept met daarin de stappen die je achtereenvolgens moet zetten om, bijvoorbeeld, groentesoep te bereiden. Je kunt in het eerste voorbeeld andere getallen nemen of in het tweede voorbeeld andere ingrediënten inzetten; de algoritmes blijven dezelfde.

Het cruciale punt in Harari's betoog is dat algoritmes volgens hem niet alleen machines aansturen – van machines voor cappuccino, auto-assemblage en het maken van chips tot aan geavanceerde computers die het vliegverkeer, ziekenhuisoperaties en het internet besturen – maar dat ook mensen en andere dieren door algoritmes worden aangestuurd. Harari geeft het voorbeeld van een baviaan die in een boom verderop bananen ziet hangen, maar die ook weet dat er een leeuw rondloopt (2017a, 97). De baviaan moet zijn honger afwegen tegen de kans dat hij tot prooi wordt voor de leeuw. Dat vereist de verwerking van heel wat gegevens: de afstand tot de bananen en tot de leeuw; hoe hard kan ik lopen en hoe hard loopt de leeuw; hoe alert of slapend is de leeuw; hoe rijp zijn de bananen; hoe groot is mijn honger? Voor de verwerking van al die gegevens zijn complexe algoritmes nodig. Een angstige baviaan, met algoritmes die de gevaren overschatten, zal verhongeren, zodat zijn genen verloren gaan. Een overmoedige baviaan, met algoritmes die gevaren onderschatten, zal door de leeuw gepakt worden, en daarmee gaan zijn genen verloren. Alleen een baviaan met de juiste algoritmes zal overleven en zijn genen kunnen doorgeven. De algoritmes ondergaan een kwaliteitscontrole door natuurlijke selectie.

Na dit beginvoorbeeld geeft Harari tal van menselijke voorbeelden van sturende algoritmes, zoals bij de verliefde man die een vrouw wil veroveren, een manager die de winst van het bedrijf moet maximaliseren, een fietser in het verkeer. Met deze voorbeelden wil Harari laten zien dat bewustzijn, gedachten, gevoelens, emoties en sensaties – zaken die we gewoonlijk toeschrijven aan een denkend en voelend subject – in feite algoritmes zijn. Gevoelde honger, angst, hoop, optimisme, overmoed – in het voorbeeld van de baviaan – zijn uitkomsten van kansberekeningen, niet door een externe rekenmachine, maar door de baviaan die zelf de rekenmachine 'is'. De baviaan wordt evenzeer door algoritmes gestuurd als de anorganische rekenmachine, alleen zijn bij de baviaan de algoritmes complexer. Nog complexer, maar niet wezenlijk anders, zijn de algoritmes die mensen besturen, bijvoorbeeld voetbalscheidsrechters.

In de mensenwereld zullen door de voortschrijdende technologie de algoritmes steeds verfijnder, perfecter en alomvattender worden, aldus Harari. En zullen onze keuzes steeds meer door algoritmes worden voorbereid, voorgekookt, en uiteindelijk door algoritmes worden gemaakt. Dat geldt niet alleen voor individuele keuzes zoals voor een studie, een baan of een liefdespartner, en de keuzes van organisaties voor bepaalde sollicitanten en de aan- en verkopen op de aandelenbeurzen, maar ook voor beleidskeuzes door de overheid en de keuze van een regering om oorlog te gaan voeren. Sterker: onze keuzes worden altijd al door algoritmes gemaakt omdat wij zelf door algoritmes gestuurde organismen *zijn*. Alleen heeft ons humanistische mensbeeld tot nu toe verhinderd dat we dit inzien. Ons humanistische mensbeeld houdt vast aan het idee van men-

sen als subjecten die *zelf* gedachten, gevoelens en handelingen voortbrengen. Dat ‘zelf’, zegt Harari, bestaat helemaal niet. Het is een moderne humanistische illusie. Hierin vindt hij hersenonderzoekers als Dick Swaab (2010) aan zijn zijde. Ook deze hersenonderzoekers zeggen dat er niet zoiets is als een ‘zelf’ dat keuzes maakt. Onze hersenen maken de keuzes, zij hebben de keuzes al gemaakt voordat wij er een besef van hebben.

2 Humanisme tussen de overbodigheid van God en de overbodigheid van de mens

Onomwonden verklaart Harari: ‘Het humanisme heeft zijn langste tijd gehad’ (2017c). Met humanisme bedoelt hij de opvatting dat de mens *zélf* de regie heeft over zijn/haar bestaan. Dat geldt niet alleen voor de individuele mens, het geldt vooral ook voor mensen-in-samenwerking. Juist het vermogen tot samenwerken heeft de mens verder gebracht dan de andere dieren en tot de dominante diersoort op aarde gemaakt, aldus Harari (2017a, 143-197). Regie betekent kunnen leren van het verleden, plannen kunnen maken voor de toekomst, en een zekere greep hebben op wat er in het heden gebeurt: het organiseren van de wereld met het oog op voedselvoorziening, ziektebestrijding, kennisverwerving, oorlogvoering (en het beperken daarvan), enzovoort. Regie is niet hetzelfde als almacht. We moeten voortdurend *dealen* met beperkingen, tegenslagen en uitdagingen, zowel natuurlijke als technische, maar dat *kún*nen we dan ook, steeds beter zelfs, in ons individuele leven en in de samenleving met elkaar.

Dat is niet altijd zo geweest. Ooit begon de mens zoals de meeste andere dieren: als jager-verzamelaar. Door de agrarische revolutie, waarin de mens bezit nam van een deel van de planten- en dierenwereld, kreeg de mens meer greep op de voedselvoorziening. Terwijl de mens als jager-verzamelaar vele natuurgoden kende, trad met de agrarische revolutie het monotheïsme in, volgens Harari (2017a, 181-182). Toen in de moderne tijd de wetenschappelijke, de industriële en de technologische revoluties zich voltrokken, werd het humanisme de dominante ideologie. In mijn lezing van Harari’s geschiedenis van de toekomst verloopt de overgang van monotheïsme naar humanisme tamelijk soepel. Vanuit het monotheïsme gezien, geeft God als vader zijn volwassen geworden kinderen steeds meer de vrije hand, al past Hij nog wel op de moraal. Vanuit het humanisme gezien, heeft de mens God niet meer nodig; wat men vroeger moest afsmecken, kan de mens nu zelf bewerkstelligen; en de moraal kan de mens zelf redelijk inzien (volgens Kant en de continentale Verlichting) dan wel haarfijn aanvoelen (volgens de Schotse Verlichting).

Met de kunstmatige intelligentie treedt echter een nieuwe technologische revolutie in werking. Niet langer is de mens het enige intelligente wezen in het universum, waar-

mee ook de alleenheerschappij van de mens ten einde komt. Robots zullen op korte termijn qua intelligentie veel meer, veel sneller, kunnen dan de mens. Maar daarmee is niet de mens ten einde. Dankzij kunstmatige intelligentie kan de mens zichzelf verbeteren: extra geheugen implanteren, beperkte ledematen vervangen, nano-sensoren inbrengen die ziektes in een vroeg stadium kunnen detecteren, enzovoort. Zo zal de *homo sapiens* langzamerhand plaatsmaken voor de *homo deus*, de mens als een god. Harari voorspelt echter dat dergelijke verbeteringen niet voor iedereen zijn weggelegd, maar alleen voor een kleine klasse rijken die zich dat kan veroorloven. De meeste mensen zullen deze overgang van *homo sapiens* naar *homo deus* gaan missen; zij zullen als arbeidskrachten, als soldaten en als professionals (artsen, leraren, kunstenaars, managers, advocaten, rechters, schrijvers) vervangen worden door robots, en dus als mens overbodig worden.

Is hiermee ook het humanisme ten einde, en vervangen door zoiets als een posthumanisme? Centraal in Harari's opvatting van humanisme staat het 'zelf' (2017a, 339v). Dat is, het zij nogmaals gezegd, niet alleen het individuele zelf, maar juist ook het zelf van de samenwerkende mens. In mijn interpretatie gaat het hier over regie. Regie betekent geen alleenheerschappij of almacht, maar wel 'het heft in handen' in het zoeken en vinden van wetenschappelijke, technische en ideologische antwoorden op (evenzeer wetenschappelijke, technische en ideologische) beperkingen, tegenslagen en uitdagingen. Regie heeft dus een 'antwoordelijk' karakter, en wordt in de moralen van de meeste ideologieën dan ook verbonden met ver-antwoordelijkheid. Zo ook in het humanisme. Regie is te vergelijken met de stuurmanpositie op een zeilboot. De stuurman is afhankelijk van wind en water, maar rekening houdend met de kracht en de richting daarvan kan hij de boot in de gewenste richting laveren en manoeuvreren. Als de stuurman een jager-verzamelaar (visser) is, is hij afhankelijk van de zon-, zee- en windgoden; als monotheïst vertrouwt hij op de zegen van zijn Bijbelse God als superregisseur; en als moderne humanist heeft hij zelf het heft in handen genomen, en vertrouwt hij op zijn rede en gevoel. Overigens heeft onze stuurman vandaag, als postmoderne *multiple-belonger*, waarschijnlijk zelf een hoogsteigen authentieke mix gemaakt van deze ideologieën.

Volgens Harari is dit sturende en regisserende 'zelf' niet een zelfstandige (denkende, voelende) instantie maar een algoritme, zij het een complex algoritme, of althans de uitkomst van complexe algoritmes. Daarmee is het 'zelf' nog steeds 'antwoordelijk', in die zin dat de menselijke algoritmes en de uitkomsten ervan antwoorden vormen op 'stimuli' – van honger en dorst tot aan *rocket science*. Maar kan dit 'zelf' als (uitkomst van) algoritmes ook verantwoordelijk zijn? Dat hangt ervan af wat met verantwoordelijkheid wordt bedoeld. De meeste ethische theorieën maken in dit verband onderscheid tussen oorzaak en reden. Als je zegt 'de storm is verantwoordelijk voor

de dijkdoorbraak' doel je op een oorzaak, maar als iemand verantwoordelijk wordt gehouden voor een weldaad of een misdaad, beoordeel je iemands redenen, en veronderstel je dat hij/zij ook anders had kunnen handelen, dat diegene de weldaad of de misdaad ook niet had kunnen plegen. Hoe zou je anders iemand kunnen prijzen (om de weldaad) of laken (om de misdaad)? Verantwoordelijkheid in deze morele zin is te omschrijven als het belast zijn met een taak tot zorg, op de vervulling waarvan je kunt worden aangesproken (Tongeren, 1986). Welnu, dit aangesproken kunnen worden, veronderstelt dat je ook anders had kunnen handelen, dat je je taak tot zorg wél had moeten uitvoeren (als je dit had nagelaten), dan wel dat je deze anders of beter had moeten uitvoeren.

Het is voorstelbaar dat ook robots kunnen worden 'aangesproken' op het beter uitvoeren van hun taken. Beter lijkt dan te betekenen: functioneel beter. Algoritmes kunnen daartoe worden herschreven. Het is zelfs voorstelbaar dat 'zelflerende' robots dit 'zelf' kunnen, via superalgoritmes die algoritmes herschrijven. Maar kunnen robots ook van hun algoritmes afwijken? Zijn er, met andere woorden, *onverwachte* uitkomsten van algoritmes mogelijk, dat wil zeggen: uitkomsten die ook de robot zelf niet voor mogelijk had gehouden?

3 De libero arbitrio: de vrije wil en de vrije scheidsrechter

Het lijkt erop dat hier een oeroude discussie wordt hernomen, namelijk de discussie over de vrije wil. Zowel Harari als de hersenonderzoekers zoals Swaab zijn hier allang klaar mee: de vrije wil bestaat niet, de vrije wil is een illusie. Voor Swaab en de zijnen zijn denken, voelen en handelen hersenprocessen (Lamme, 2011; Swaab, 2010, 379v); voor Harari zijn het algoritmes (2017a, 348v). Er is voor hen geen zelfstandige wil, laat staan een vrije wil. In de zestiende eeuw, in de overgang naar de moderne tijd, voerden Erasmus (1469-1536) en Luther (1483-1546) dit debat in religieuze termen (Bell, 2005). Terwijl Luther in zijn *De arbitrio servo (Over de onvrije wil)*, in de context van de genade van God, de onvrije of gebonden wil verdedigt als door God in Christus geopenbaard, verdedigt Erasmus in *De libero arbitrio (Over de vrije wil)* de vrije wil als 'de kracht waarmee de mens zich tot datgene wat tot eeuwig heil voert, kan toewenden of afwenden' (Bell, 2005, 32-38). In de achttiende-eeuwse Verlichting laat Kant in zijn magistrale ethiek zien dat de vrije wil niet kan worden bewezen maar wel moet worden verondersteld, wil ethiek überhaupt mogelijk zijn. Concepten als autonomie, morele verantwoordelijkheid en morele plicht worden zinloos zonder veronderstelde vrije wil, dat wil zeggen zonder de mogelijkheid van plichtsverzaking en aangesproken kunnen worden op verantwoordelijkheid (Kant, 2008, 129v).

Ook in deze tijd kan de vrije wil, tegen alle wetenschappelijke weerlegging van Harari en Swaab c.s. in, niet worden bewezen, maar wel worden verdedigd. Om te beginnen moet de absolute tegenstelling tussen de vrijheid en onvrijheid van de wil worden gereduceerd. Niemand zal ontkennen dat onze wil wordt beïnvloed door sturende factoren vanuit ons lichaam, ons brein, de maatschappij, de cultuur, en niet in de laatste plaats door de wil van anderen. De mimetische theorie van René Girard (2008) toont de besmettelijke invloed van de interesse of begeerte van anderen op de kracht en richting van onze wil. Uiteraard is de wil niet absoluut vrij. Maar misschien kan de wil wel vrij zijn als responsief, (ver)antwoordelijk, relatief vrij antwoord op alle heteronome factoren die de wil beïnvloeden en bepalen – zo heb ik elders verdedigd (Duyndam, 2017; 2011). Doet Harari dat trouwens niet zelf ook (2017a, 356), als hij als motief opgeeft voor zijn geschiedschrijving van de toekomst en voor zijn voorspellingen over het overbodig worden van de meeste mensen dat we moeten weten wat er aan het gebeuren is en wat ons te wachten staat, zodat we er tijdig nog iets aan kunnen doen? Het is vooral de onwetendheid die hij als historicus van de toekomst bestrijdt, de onwetendheid die volgens hem het gevolg is van het modern-humanistische mensbeeld van zelfbeschikking. Ons modern-humanistische zelfbeeld versluiert ons zicht op de toekomst, die al aan het gebeuren is zonder wij dit ten volle beseffen.

Zelf denk ik ook dat de vrije wil mogelijk is en verdedigd moet worden, niet als vrijheid die losstaat (ab-soluut is) van wat de wil beïnvloedt, uitdaagt of tracht te onderwerpen, maar als relatief of relationeel veerkrachtig en weerbaar *antwoord* op de druk, tegenslag en kwetsbaarheid waaraan de wil blootstaat. Ik zou precies dit humanistisch noemen: de relatief vrije mogelijkheid van een eigen antwoord. De volgende ervaringsargumenten kunnen deze visie – de relatief vrije mogelijkheid van een eigen antwoord – ondersteunen. Ik formuleer deze argumenten in algemene ik-vorm van de ervaring, omdat zij de algemene vrije mogelijkheid van een *eigen* antwoord verdedigen.

- 1 Ik ben in staat tot geduld. Terwijl robots steeds sneller en zo efficiënt mogelijk te werk gaan, kan ik aarzelen, talmen, terughouden, verwijlen, wachten. Dankzij dit geduld ben ik in staat tot zelfreflectie. Ik kan in een dagboek aan zelfonderzoek doen. Ik kan door het lezen van literatuur en poëzie, en door het opgaan in theater, speelfilms en documentaires, mijn leven en wat daarin gebeurt in het licht van andere levens en grotere verhalen zien, begrijpen en duiden.
- 2 Niet alleen ben ik in staat tot zelfreflectie, ik word mijzelf ook 'irreflectief' gewaar (Sartre, 1987) in mijn emoties. Naast een cognitieve kant – de communicatieve (berekenbare, bediscussieerbare) relatie met de aanleiding of reden voor een emotie – heeft een emotie ook een zelfgewartwordingsaspect. Als ik bang ben voor iets of iemand, of boos ben op de buurman, of verdrietig, of opgelucht, of blij ben over iets, *voel ik mijzelf* als angstig, boos, verdrietig, opgelucht, en blij. Dit voelen van

mijzelf is geen terugblikkende reflectie, maar een gelijktijdige zelfgewaarwording – gelijktijdig met het optreden van de aanleiding of reden van de emotie (Duyndam, 1997).

- 3 Ook in mijn gedachten kan ik mij van mijzelf bewust zijn: terwijl ik over iets denk, weet ik dat ik daarover denk.
- 4 Dankzij dit zelfgewaarwordingsaspect van gedachten en gevoelens kan ik mij vergissen, terwijl ik ervan overtuigd ben gelijk te hebben. Als ik mijzelf als angstig, boos of jaloers gewaarword, ‘weet ik zeker’ dat ik gelijk heb. Maar ik kan uit deze ‘zekerheid’ worden gehaald doordat ik word gecorrigeerd door anderen. Hoewel vergissingen bij een robot moeilijk voorstelbaar zijn, schijnen ze toch voor te komen. Bij geautomatiseerde medicijnenuitgifte in ziekenhuizen zien verpleegkundigen soms fouten optreden (Boonen, 2017), die zij vervolgens slechts met grote moeite kunnen tegenhouden (zij moeten daartoe tegen alle protocollen ingaan, wat tot enorme vertraging en frustratie leidt).
- 5 Niet alleen kan ik mij vergissen door de ‘zekerheid’ van mijn zelfgewaarwording en zelfbewustzijn (waaruit anderen mij kunnen bevrijden), maar ik doe ook dingen fout, in het bijzonder tegenover anderen. Ik kan anderen krenken, beledigen en zelfs vernederen, en dan naderhand spijt, berouw of wroeging hebben. Wat zouden dergelijke gevoelens – zeker als deze niet algoritmisch tot actie leiden – voor een robot kunnen betekenen? Zij zijn inefficiënt en leiden waarschijnlijk tot stilstand.
- 6 Mijn zojuist genoemde fouten die ik anderen aandoe, leiden tot schuld tegenover die anderen. Voor morele schuld kan ik worden vergeven, en ook kan ik anderen vergeven voor wat zij mij hebben aangedaan (Duyndam, 2004; 2003; 1999). Welke betekenis zouden morele schuld en vergeving echter kunnen hebben voor een robot? Een robot vergeven voor een fout is niet nodig, of in elk geval niet efficiënt. Ik kan ’m beter herprogrammeren.
- 7 Ik kan weemoedig of melancholisch worden over hoe mijn leven is gelopen, over de keuzes die ik daarin heb gemaakt, en over wat mij is overkomen. Voor een robot zouden dergelijke stemmingen volkomen inefficiënt en contraproductief zijn. Voor een mens misschien ook, maar ze horen wel bij het leven. Ze verdiepen het leven, en ze bepalen het verdere verloop van mijn leven, en mijn interpretaties van volgende ervaringen en gebeurtenissen.
- 8 Ik kan genieten, dat wil zeggen (met Levinas) een zekere onafhankelijkheid houden ten opzichte van datgene waarvan ik afhankelijk ben, in het bijzonder van mijn eerste levensbehoeften. In de afhankelijkheid van voedsel (noodzakelijk om te overleven) een zekere onafhankelijkheid bewaren door van je eten en drinken te genieten (Duyndam, 2009). Zo houd ik het midden tussen de asceet (die ernaar streeft zo onafhankelijk mogelijk te zijn) en de verslaafde (die volledig afhankelijk is van waaraan hij/zij verslaafd is). Een robot kent die balans niet, omdat hij én volledig afhan-

- kelijk is (van zijn programmatuur en algoritmes) én volledig onafhankelijk te werk gaat.
- 9 Ik ben in staat tot empathie met anderen. Niet door, zoals een robot, een aangeleerde en mogelijk geperfectioneerde reactie te geven, maar door werkelijk (inclusief zelfgewaarwording) *potentieel* te voelen wat de ander *actueel* voelt (Duyndam, 2012).
 - 10 Mij bewust van mijn feilbaarheid, fouten en schuld (5, 6), en mijzelf aanvaardend in de goede en verkeerde keuzes die ik in mijn leven heb gemaakt (7), en door een zekere onafhankelijkheid ten opzichte van mijn afhankelijkheden (8) empathisch tegenover anderen (9), kan ik hoopvol zijn ten aanzien van de toekomst. Ook 'tegen beter weten in', als hoop rationeel niet gerechtvaardigd is.
 - 11 Ik kan liefhebben en geliefd worden. Met een robot kun je seks hebben, misschien een gesprek voeren, en er zelfs aan gehecht raken. Maar een robot is principieel vervangbaar, en in die zin niet uniek. Liefde en vriendschap betekenen echter dat je vriend of je geliefde uniek is, in de zin van de enige, en onvervangbaar, en ook dat ik voor mijn vrienden en mijn geliefde(n) de enige ben, en onvervangbaar. Als een vriend uit mijn leven verdwijnt, of als een liefde uit raakt, houd ik wellicht andere vrienden over, die evenzeer uniek zijn, en ik vind misschien een nieuwe liefde, maar die ene vriend of die ene geliefde is er dan niet meer. Die is uniek en onvervangbaar (Duyndam, 2007).
 - 12 Hoewel de liefde natuurlijk het belangrijkste argument is, sluit ik af met het oordeel, omdat ik daar ook mee begon. Ik kan een oordeel vellen. Ik ben niet alleen in staat tot (wat Aristoteles noemt: de ethische deugd van het) geduld (argument 1), maar ook tot *phronèsis* – volgens Aristoteles de deugd der deugden. De verstandsdeugd *phronèsis* (praktische wijsheid) stelt mij in staat het algemene van een wet of regel te verbinden met het concrete van een casus. Wat betekent het in deze concrete situatie om geduldig te zijn, als het juiste midden tussen een tekort aan geduld (ongeduld) en een te veel aan geduld (lankmoedigheid)? *Phronèsis* (Latijn: *prudentia*) leert mij dit juiste midden te bepalen in een concrete situatie, leert mij de algemene deugd (geduld, moed, vrijgevigheid) te verbinden met de actuele situatie. Zo oordeelt een rechter ook die een situatie (een arme die een brood steelt) moet verbinden met een regel (stelen is verboden).

En zo gaat het ook in de beoordeling van een buitenspelsituatie in een voetbalwedstrijd. Waarmee we weer terug zijn bij ons openingsvoorbeeld van de scheidsrechters. Ook zij verbinden een concrete situatie in het veld met de algemene buitenspelregel. Dat doen zij goed als zij het prudent doen (12), met liefde voor hun vak en voor het spel (11), met leiderschap (10), met begrip voor de gevolgen van hun beslissing (9), in een zekere onafhankelijkheid (8), met zelfkennis en in het volle besef van hun feilbaarheid (7-4), zelfbe-

wust (3), vertrouwend op hun intuïtie (2), en vóór alles niet opgejaagd, maar met geduld (1). Kortom: als *arbiter cum arbitrio libero*, als arbiters met vrije wil.

Van de twaalf hier opgesomde kwaliteiten is het dus – voorzichtig gezegd – op z'n minst sterk de vraag of een robot hierin een scheidsrechter kan vervangen. Deze vraag is van groot belang voor humaniteit, waarvan we in het begin van dit artikel hebben gesteld dat het menselijk oordeelsvermogen er onderdeel van is: het vermogen om te oordelen over de waarheid, goedheid, schoonheid, en geloofwaardigheid van gebeurtenissen, situaties, teksten, opvattingen, enzovoort. De twaalf genoemde kwaliteiten veronderstellen allemaal, zij het in verschillende mate, zo'n oordeelsvermogen: (1) mijn leven en wat daarin gebeurt in het licht van andere levens en grotere verhalen zien, begrijpen en duiden; (2) irreflectieve zelfgewaarwording; (3) zelfbewust denken; (4) de mogelijkheid van vergissen en correctie; (5) spijt van, berouw en wroeging over misstappen; (6) schuld en vergeving; (7) de betekenis van weemoed en melancholie voor het verdere leven; (8) de balans van het genieten tussen afhankelijkheid en onafhankelijkheid; (9) empathie; (10) hoop; (11) liefde; en uiteraard (12) het wijze oordeel van de *phronèsis*.

Op basis van het onderzoek van onder anderen Frans de Waal en Harry Kunneman lijkt het waarschijnlijk dat ook andere dieren in zekere mate zo'n oordeelsvermogen hebben. Op basis van de casus van de scheidsrechter, en de uitwerking in de twaalf genoemde kwaliteiten, concludeer ik dat robots niet over een dergelijk vermogen beschikken.

Het humanisme is niet ten einde. Het humanisme verdedigt in mijn opvatting de vrije mogelijkheid van een oordeel. De mogelijkheid dus van een relatief of relationeel veerkrachtig en weerbaar *antwoord* op de druk, tegenslag en kwetsbaarheid waaraan onze wil blootstaat. Precies dit noem ik humanistisch: de relatief vrije mogelijkheid van een eigen antwoord. Zo kan het humanisme de intellectuele en artistieke beweging zijn die de manifestaties van cultuur (wetenschap, politiek, economie, religie, enzovoort) kritisch én zelfkritisch onderzoekt vanuit de bronnen en uitgangspunten van de cultuur, waarbij humaniteit de inzet, het omwille-waarvan is van de humanistische kritiek en zelfkritiek.

Literatuur

- Bell, Theo (2005), "Erasmus en Luther. Over de vrijheid van de wil", in: J. Duyndam, M. Poorthuis & T. de Wit (red.), *Humanisme en religie. Controverses, bruggen, perspectieven*, Eburon, Delft, 29-45.
- Boonen, Marcel (2017), *Nurses in Space. A Qualitative Empirical and Conceptual Study into the Use of a Drug Safety System by Nurses in an Orthopedic Ward of a General Hospital*, UvH, Utrecht.
- Duyndam, Joachim (2019), "Het humanisme ten einde?", in: *Tijdschrift voor Theologie*, 59/1, 53-63.
- Duyndam, Joachim (2017), "Resilience Beyond Mimesis. Humanism, Autonomy, and Exemplary Persons", in: B. Becking, A.-M. Korte & L. van Liere (red.), *Contesting Religious Identities. Transformations, Disseminations and Mediations*, Brill, Leiden/Boston, 175-193.
- Duyndam, Joachim (2012), "Emotionele veerkracht tussen existentiële eenzaamheid en empathische verbondenheid", in: T. Jorna (red.), *Mag een mens eenzaam zijn? Studies naar existentiële eenzaamheid en zingeving*, SWP, Amsterdam, 49-79.
- Duyndam, Joachim (2011), *De liefde van Alcestis. Over relationele weerbaarheid en hermeneutische levensbeschouwing*, SWP, Amsterdam.
- Duyndam, Joachim (2009), "Sincerely Me. Enjoyment and the Truth of Hedonism", in: B. Hofmeyr (red.), *Radical Passivity. Rethinking Ethical Agency in Levinas*, Springer, Heidelberg, 67-78.
- Duyndam, Joachim (2007), "Credible Fatherhood and Unique Identity. Toward an Existential Concept of Adoption", in: *The European Legacy*, 12, 729-735.
- Duyndam, Joachim, Poorthuis, Marcel & Wit, Theo de (red.) (2005), *Humanisme en religie. Controverses, bruggen, perspectieven*, Eburon, Delft.
- Duyndam, Joachim (2004), "Vernedering en herstel", in: *Tijdschrift voor Herstelrecht*, 4/1, 28-37.
- Duyndam, Joachim (2003), "Voorwaarden voor vergeving. Een stappenplan", in: *Justitiële Verkenningen* 29/3, 88-97.
- Duyndam, Joachim (1999), "Vergeving en de begeleiding daarbij", in: *Praktische Humanistiek*, 8/3, 114-124.
- Duyndam, Joachim (1997), *Denken, passie en compassie. Tijdreizen naar gemeenschap*, Kok Agora, Kampen.
- Girard, René (2008), *Mimesis and theory. Essays on Literature and Criticism, 1953-2005*, Stanford UP, Stanford (CA).
- Harari, Yuval Noah (2018), *21 lessen voor de 21^{ste} eeuw*, Rap, Amsterdam.
- Harari, Yuval Noah (2017a), *Homo Deus. Een kleine geschiedenis van de toekomst*, Rap, Amsterdam (2015).
- Harari, Yuval Noah (2017b), *Sapiens. Een kleine geschiedenis van de mensheid*, Rap, Amsterdam (2011).
- Harari, Yuval Noah (2017c), "Het humanisme heeft zijn langste tijd gehad", in: *NRC Handelsblad*, 25 maart 2017.
- Kant, Immanuel (2008), *Fundering voor de metafysica van de zeden*, Boom, Amsterdam (1785).
- Kunneman, Harry (2017), "Radicaal humanisme", in: *Waardenwerk*, 70/71, 18-34.
- Lamme, Victor (2011), *De vrije wil bestaat niet. Over wie er echt de baas is in het brein*, Bakker, Amsterdam.
- Sartre, Jean-Paul (1987), "Magie en emotie. Schets van een theorie van de gemoedsbewegingen", Boom, Meppel (1939).
- Swaab, Dick (2010), *Wij zijn ons brein. Van baarmoeder tot Alzheimer*, Contact, Amsterdam.
- Tongeren, Paul van (1986), "Kollektieve verantwoordelijkheid", in: *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 78, 117-134.
- Zalasiewicz, Jan, Williams, Mark, Steffen, Will & Crutzen, Paul (2010), "The New World of the Anthropocene", in: *Environmental Science & Technology*, 44, 2228-2231.