

PIM HASELAGER

Wat je maakt ben je zelf

Over kunstmatige intelligentie en menselijk
zelfbeeld

INLEIDING

Mensen kunnen meer over zichzelf te weten komen door zich te vergelijken met andere mensen (Gelukkig heb ik niet dezelfde woedeaanvallen als Flip. Ik kan me beter verplaatsen in anderen dan Marie, maar Jodokus is wel wat praktischer ingesteld dan ik ben). Maar zodra dergelijke vergelijkingen tot meer moeten dienen dan de vaststelling van (oppervlakkige) overeenkomsten of verschillen in gedrag wordt hun nut twijfelachtig. Als het erom gaat de *oorzaak* of aanzet tot het gedrag vast te stellen dan helpen vergelijkingen alleen als de interne werking van de ander of het andere bekend is. Als ik wil weten *waarom* ik zo snel geïrriteerd raak helpt het niet om te weten dat het bij Petrunia ook zo is, zolang ik niet weet waarom zij zo is. Het is (mede) om deze reden dat mensen zichzelf vergelijken met hun technologische producten. Onze machines hebben we zelf in elkaar gezet (in elk geval zijn er mensen die dat gedaan hebben en die we daarover vragen kunnen stellen). Door onszelf te vergelijken met machines kunnen we onszelf niet alleen plaatsen qua gedrag, maar ook een beter inzicht krijgen in de oorzaken daarvan. Een dergelijke praktijk kan heel nuttig zijn, maar heeft ook een keerzijde. Een vergelijking tussen mens en machine kan ook geforceerd werken, op het moment dat we per se iets van onszelf in machinetermen willen duiden terwijl dat eigenlijk niet kan. Een ander gevaar is dat we door de overeenkomsten

de verschillen niet meer zien. Wat begint als een hulpmiddel tot zelfbegrip kan als een dwangbuis eindigen.

In dit hoofdstuk zal ik nader ingaan op de gevolgen voor het menselijk zelfbeeld van vergelijkingen tussen mens en de producten van kunstmatige intelligentie, en dan met name op de relatie tussen de begrippen 'wil' en 'autonomie' en robots. Uitgangspunt is dat menselijke technologie mede het beeld vormt dat we van onszelf hebben, en dat we dus in de paradoxale situatie verkeren dat we deels ons zelfbeeld letterlijk *maken*.

AUTONOME ROBOTS?

Onze verbeelding wordt met grote regelmaat geprikkeld door boeken en films waarin door mensen gemaakte machines hun eigen weg proberen te gaan of zelfs de macht proberen over te nemen. Hun oncontroleerbaarheid en hun bovenmenselijke krachten en denkvermogens zijn voldoende fascinerend om een miljoenenpubliek te trekken, getuige de successen van science fiction verhalen (o.a. 'Ik robot' van Asimov), en films als Terminator, Bladerunner, A.I., I robot, Matrix, etc¹. In deze films worden vaak angstwekkende of verwarrende consequenties getrokken uit de ontwikkelingen van Kunstmatige Intelligentie (KI).

Angstwekkend zijn de plots waarin mensen te ver zijn gegaan in hun streven intelligentie te creëren en hun producten onbeheersbaar worden. Robots proberen met bruto geweld de wereld over te nemen van zich dapper verwerende mensen (Terminator). Machines hebben de mensen onderworpen en houden ze als energiebron (Matrix). Een ander thema betreft de soms grote overeenkomsten tussen robots en mensen, hetgeen verwarrend werkt op ons zelfbeeld omdat de scheidslijn tussen mens en machine niet zo absoluut is als hij misschien lijkt (Bladerunner, A.I., en commander Data van de tv-serie StarTrek: The next generation).

Een deel van deze zorgen en verwarring wordt misschien mede gevoed door steeds talrijkere claims van KI in verband met het bestaan van ‘autonome agenten’. Een ‘agent’ is de gebruikelijke aanduiding van virtuele (internet), kunstmatige (robots) of organische (mensen, dieren) systemen die actief kunnen optreden en zaken tot stand kunnen brengen of doelen bereiken. Op het eerste gezicht zijn claims over autonome agenten zeker verdedigbaar. Robots ondergaan niet slechts invloeden die tot reacties leiden (zoals bij stenen die van een heuvel afrollen bv.), maar ze kunnen ook zelf dingen bewerkstelligen. Ze zijn mobiel en hebben (kunstmatige) zintuigen. Ze zijn niet alleen reactief, maar ook proactief, doelgedreven en niet louter vanuit de omgeving aangestuurd. Ze zijn in staat om complexe handelingen in de werkelijkheid uit te voeren, vaak in een specifiek en zelfstandig geplande volgorde, en doen dat steeds vaker en onder steeds uitdagender omstandigheden, tot op Mars toe. Bovendien is hun gedrag vaak emergent in die zin dat het niet specifiek voorgeprogrammeerd is. Robots kunnen zich onvoorspelbaar gedragen en lokken dus soms beschrijvingen uit als ‘verrassend’, ‘origineel’, of zelfs ‘ongewenst’. Bovendien kan het gedrag van robots gestoeld zijn op de geschiedenis van hun interacties met de omgeving, zij kunnen leren van hun ervaringen in plaats van blind steeds dezelfde gedragingen te herhalen. Zodoende kan gezegd worden dat robots niet alleen dingen doen, maar ook dat ze ze ‘op hun eigen manier’ doen. Een ander belangrijk punt voor de discussie over autonome robots is dat mensen die robots observeren vaak onmogelijk een bepaalde mate van empathie buiten kunnen sluiten. ‘Meeleven met een robot’ klinkt misschien vreemd, maar bij het zien van een voortploeterende robot vragen de meeste toeschouwers zich bijna onwillekeurig af wat de robot wil doen. Ze raken betrokken bij zijn pogingen en zijn geneigd te zeggen, zelfs als de robot gefaald heeft, dat het op z’n minst ‘iets geprobeerd’ heeft.

Men zou natuurlijk kunnen beweren dat veel van de

hierboven gebruikte sleuteltermen (zoals handelen of doen, origineel, leren, iets willen, proberen een doel te bereiken) eigenlijk alleen in metaforische, en niet in letterlijke, zin van toepassing zijn op robots. Uiteindelijk, zo zou de gedachte kunnen zijn, zijn robots teveel van mensen afhankelijk, in verband met hun ontwerp, programma en functioneren, om daadwerkelijk autonoom te zijn. Ook deze tegenwerping bevat een kern van waarheid, maar een probleem bij een nader onderzoek ervan is dat termen als 'autonomie' en 'agent' niet een erg duidelijke, of zelfs maar eenduidige, betekenis hebben. Ze kunnen van meerdere definities voorzien worden, waarin dan weer termen opduiken (bv. doelen, kiezen, wil, of bewust) die zelf ook niet volledig helder zijn. Het lijkt er dan ook op dat autonomie een variëteit aan betekenissen kan hebben, die elk op z'n best een soort familiegelijkenis ten opzichte van elkaar vertonen, en die op uiteenlopende voorbeelden van toepassing kunnen zijn. Het is onder andere hierom dat een nadere analyse van de rol van technologie op ons zelfbegrip verhelderend kan zijn. Door te bekijken in welke opzichten het begrip autonomie al dan niet van toepassing kan zijn op robots kunnen we meer inzicht krijgen in de verschillende mogelijke betekenis ervan. Belangrijk daarbij is dat al te extreme (te ruime of te beperkte) interpretaties van autonomie in ieder geval bij aanvang vermeden moeten worden omdat daardoor onderzoek naar autonomie al bij voorbaat een kant op gestuurd wordt.

De recente robotica is voortgekomen uit teleoperatie, het op afstand bedienen van werktuigen (vaak een soort tangen). Op deze manier konden mensen handelingen uitvoeren op veilige afstand van gevaarlijke omgeving, bv. met hoge temperaturen. Rond het eind van de jaren '40 werd teleoperatie gebruikt in de eerste nucleaire reactoren. Het manipuleren van de werktuigen was vaak een vermoeiende bezigheid, onder andere wegens de gebrekkige feedback (met een hamer een spijker in een muur slaan is iets heel anders dan hetzelfde proberen te doen met een

hamer die niet in je hand maar in een op afstand bediende tang zit). Al snel ontstonden dus suggesties om de eenvoudige en repetitieve handelingen zich automatisch, dus zonder menselijke aansturing of interventie, te laten herhalen. Dit wordt wel ‘supervisory control’ genoemd. De mens geeft commando’s van hoog niveau (‘draai’ of ‘sla’) en laat dat het werktuig een gewenst aantal keer uitvoeren, terwijl de mens de uitvoering verder alleen maar gadeslaat.

Het is niet overdreven om in deze ontwikkeling de oorsprong van het begrip ‘autonome agenten’ binnen KI te zien. Een reductie van de noodzaak voor menselijke supervisie en interventie komt feitelijk neer op het vergroten van de autonomie van het werktuig. Gegeven deze geschiedenis is het begrijpelijk dat in handboeken over robotica definities gegeven worden die als volgt zijn te parafraseren:

Autonome agenten opereren onder diverse (uitdagende maar niet onmogelijk extreme) omstandigheden zonder hulp van een ontwerper of bestuurder, terwijl ze adequaat omgaan met onvoorspelbare gebeurtenissen in een realistische omgeving.

Het komt er dus op neer dat vanuit het perspectief van de robotica, de mate van autonomie van de robot toeneemt naarmate de supervisie en interventie door mensen afneemt.

AUTONOMIE, WIL, EN HET KIEZEN VAN DOELEN

In het dagelijks leven vindt men echter een duidelijk ander accent in verband met autonomie. Er wordt grotere nadruk gelegd op de *redenen waarom* men handelt, i.c. de doelen die men gekozen heeft, dan op de manier waarop de doelen bereikt worden. De *zelfstandige keuze* van doelen vormt de essentie van autonomie (‘Ik doe wat ik zelf wil’).

Deze nadruk op het zelfstandig kiezen van doelen vindt men ook in de filosofie. *Auto-nomos*, vanuit het Grieks vrij te vertalen als ‘een wet voor zichzelf zijn’, geeft het belang van zelf-regulatie of zelf-bestuur aan. Het hangt nauw samen met handelen vanuit eigen belang, en het maken van eigen keuzes, in plaats van het volgen van doelen die door anderen opgedragen worden. Het is uiteraard onmogelijk om hier recht te doen aan de lange wijsgerige geschiedenis in verband met autonomie. Duidelijk is in elk geval dat reeds Plato een duidelijk onderscheid maakt tussen veroorzaking middels keuze en fysische veroorzaking. In de *Phaedo* (98c-99b) zegt Socrates dat een verklaring van zijn zitten of liggen die zuiver in termen van zijn botten en ledematen is geformuleerd juist de echte oorzaak van zijn gedrag zal missen. Die bestaat namelijk uit Socrates’ doel, het resultaat van zijn keuze voor wat voor hemzelf het beste is om te doen.

Vanuit dit alledaagse en filosofische perspectief gezien schiet het autonomiebegrip binnen de robotica te kort. Robots mogen nog zo onafhankelijk handelen, en misschien zelfs ‘vrij’ bepalen *hoe* ze doelen willen bereiken, maar de *doelen zelf* zijn geselecteerd door menselijke ontwerpers en programmeurs. Dit essentiële verschil valt nog wat duidelijker te illustreren door de volgende serie (denkbeeldige) voorbeelden:

Een geval van *volledige autonomie* zou het volgende kunnen zijn: ik denk na over de voor- en nadelen van het drinken van bier, besluit dat ’t het beste zou zijn als ik het zelf zou brouwen, en doe dat vervolgens om het daarna op te drinken. Doelen, verlangens, argumenten, keuzes en handelingen zijn volledig met elkaar in overeenstemming en door mij bepaald.

Een *sterke wil* zou zich kunnen tonen als ik bier zou willen drinken, maar tegelijkertijd het verlangen heb om af te vallen, en ik uiteindelijk besluit om water te nemen. Hier is sprake van een zelfstandig opgelost (intern) conflict.

Wellicht vaker voorkomend is het geval van de *zwakke*

wil: ik wil gezond zijn, en denk dat bier slecht voor mijn gezondheid is, maar toch drink ik het.

Extremer is het geval van *waandenkbeelden*, waarbij ik bijvoorbeeld mijzelf voorhoudt dat buitenaardse wezens het drinkwater vergiftigd hebben en alleen het drinken van bier mij daarvan vrijwaart.

Tenslotte is er de mogelijkheid van *hersenspoeling*: zonder dat ik het door heb doen anderen mij denken dat bier goed voor me is (of me aan buitenaardse wezens doet ont-komen, of wat dan ook) en daarom drink ik het.

In zo'n reeks van voorbeelden verschuiven we langzaam van autonomie van wil (doel keuze) en handeling naar nog slechts autonomie van handelen (zwakke wil) tot uiteindelijk geen enkele vorm van autonomie meer. Daarmee neemt ook de mate van verantwoordelijkheid voor het handelen af. Met het oog op een dergelijke reeks van voorbeelden zou een willekeurige wetenschapper in de robotica een pijnlijke conclusie voor de voeten geworpen kunnen krijgen: robots zijn vergelijkbaar met gehersenspoelde wezens. Ze weten niet eens dat hun doelen door ons worden vastgesteld en daarom kunnen ze niet in de buurt komen van enige serieuze vorm van autonomie.

DE MECHANISERING VAN LICHAAM, DENKEN EN WIL

Een dergelijke conclusie is echter niet de enig mogelijke. Want hoe zeker kunnen wij er eigenlijk van zijn dat de door ons geselecteerde doelen daadwerkelijk volledig autonoom, volledig door ons zelf, gekozen zijn? Bestaat er wel zoiets als complete autonomie? Over dit onderwerp zijn hele bibliotheken volgeschreven, dus ook hier moet ik mij beperken. Ik wil me vooral richten op een interessante ontwikkeling die zich afspeelde in de 17de en 18de eeuw die een illustratie vormt van de relatie tussen technologie en wijsgerig denken over menselijke wilsvrijheid.

In een periode van ruwweg twee eeuwen voltrekt zich een verandering in het menselijk zelfbeeld die wel aangeduid kan worden als de mechanisering van het lichaam en, in toenemende mate, van het denken. Een redelijk beginpunt vormt het in 1543 verschenen *De humani corporis fabrica* waarin Vesalius (1514-1564) uiterst gedetailleerd diverse onderdelen van het menselijk lichaam in kaart bracht. Harvey's (1578-1657) these dat het hart als een pomp gezien moet worden vormt een treffende illustratie van de vergelijking tussen mens en machine. Niet veel later maakte Borelli (1608-1679) in zijn *De motu animalium* expliciete vergelijkingen tussen een gewicht dragend menselijk lichaam en een hefboom. Hij observeerde verder dat spieren zich gedragen als natgemaakte touwen: ze trekken samen en kunnen zo gewichten heffen. Hoewel vergelijkingen tussen machines en het menselijk en dierlijk lichaam steeds meer gemeengoed worden lijkt er een fundamentele kloof tussen machines en denken te blijven bestaan. Toch bleef dat niet heel lang zo.

Een invloedrijke technologie in die genoemde periode betrof het uurwerk. Klokken vormden geavanceerde kunststukjes maar waren tegelijkertijd steeds veelvuldiger aanwezig in de maatschappij.² Het was Descartes (1600-1648) die een beroemde en beruchte vergelijking formuleerde tussen dergelijke machines en dieren. Volgens zijn 'dier-machine' these functioneren dieren als uurwerken:

'Ik kan de mening van Montaigne en anderen die begrip of denken toeschrijven aan dieren niet delen. (..) Ik weet dat dieren veel dingen beter doen dan wij, maar dat verrast me niet. Dat kan zelfs gebruikt worden om aan te tonen dat zij natuurlijk en mechanisch handelen, zoals een klok die beter dan wij tijd kan aangeven. Als de zwaluwen komen in de lente, opereren zij ongetwijfeld als uurwerken'.³

Malebranche (1638-1715) schreef in dezelfde tijd:

‘Dieren zijn zonder verstand of bewustzijn in de gebruikelijke betekenis. Zij eten zonder eetlust, zij schreeuwen zonder pijn, zij groeien zonder dit te begrijpen, zij begeren niets, zij vrezen niets, zij beseffen niets. Indien zij zich misschien eens zo gedragen dat dit verstandig schijnt, dan is dat het gevolg van een lichaamsinrichting, die God zo heeft besteld, dat ter wille van zelfbehoud zij redeloos, zuiver mechanisch, alles ontlopen wat hen dreigt te verderven’.⁴

Hoewel volgens Descartes het voor mensen kenmerkende denkvermogen buiten het bereik van de mechanische benadering lag, ontstonden er in zijn tijd wel degelijk machines die gezien kunnen worden als voorzichtige pogingen om aspecten van het denken te mechaniseren, namelijk het rekenen. Kort na elkaar ontwikkelen Von Napier (1617), Schickard (1623), Pascal (1642) en Leibniz (1671) rekenmachines.⁵ In 1642 was Pascal de eerste die het mechanisch rekenen publiekelijk kon tonen; zijn ‘Pascaline’ kon optellen en, met enige interventie, aftrekken. Vroeg in de 18de eeuw waren de machines van Vaucanson een ware sensatie omdat ze zo ‘levensecht’ waren. Deze robots *avant la lettre* konden trommelen of fluitspelen, en de mechanische eend kon zwemmen en eten. Niet veel later trok de Franse filosoof en arts De la Mettrie in 1747 een conclusie uit al deze ontwikkelingen met zijn boek *L’Homme Machine*. De mens is een machine: ‘de mens is een samenstelsel van tandraden die elkaar activeren’.

Automaten en rekenmachines lieten zien dat belangrijke onderdelen van het menselijk lichaam en het menselijk denken zuiver mechanisch nagebootst konden worden. Toch werden daaruit in eerste instantie nog geen conse-

quenties getrokken voor het denken over autonomie en de wil. Zo schreef Pascal:

‘De rekenkundige machine produceert effecten die dichter in de buurt komen van denken dan wat dieren ook maar kunnen doen; maar het kan niets doen wat ons ertoe zou kunnen leiden om te zeggen dat het een vrije wil heeft, zoals de dieren hebben’.⁶

Volgens Descartes verwees het begrip ‘wil’ dan ook naar een specifiek vermogen om bewegingen te veroorzaken, zodat bijvoorbeeld het verlangen om te lopen leidt tot de bewegingen van onze benen.⁷ De wil bestaat uit ons vermogen om iets wel of niet te doen en is *niet beperkt in enig opzicht*.⁸ Maar deze positie van Descartes heeft niet heel lang standgehouden. De Schotse filosoof Hume stelde in zijn ‘Een verhandeling over de menselijke natuur’ (III, ii, 1739) dat de keuzes van de wil veroorzaakt worden door motieven, en dat hoewel we een gevoel van vrijheid in onszelf ervaren, een toeschouwer over het algemeen onze handelingen kan *afleiden* uit onze motieven en ons karakter. Kort daarop verdedigde Hartley de stelling dat handelingen *mechanisch* veroorzaakt werden door lichamelijke en geestelijke toestanden:

‘elke handeling resulteert uit voorafgaande omstandigheden van lichaam en geest, op dezelfde wijze en met dezelfde zekerheid, als andere effecten ontstaan vanuit hun mechanische oorzaken, zodat een persoon niet onverschillig een handeling A, of de daartegenover gestelde handeling a kan doen, terwijl de voorafgaande omstandigheden gelijk zijn; in plaats daarvan moet hij noodzakelijkerwijs een van de twee doen, en die alleen’.⁹

In de eeuw die Hartley van Descartes scheidt werd het contrast tussen handelingen veroorzaakt door keuzes en door mechanische oorzaken onduidelijker. De wil wordt

niet langer als volledig vrij gezien, maar als bepaald door andere factoren, zoals motieven.

Hoewel de filosofie in de 17de eeuw begon met het Cartesiaanse idee van een onbeperkt vrije wil eindigde het, in ieder geval in het werk van vooraanstaande filosofisch-wetenschappelijke denkers, in de 18de eeuw met een opmerkelijk mechanistisch beeld. Maar als motieven als mechanische oorzaken van handelingen gezien kunnen worden, dan kan al snel de vraag ontstaan waardoor motieven op hun beurt veroorzaakt worden. Een voor de hand liggende suggestie is dan dat motieven bepaald zouden kunnen worden door karakter, en dat karakter weer wordt bepaald door erfelijke en omgevingsfactoren. Dergelijke vragen houden nooit meer op, en in ons huidige tijdvak heeft de bestudering van hersenen en genetische factoren weer een diepere laag in kaart gebracht.

CONCLUSIE

De willekeurige vertegenwoordiger van de robotica, die wij eerder om de oren hebben geslagen met de stelling dat zijn of haar robots eigenlijk gehersenspoelde creaturen zijn, en niet in het bezit van een werkelijke vorm van autonomie, zou dus fijntjes kunnen opmerken dat hetzelfde eigenlijk ook wel eens zou kunnen gelden voor ons mensen. Wij denken onze doelen zelf te kiezen, maar wat garandeert ons dat onze ultieme beweegredenen niet bepaald zijn door oorzaken die zich buiten ons gezichtsveld en onze invloed afspelen? Wij zouden dus wel degelijk vergelijkbaar kunnen zijn met onze producten, zoals robots. Als er al een verschil tussen mens en robot zou zijn dan is dat niet per se dat robots 'ware' autonomie zouden ontberen, maar wellicht dat mensen de ontorechte 'overtuiging' hebben dat zij wel waarachtig autonoom zijn. Sterker nog, zou deze roboticus kunnen vragen; laten onze tech-

nologische producten door hun complexe, doelgerichte maar soms onvoorspelbare en vaak empathie opwekkende gedragingen niet juist zien dat het hele idee van ‘echte’ autonomie helemaal niet noodzakelijk is om vergelijkbaar menselijk handelen na te bootsen en te verklaren?

Het is in dit verband goed om erop te wijzen dat er tegenwoordig steeds meer gesproken wordt over het idee dat de menselijke wil in verband met ons gedrag niet exact de rol speelt of zou kunnen spelen die we er traditioneel aan toekennen, namelijk die van de ‘bestuurder’ van ons gedrag. Een belangrijke representant van deze stroming is Daniel Wegner, die stelt dat het idee van een wil die gedrag veroorzaakt een *illusie* zou kunnen zijn. Hij wijst onder andere op neurowetenschappelijk onderzoek dat aangeeft dat de hersenprocessen die correleren met het gedrag waartoe een persoon besluit, reeds aanwijsbaar zijn in het brein van die persoon nog *voordat* hij of zij zich bewust is van datzelfde wilsbesluit. Met andere woorden, de bewuste wil hobbelt slechts wat achter de (hersenen)feiten aan. Hoewel er altijd commentaar mogelijk is op het specifieke bewijsmateriaal dat wordt aangevoerd, valt niet te ontkennen dat de empirische ondersteuning voor een dergelijke positie in de laatste jaren flink toeneemt.

In de inleiding zei ik dat we onszelf beter kunnen leren begrijpen door onszelf te vergelijken met machines. Een beter zelfbegrip *kan* tot gevolg hebben dat we een illusie armer worden. Of dat in verband met wil en autonomie ook het geval zal zijn staat nog te bezien, al zijn er aanwijzingen dat het die kant op zou kunnen gaan. Maar in de inleiding heb ik ook aangegeven dat een mens-machine vergelijking *geforceerd* kan werken en ons blind kan maken voor daadwerkelijk bestaande verschillen. Het is misschien goed dit punt nog even te benadrukken door een uitspraak te memoreren van de Griekse geneeskundige Hippocrates (460-377), die reeds lang geleden zei dat het vergelijken van mensen met hun producten getuigt ‘van een buitengewoon armoedig mensbeeld’.

Als het credo van dit stuk ‘Wat je maakt ben je zelf’ ook maar ten dele juist zou zijn, dan is het zaak de ontwikkelingen binnen de KI, en aanverwante wetenschappen zoals de neurowetenschap, nauwlettend te blijven volgen. Aan de ene kant dienen we bereid te zijn ons zelfbeeld te corrigeren vanwege voortschrijdend wetenschappelijk inzicht in onze machinerie, terwijl we aan de andere kant moeten waken voor het overhaast terzijde schuiven van diepgewortelde concepten zoals wil en autonomie die deel lijken uit te maken van de kern van wat en wie we zijn. De *vernieuwing* van het menselijk zelfbeeld is niet alleen een kwestie van technologische vooruitgang, maar ook van onze reflectie daarop. Voor allebei is wijsheid nodig.

DANKWOORD

Rolf Hoekstra dank ik voor zijn aanmoediging en constructief commentaar. Dit hoofdstuk bevat enkele fragmenten van een voor uitgeverij Boom te schrijven manuscript ‘Filosofie van de Kunstmatige Intelligentie’.

NOTEN

- 1 Er waren overigens al verhalen en zelfs al films met het thema van KI voordat die wetenschap officieel bestond, zie bv. klassiekers als *Metropolis* (1926), *The day the earth stood still* (1951) en *Forbidden planet* (1956). Capeks toneelstuk R. U. R. (Rossum’s Universele Robots) werd voor het eerst opgevoerd in Praag in 1921. In dit stuk werd het woord ‘robot’ (Tsjechisch voor dienstbaar, slaafs) voor het eerst gebruikt. Rossum is de hoofdpersoon die een vorm van intelligente maar dienstbare arbeiders maakt, die uiteindelijk in opstand komen.

- 2 Klokken zijn bovendien interessant omdat ze, net als computers en robots, tot op zekere hoogte het vermogen hebben om zelfstandig te kunnen functioneren en hun rol in het toenmalige denken over onszelf is daarom wel vergeleken met die van computers in onze tijd; zie D. Draaisma, 'De machine achter het gordijn: Drie eeuwen debat over bewustzijn en mechanisme', in: P. Hagoort en R. Maesen (red.), *Geest, Computer, Kunst*, Utrecht, Grafiet, 1986.
- 3 Descartes, 'Brief aan de Markies van Newcastle', 23 november 1646.
- 4 H.C.D. de Wit, *Ontwikkelingsgeschiedenis van de biologie*, dl. 1, Wageningen, Pudoc, 1982.
- 5 Von Napiers 'beenderen' zijn eigenlijk meer een hulpmiddel dan een machine: tien houten stokjes waar getallen en hun meervouden op stonden, zodat een rekentabel kon worden gevormd door de stokjes naast elkaar te plaatsen. Het betrof geen mechanisme, menselijke interventie en supervisie waren nog steeds nodig.
- 6 G. Ifrah, *The Universal History of Numbers: From Prehistory to the Invention of the Computer*, John Wiley and Sons, 2001.
- 7 Descartes, *De hartstochten van de ziel*, xvii, 1649.
- 8 Descartes, *Meditaties over de eerste filosofie*, AT 56-57, 1641.
- 9 D. Hartley, *Observaties over de mens*, Conclusie, 1741.