Denkend aan robots zien de meeste mensen apparaten voor zich die auto's in elkaar zetten in grote, anonieme fabriekshallen. De robots uit Hollywoodfilms daarentegen zijn bijna-menselijke, gevoelige apparaten die emoties tonen en waarmee we relaties kunnen hebben. Is dit een realistisch beeld van de toekomst? De meningen van futurologen lopen uiteen, maar Roy de Kleijn en Bernhard Hommel geloven dat humanoïde robots een kleurrijke toekomst tegemoet gaan.

GESCHIEDENIS EN TOEKOMST VAN DE MENS-ROBOTINTERACTIE

VERLIEFD OP EEN ROBOT

DE GESCHIEDENIS VAN ROBOTS

De eerste robots werden in de jaren 50 gemaakt, toen de massaproductie van auto's op gang kwam en ze nuttig waren als vervangers van menselijke arbeid. Lopendebandwerk waar nauwkeurigheid én kracht voor nodig waren, zoals het fabriceren van auto's, bleek een uitstekende niche voor deze apparaten. Rond dezelfde tijd vond in de academische psychologie de *cognitieve revolutie* plaats, waaruit de cognitieve psychologie zoals we die nu kennen ontstond. In dit nieuwe werkveld van de cognitieve wetenschap, waar disciplines als psychologie, neurowetenschap en computerwetenschappen samenkomen, groeide de *computeranalogie* in populariteit. Misschien, dacht men, kunnen we de mens beschouwen als een enorm ingewikkelde computer, met invoer (zintuiglijke waarneming), verwerking (cognitieve processen), en uitvoer (gedrag).

Dit nieuwe inzicht bood niet alleen mogelijkheden voor psychologen, die met een frisse blik de mysteries van deze mentale processen wilden onderzoeken, maar opende ook de deur voor onderzoekers die intelligente systemen wilden bouwen. De cognitieve robotica werd geboren: door gebruik te maken van robots konden alle drie onderdelen van de computeranalogie onderwerp van onderzoek worden. Invoer kon worden geleverd door sensoren (camera's, microfoons), uitvoer kon plaatsvinden door beweging van de robot en manipulatie van de omgeving, en de verwerking door een regelsysteem of een computerprogramma. Hoewel sensoren en regelsystemen in deze tijd nog betrekkelijk rudimentair waren, legde een architectuur als deze de fundering voor autonome, intelligente systemen, en optimisten voorspelden dat het niet lang zou duren voordat communicatie en interactie met machines niet te onderscheiden zou zijn van interactie tussen mensen.

Onderzoek in de robotica boekte met rasse schreden vooruitgang, en in de jaren 70 begonnen de eerste robots enigszins op mensen te lijken, de zogenoemde *humanoïde* robots. Humanoïde robots kenmerken zich door hun gelijkenis met het menselijk lichaam. Vaak beschikken ze over twee armen (effectoren) waarmee ze de omgeving kunnen manipuleren en een hoofd met ogen (camera's), en soms zelfs twee benen om zich voort te bewegen.

Het ontwikkelen van robots die net zo flexibel en slim zijn als mensen biedt legio mogelijkheden. Zo zouden ze ons kunnen helpen bij praktische zaken als boodschappen doen, koken en ander handmatig werk, maar ze zouden ook ouderenzorg op zich kunnen nemen of humanitaire hulp in ramp- of oorlogsgebieden. Als robots inderdaad zo veel dingen gaan kunnen en een belangrijke rol gaan vervullen in onze samenleving, is het belangrijk om kritisch te kijken naar de interactie tussen ons en onze robotische 'medemens'.

ROBOTS IN DE MAATSCHAPPIJ Zonder het te beseffen komen we dagelijks in aanraking met robots. Afhankelijk van de gehanteerde definitie hebben we te maken met robotstofzuigers en drones, maar ook met een automatische piloot in het vliegtuig. Deze robots zijn sterk gespecialiseerd en onze

FOTO'S: HOLLANDSE HOOGTE

interactie met ze heeft geen sterke emotionele lading. We zien ze niet als gelijken, voelen ons niet beledigd als ze een 'fout' maken, en vinden ze niet dom als ze iets niet kunnen (hoewel we allemaal wel eens tegen onze computer schreeuwen als die weer eens vastloopt).

Toch betekent dit niet dat we geen sociale band met ze kunnen opbouwen. In een beroemd geworden experiment lieten Heider en Simmel (1944) zien dat we intenties toeschrijven aan simpele bewegende geometrische vormen, en Braitenberg (1984) opperde met zijn gedachte-experimenten dat extreem simpele robots dezelfde gevoelens opwekken. Een ander goed voorbeeld hiervan is de in de jaren 90 razend populaire Tamagotchi: een 'elektronisch huisdier' dat gevoerd moest worden en aandacht eiste, waarbij opviel dat mensen zich er razendsnel aan hechtten. Ook uit het veld van de neurowetenschappen komt bewijs dat we robots als méér zien dan domme machines. Zowel Oberman et al. (2007) als Gazzola et al. (2007) hebben laten zien dat het observeren van acties uitgevoerd door robots het menselijke spiegelneuronsysteem activeert, iets wat voorbehouden leek te zijn aan het observeren van *biologische* acties van mensen of dieren. De mate van deze activering bleek zelfs niet te verschillen bij het observeren van menselijke of robotische actie. Naast intentionaliteit kunnen wij, afhankelijk van de complexiteit van de robot, ook andere eigenschappen toeschrijven aan robots.

Zo blijkt dat als mensen in contact komen met humanoide robots, we nog veel meer geneigd zijn om menselijke eigenschappen aan ze toe te schrijven – een fenomeen dat *antropomorfisme* wordt genoemd. Zo worden robots eigenschappen als creativiteit en het hebben van een eigen wil toegeschreven. Heel vreemd is dit eigenlijk niet: het gedrag van een complexe robot is, net als bij mensen, het product van zowel een interne staat als fysieke mogelijkheden en beperkingen. Mensen verschillen in de mate waarin ze robots antropomorfiseren. Zo vertonen mensen met schade aan de amygdala en mensen met autisme minder antropomorfisme dan neurologisch gezonde personen (Waytz et al., 2010).

Dit antropomorfiseren van robots heeft interessante sociale gevolgen. Zo leidt het toeschrijven van een 'mind' aan robots noodzakelijkerwijs tot het toeschrijven van verantwoordelijkheid aan robots. Als een robot beslissingen kan nemen en hiernaar kan handelen, kan de robot verantwoordelijk worden gehouden voor de gevolgen hiervan. Een ander gevolg van het hebben van een 'mind' is een bewustzijn, wat leidt tot het verworden tot een morele 'agent', met alle

rechten en plichten die daarbij horen (Gray et al., 2007).

Convergent bewijs voor deze 'vermenselijking' van robots wordt gevonden in een bekend paradigma uit de sociale psychologie en economie, de *economische spellen*. In een aantal van deze spellen wordt de proefpersoon een bedrag gegeven, dat hij of zij vervolgens mag verdelen naar eigen inzicht. Hierbij wordt duidelijk dat vaak vrijwillig een deel van een toegekend bedrag afgestaan wordt aan robots, die hier overduidelijk niets aan hebben. Ook tonen mensen een verhoogde fysiologische respons wanneer ze een video te zien krijgen waarin een robot wordt mishandeld. Met andere woorden, mensen lijken *empathie* te voelen voor robots (Van Es, 2015; Rosenthal-Von der Pütten et al., 2013). Maar hoe ver gaat deze empathie? Als ons huisdier wordt mishandeld roept dat ongetwijfeld een fysiologische respons op, en we zijn ook zeker bereid te betalen voor medische zorg voor ons huisdier. Toch betekent dat niet dat huisdieren gelijk staan aan mensen wat betreft onze empathische gevoelens.

Hoe zit dat met robots? We vinden het zielig als ze pijn hebben, en we zijn bereid ze geld te geven. Toch zouden we niet zeggen dat we 'bevriend' met ze zijn, en we hebben geen romantische relaties met ze. Waarom eigenlijk niet? Wat is daar voor nodig? Zonder in trivialiteiten als voorplanting te vervallen (we kunnen tenslotte wel verliefd worden op onze onvruchtbare maar ontzettend knappe buurvrouw), is het duidelijk dat menselijke relaties méér vergen dan alleen empathie. Wanneer aan mensen wordt gevraagd wat ze zo leuk vinden aan hun partner, valt op dat de partner vooral betrouwbaar, grappig, en aantrekkelijk moet zijn (Buss, 1994). Betrouwbaar zijn robots nu al, en grappig kunnen ze ook zijn (we komen hier later op terug), maar aantrekkelijk?

REALISME VAN ROBOTS Ontwikkelingen in het maken van robots die vooral fysiek sterk op mensen lijken – ook androiden genoemd – hebben vooral in de afgelopen tien jaar een grote vlucht genomen. Er zijn sinds een aantal jaar robots die fysiek, in stilstand, bijna niet te onderscheiden zijn van mensen. Het is gemakkelijk voor te stellen dat het fysiek realisme van robots zou leiden tot grotere acceptatie, of in elk geval meer empathische gevoelens voor dit soort robots. Het mes lijkt echter aan twee kanten te snijden.

De Japanse hoogleraar robotica Masahiro Mori (1970) kwam er in de jaren 70 achter dat de relatie tussen het fysiek realisme van een humanoïde robot en de empathie die hij opwekt geen monotone relatie is. Mensen vinden robots leuker naarmate ze meer op mensen lijken, maar slechts tot op zekere hoogte. Wanneer een robot sterk op een mens lijkt maar er nog wel van te onderscheiden is, roept dit een gevoel van afkeer op bij mensen. Mori doopte dit gebied van 'bijna-maar-net-niet' menselijkheid de *uncanny valley*. Verklaringen voor dit verschijnsel variëren, maar Mori zelf dacht dat het te maken zou hebben met een evolutionair beschermingsmechanisme om ziekteverwekkers te vermijden, te vergelijken met de afkeer die we voelen bij het zien (of ruiken) van dode mensen en uitwerpselen.

We moeten kritisch kijken naar de interactie tussen ons en onze robotische 'medemens'

Niet alleen in fysiek opzicht worden robots steeds menselijker. Sinds de jaren 60 wordt er alles aan gedaan om computersystemen – en daarmee robots – steeds intelligenter en daarmee menselijker te maken. Waar het schaakspel door sommigen werd gezien als het summum van intelligentie (intelligente mensen houden tenslotte van schaken) had de wiskundige Alan Turing (1950) hier een ander idee over: hij opperde dat een computersysteem pas écht intelligent is als je er een gesprek mee kunt voeren dat niet te onderscheiden is van een normaal mens-tot-mens-gesprek.

Hij formaliseerde een test waarmee duidelijk zou worden of een computer intelligent is of niet, de zogenoemde *imitation game*, die we nu kennen als de *Turing-test*. In deze test chat een persoon (via tekstberichten) met twee andere *agents*: één daarvan is een computer, de ander een mens. Na een klein halfuur wordt gevraagd om te bepalen welke van de twee gesprekspartners de computer was, en welke de mens. Deze test wordt minstens één keer per jaar uitgevoerd als de *Loebner Prize contest*. Met een beloning van $100.000 trekt deze wedstrijd ieder jaar de beste Kunstmatige Intelligentie (KI)-bots aan om zich te bewijzen, maar nog zonder resultaat: de Grand Loebner Prize is nog nooit uitgereikt.

Maar alle hoop is nog niet verloren, beweren sommige mensen. In een aangepaste Turing-test is het computerprogramma *Eugene Goostman* erin geslaagd om na een Engelstalig gesprek van vijf minuten een derde van de juryleden ervan te overtuigen dat hun gesprekspartner een dertienjarige Oekraïense jongen was, met Engels als tweede taal. In dit

kader is de test zodanig ingeperkt dat een kritische lezer zich af kan vragen of dit nu echt een demonstratie van intelligentie is. Ook Hugh Loebner zelf, de organisator van de Loebner Prize, is niet overtuigd: *Eugene Goostman* haalde in Loebners test niet eens de top vijf.

Een normaal gesprek valt dus nog niet te voeren met een robot. Maar er wordt wel degelijk vooruitgang geboekt, ook op andere gebieden die we tot nu toe 'typisch menselijk' noemden. Zo zijn algoritmen die muziekstukken componeren, en zelfs improviseren in samenwerking met menselijke muzikanten, behoorlijk goed in wat ze doen. En zelfs humor is niet langer alleen aan mensen besteed. Geen geringe prestatie, aangezien veel vormen van humor erg afhankelijk zijn van de mogelijkheid tot natuurlijke-taalverwerking, een vaardigheid waarin computers nog zeer goed te onderscheiden zijn van mensen, zoals de Turing-test laat zien. Recente algoritmen kunnen woordgrappen maken en raadsels verzinnen (Weinberg et al., 2007; Ritchie et al., 2007).

Maar ook op gedragsniveau lopen we tegen een hindernis aan. De uncanny valley die zichtbaar is bij toenemend fysiek realisme lijkt ook een rol te spelen bij het *gedrag* van een robot. In een uitgekiend experiment werden twee dimensies van gedragsmatig realisme (lichaamsbeweging en lipsynchronisatie) gemanipuleerd terwijl fysiek realisme constant werd gehouden. Ook hier werd een uncanny valley gevonden (Groom et al., 2009).

Later onderzoek heeft laten zien dat de uncanny valley zeker niet allesbepalend is. Een net zo belangrijk aspect lijkt de *congruentie* tussen uiterlijk en gedrag: als een robot er menselijk uitziet willen we dat het gedrag ook menselijk is, als een robot er niet menselijk uitziet willen we geen menselijk gedrag observeren. En er lijkt meer bewijs voor het belang van congruentie. Uit een studie naar de congruentie tussen realisme van gezicht en stemgeluid in robots blijkt dat ook incongruentie tussen deze twee eigenschappen een uncanny valley veroorzaakt (Saygin et al., 2012; Mitchell et al., 2011). Dit biedt handvatten voor de ontwikkeling van moderne humanoïde robots, het is dus zaak om – zolang de wetenschap nog geen perfecte robotlichamen kan voortbrengen – ook het uiterlijk en gedrag van robots te differentiëren van dat van mensen.

DE INTERACTIE TUSSEN MENS EN ROBOT

Mensen reageren dus verschillend op robots. Ze verschillen in de mate van antropomorfisme, maar ook het realisme van uiterlijk en gedrag van een robot beïnvloedt wat we van ze vinden. Hoe we op robots reageren lijkt ook een functie te zijn van hoe bekend we met ze zijn. In een studie hebben proefpersonen geïnteracteerd met een sociale robot, door proto-conversaties te voeren waarbij de robot met gezichtsuitdrukkingen terugkoppeling gaf aan de menselijke partner. Het werd duidelijk dat mensen snel de 'lichaamstaal' van zulke robots leren interpreteren. Communicatie verliep soepeler en sneller naarmate de tijd, en dus blootstelling, vorderde. Mensen zijn dus in staat om bij interactie met robots dezelfde sociale 'cues' te gebruiken als ze gewend zijn bij andere mensen te gebruiken (Breazeal, 2003).

Maar wat betekent dit voor het praktische sociale nut van robots? Hebben zorgrobots inderdaad een toegevoegde waarde? Omdat het om een relatief moderne ontwikkeling gaat – en zorgrobots nog niet op grote schaal worden ingezet – zijn er nog niet veel empirische studies die ons het antwoord kunnen geven. In de studies die er wel zijn gedaan lijkt het inzetten van zorgrobots wel degelijk een verbetering in de kwaliteit van leven te bewerkstelligen. In één studie werden robot-honden ingezet in drie verzorgingstehuizen, en werd het effect vergeleken met diertherapie met echte honden. Er werd gevonden dat bewoners zich hechtten aan zowel de robot- als de echte honden, en dat beide een significante verlaging van eenzaamheidsklachten teweegbrengen (Banks et al., 2008).

Ook in andere sociale situaties, zoals in een therapiesetting, worden robots al beperkt ingezet. Er zijn nog geen grootschalige studies naar de effectiviteit van zulke therapieën gepubliceerd, maar de eerste meta-analyses van de beschikbare kleinschalige studies tonen aan dat psychotherapie met robots een beter resultaat laat zien dan traditionele, niet-robotische therapie (Costescu et al., 2014). Ook tonen kinderen met autisme vaker imitatiegedrag bij het zien van een robotarm die iets pakt dan een menselijke arm, terwijl gezonde kinderen het omgekeerde effect laten zien; een effect dat verklaard kan worden door juist het ontbreken van sociale cues. Ook lijken robots die wél emotie tonen, maar in mindere mate dan mensen, een waardevolle toevoeging te zijn bij therapie voor kinderen met autisme (Pierno et al., 2008; Kozima et al., 2007).

HOE KAN HET DAT ROBOTS SOCIAAL ZIJN?

Om robots te maken die goed met mensen omgaan, is het van belang dat ze het gedrag van mensen kunnen interpreteren. Bevredigend contact met iemand is geen eenrichtingsverkeer, de kwaliteit van interactie tussen twee sociale partners is minstens net zo belangrijk als de kwaliteiten van één van hen. Een belangrijke vaardigheid bij de interactie tussen mensen

Wij geloven dat de robot, op lange termijn, als romantische partner uitkomst kan bieden voor mensen die geen partner kunnen krijgen

is emotieherkenning. Robots worden gezien als levenloze computersystemen die maar moeilijk iets met emotie kunnen.

In omgevingen als ziekenhuizen of verzorgingstehuizen, waar de 'zorgrobot' een rol zou moeten gaan spelen, is het van belang dat de interactie tussen robot en mens betrekkelijk natuurlijk verloopt. Patiënten en bewoners zouden niet speciaal opgeleid hoeven worden om met deze robots om te gaan, maar zouden met hen moeten kunnen interacteren op menselijke wijze. Hiervoor zijn niet alleen natuurlijke-taalverwerking en het autonoom kunnen navigeren van belang. Ook het kunnen omgaan en inspelen op de gemoedstoestand van de menselijke partner is – vooral in situaties als deze, waar de partner een afhankelijke rol heeft – essentieel voor het opbouwen van een sociale band.

Onderdeel van deze 'emotionele robots' is gezichtsherkenning zoals we dat kennen van onze fotocamera's en sociale netwerksites. Het is van belang dat de robot weet of hij te maken heeft met een levenloos voorwerp of een mens. Voor mensen lijkt dit een bijna moeiteloos proces: al na twee maanden na de geboorte laten baby's een voorkeur zien voor gezichten boven andere potentieel interessante objecten. Er bestaan verschillende theorieën over de mechanismen die baby's hiertoe in staat stellen. Eén van de meest invloedrijke daarvan stelt dat baby's geboren worden met een schema van visuele informatie dat gezichten kenmerkt, zoals de relatieve locatie van drie gebieden van hoog contrast (ogen en mond) binnen een groter gebied van ongeveer de juiste grootte (Valenza et al., 1996). Een van de populairste algoritmen voor kunstmatige gezichtsherkenning, het Viola-Jonesalgoritme (Viola en Jones, 2001), vertoont veel overeenkomsten met deze theorie. Dit algoritme scant een visuele invoer op het voorkomen van bepaalde kenmerken zoals de drie gebieden van hoog contrast zoals hierboven beschreven.

Een robot weet dus of hij te maken heeft met een mens of niet. Een volgende stap in het sociale mens-robotproces is het interpreteren van het gezicht: wie is het (identificatie) en wat zegt het (emotieherkenning)? De algoritmen achter deze processen zijn veelal op kunstmatige neurale netwerken gebaseerd, en nemen toe in betrouwbaarheid en nauwkeurigheid. In 2014 kreeg het algoritme *GaussianFace* aardig wat media-aandacht, omdat het het eerste algoritme zou zijn dat beter scoorde dan mensen op een gezichtsherkenningsbenchmark. Het algoritme wist een indrukwekkende 98,5 procent van gezichten te herkennen, tegenover de 97,5 procent die mensen konden herkennen (Lu en Tang, 2014). Met hulp van al dit soort algoritmen zijn robots dus in staat om nog beter dan mensen gezichten te herkennen en identificeren. Robots die in staat zijn om mensen en hun emoties te herkennen kunnen zo worden ingezet bij afhankelijke patiënten, en bijvoorbeeld pijn of verdriet herkennen, en deze informatie vervolgens doorgeven of verzamelen. Een succesvolle demonstratie van deze techniek in twee verzorgingstehuizen is geleverd door Swangnetr en Kaber (2013), die lieten zien dat zorgrobots die medicijnen afleveren nauwkeurig emoties konden identificeren in deze populatie.

Ook emotieherkenning in stemgeluid kan een manier zijn om de interactie tussen mens en robot te verbeteren. Net zoals een kind kan leren dat iets niet mag door het 'boos' toe te spreken, of juist wel door het aanmoedigend toe te spreken, kunnen robots iets leren door classificatie-algoritmen te trainen op de prosodie van een menselijke stem (Breazeal en Aryananda, 2002).

ROBOMANTIEK Het lijkt misschien een vreemd idee, intimiteit met een robot. Waarom zou iemand dat willen? Er zijn genoeg aparte verhalen bekend over ongebruikelijke relaties, van een Amerikaanse vrouw die trouwt met de Eiffeltoren tot een Japanse man die getrouwd is met zijn favoriete karakter uit een Nintendospel. Ook is er een kleine, maar groeiende groep mannen die relaties hebben met sekspoppen (of, zoals het bedrijf ze meer beschrijvend noemt, *love dolls*). Er gaan zelfs stemmen op die beweren dat de seksindustrie de drijvende kracht zal worden achter snelle ontwikkelingen in kunstmatige intelligentie, waarbij sekspoppen uitgerust zullen worden met complexe KI, om zo ook interessante

gesprekken te kunnen voeren. Deze ontwikkelingen zouden, bovenop het feit dat ze technologisch uitermate interessant zijn, ook kunnen bijdragen aan het oplossen van sociale problematiek.

De Zweedse tv-serie *Real Humans* gunt ons een blik in een toekomst waar de robotica sociaal beter acceptabele oplossingen met betrekking tot prostitutie creëert. Levy (2007) denkt dat de robotprostituee (m/v) verschillende voordelen biedt boven de menselijke variant. Hoewel de cijfers behoorlijk fluctueren, heeft ongeveer 14% van alle Nederlandse volwassen mannen weleens betaald voor seks (Francoeur, 2001). Het lijkt er dus op dat er behoorlijke vraag is naar seks tegen betaling, maar er zijn ook de nodige ethische bezwaren. Een reden om seksrobots te verkiezen boven menselijke prostituees is het stigma dat verbonden is aan prostitueebezoek. Een meerderheid van prostitueebezoekers geeft aan het bezoek verborgen te houden voor vrienden en familie. Mogelijk wordt dit stigma minder wanneer levenloze robots worden gebruikt in plaats van mensen, hoewel het nog lastig te voorspellen is of en wat voor stigma verbonden zal zijn aan seks tussen mens en robot. Daarbij wijzen tegenstanders van prostitutie op de slechte omstandigheden en criminele activiteiten die er vaak mee gepaard gaan: mensenhandel, harddruggebruik en seksueel overdraagbare aandoeningen worden vaak in één adem genoemd met de seksindustrie. Seksrobots zouden hier een groot verschil kunnen maken.

Maar met snelle vooruitgang in KI is seks hebben met een robot nog maar het topje van de ijsberg. Lange-afstandsrelaties tussen mensen die elkaar nog nooit hebben ontmoet zijn niet ongewoon (zie het tv-programma *Catfish*), en kunnen maanden, zo niet jaren duren. Het belang van goede gesprekken lijkt dus ook van behoorlijk belang. Zodra KI éindelijk de Turing-test doorstaat, zouden we dan misschien ook een duurzame, romantische relatie kunnen hebben met een robot? En waarom zouden we dat willen? Wij geloven dat de robot, op lange termijn, als romantische partner uitkomst kan bieden voor mensen die om wat voor reden dan ook geen partner kunnen krijgen. Ook blijkt uit onderzoek dat single-zijn niet bepaald gezond is. Een meta-analyse laat zien dat alleenstaande mannen acht tot zeventien jaar eerder sterven dan getrouwde mannen, en alleenstaande vrouwen zeven tot vijftien jaar eerder dan getrouwde vrouwen (Roelfs et al., 2011). Of robots een verschil kunnen maken in deze cijfers valt nog te bezien, maar het is niet ondenkbaar dat ze een rol kunnen vervullen bij het bestrijden van eenzaamheid.

Wanneer een robot sterk op een mens lijkt maar er nog wel van te onderscheiden is, roept dit een gevoel van afkeer op bij mensen

WAT STAAT ONS TE WACHTEN? We kunnen concluderen dat robots steeds menselijker worden, zowel wat betreft uiterlijk als gedrag. Wat de grote volgende stap zal zijn in de kunstmatige intelligentie kan alleen de tijd uitwijzen, maar we kunnen wel een *educated guess* maken op basis van huidige ontwikkelingen.

Supercomputers als Watson, zoals elders in dit nummer door Frank van der Velde besproken, bieden enorme rekenkracht. Gekoppeld aan slimme algoritmen kunnen mensen ermee worden verslagen met schaken, worden er nieuwe medicijnen mee ontwikkeld en kan er zelfs *Jeopardy* mee worden gespeeld. Maar in de cognitieve wetenschap zijn er steeds meer mensen van overtuigd dat niet de enorme rekenkracht van ons brein, maar haar *architectuur* ons 'menselijk' maakt. Relatief simpele elementen, neuronen, die veel onderlinge verbindingen hebben en parallel kunnen werken zorgen blijkbaar voor de rijke ervaringen en complexe gedragingen die mensen hebben.

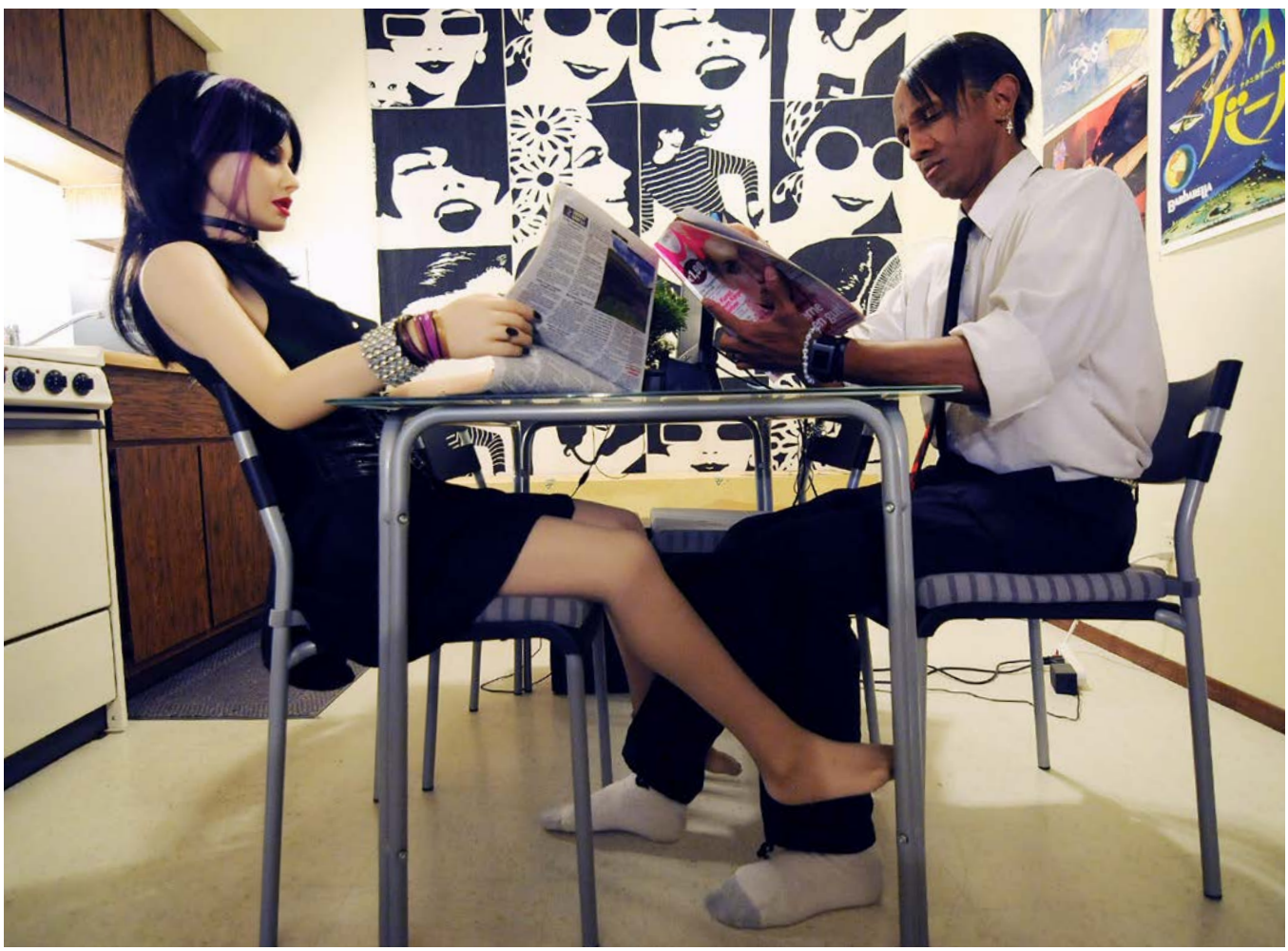
Dit idee is in conceptuele vorm al in gebruik in een hoop KI-toepassingen, maar het Blue Brain Project, geleid door de École Polytechnique Fédérale de Lausanne gaat nog een stap verder. In de zoektocht naar *ware* kunstmatige intelligentie willen zij een nauwkeurige simulatie maken van een mensenbrein op neuronaal niveau. Tenslotte, als het gedrag van mensen voortkomt uit onze hersenen, zou een digitale kopie van deze hersenen dan niet even slim, grappig, en gevoelig zijn als wijzelf?

Afgezien van de vraag of dat inderdaad zo is, roept het aardig wat vraagstukken op. Wat als we ooit in staat zijn om zo'n simulatie in een levensechte robot te draaien? Het lijkt moeilijk voor te stellen dat zo'n robot niet precies dezelfde

emoties – en misschien zelfs romantische gevoelens – bij ons oproept als een mens. Het is in dat geval noodzakelijk om ook stil te staan bij de ethische dilemma's die zo'n conclusie met zich meebrengt. Wanneer aan een robotische partner dezelfde emotiebeleving (en eigenlijk álle qualia) toe te schrijven valt, kunnen de voordelen die een robotpartner in eerste instantie lijkt te hebben wel eens behoorlijk tegenvallen. Bij het hebben van een vrije wil snijdt het mes natuurlijk aan twee kanten. Zo wordt het lastig te verdedigen dat we een romantische relatie zouden willen met een robot met een eigen wil en gevoelens, die eigenlijk liever een andere partner zou hebben. De groep mensen die robots willen gebruiken om de leegte te vullen die is ontstaan door het ontbreken van menselijk contact, zouden er wel eens achter kunnen komen dat robots zich eigenlijk ook niet zo tot ze aangetrokken voelen.

Op het moment dat we gaan twijfelen aan de innerlijke belevingswereld van robots moet ook kritisch gekeken worden naar de rol die robots in onze maatschappij moeten vervullen. Als we niet langer een meester-slaafverhouding hebben met robots, maar ze een volwaardige rol hebben in de maatschappij, wat is dan het nut van zulke robots behalve een *tour de force* van de cognitiewetenschap? Willen we sociale robots slechts *gebruiken* – voor romantische of seksuele relaties, of in andere contexten – dan is het van belang dat ze zich menselijk gedragen zónder de gevoelens te ervaren die daarbij horen, ze zouden een soort *filosofische zombies* moeten zijn.

Of het nu tot stand komt door simulatie van het brein op neuronaal niveau, of door symbolische KI, feit blijft dat robots meer op mensen gaan lijken, met complexer gedrag en menselijke lichamen. Om ons voor te bereiden op een



samenleving waar robots en mensen op grote schaal met elkaar samenwerken hebben we een sterk onderzoeksveld van mens-robotinteractie (HRI) nodig, dat in samenwerking met de cognitieve wetenschappen deze ontwikkeling in goede banen kan leiden. Doen we dit niet, dan moeten we niet klagen als over twintig jaar tehuizen vol zitten met verwaarloosde, depressieve robots. Maar ach, dat biedt tenminste meer werk voor psychologen.

OVER DE AUTEURS

Roy de Kleijn (1982) is als promovendus werkzaam op de afdeling Cognitieve Psychologie van Universiteit Leiden. Zijn promotie-onderzoek behelst het snijvlak van de cognitieve wetenschap en robotica. E-mail: kleijnrde@fsw.leidenuniv.nl.

Bernhard Hommel (1958) is hoogleraar Algemene Psychologie aan Universiteit Leiden. Zijn onderzoek richt zich op perceptie, aandacht, en actiecontrole. Tevens is hij oprichter van het Leiden Institute for Brain and Cognition.

Literatuur

Banks, M. R., Willoughby, L. M., & Banks, W. A. (2008). Animal-assisted therapy and loneliness in nursing homes: Use of robotic versus living dogs. *Journal of the American Medical Directors Association*, 9, 173-177.

Braitenberg, V. (1984). *Vehicles – Experiments in synthetic psychology*. MIT Press, Cambridge, MA.

Breazeal, C. (2003). Toward sociable robots. *Robotics and Autonomous Systems*, 42, 167-175.

Breazeal, C. & Aryananda, L. (2002). Recognition of affective communicative intent in robot-directed speech. *Autonomous Robots*, 12, 83-104.

Buss, D. (1994). *The evolution of desire: Strategies of human mating*. Basic Books, New York.

Costescu, C. A., Vanderborght, B., & David, D. O. (2014). The effects of robot-enhanced psychotherapy: A meta-analysis. *Review Of General Psychology*, 18, 127-136.

Francoeur, R. T. (2001). *The international encyclopedia of sexuality*. Continuum Publishing, New York, NY.

Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, 35, 1674-1684.

Gray, H., Gray, K., & Wegner, D. (2007). Dimensions of mind perception. *Science*, 315, 619.

Groom, V., Nass, C., Chen, T., Nielsen, A., Scarborough, J. K., & Robles, E. (2009). Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human-Computer Studies*, 67, 842-849.

Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243-259.

Kozima, H., Nakagawa, C., & Yasuda, Y. (2007). Children-robot interaction: a pilot study in autism therapy. In von Hofsten, C. & Rosander, K., editors, *From Action to Cognition*, volume 164 of *Progress in Brain Research*, pages 385-400. Elsevier.

Levy, D. (2007). Robot prostitutes as alternatives to human sex workers. In *Proceedings of the 2007 IEEE International Conference on Robotics and Automation*, Rome, Italy.

Lu, C. & Tang, X. (2014). Surpassing human-level face verification performance on LFW with GaussianFace. *CoRR*, abs/1404.3840.

Mitchell, W. J., Szerszen, Sr., K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-PERCEPTION*, 2, 10-12.

Mori, M. (1970). The uncanny valley. *Energy*, 7, 33-35.

Oberman, L. M., McCleery, J. P., Ramachandran, V. S., & Pineda, J. A. (2007). EEG evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots. *Neurocomputing*, 70, 2194-2203.

Pierno, A. C., Mari, M., Lusher, D., & Castiello, U. (2008). Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia*, 46, 448-454.

Ritchie, G., Manurung, R., Pain, H., Waller, A., Black, R., & O'Mara, D. (2007). A practical application of computational humour. In *Proceedings of the 4th International Joint Workshop on Computational Creativity*, pages 91-98, London, UK.

Roelfs, D. J., Shor, E., Kalish, R., & Yogev, T. (2011). The rising relative risk of mortality for singles: Meta-analysis and meta-regression. *American Journal of Epidemiology*, 174, 379-389.

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*, 5, 17-34.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, 7, 413-422.

Swangnetr, M. & Kaber, D. (2013). Emotional state classification in patient-robot interaction using wavelet analysis and statistics-based feature selection. *IEEE Transactions on Human-Machine Systems*, 43, 63-75.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460.

Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception And Performance*, 22, 892-903.

van Es, L. (2015). Empathic response to cute and scary robots. Master's thesis, Leiden University.

Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages I-511-I-518.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? the stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5, 219-232.

Weinberg, G., Godfrey, M., Rae, A., & Rhoads, J. (2007). A real-time genetic algorithm in human-robot musical improvisation. In *Computer Music Modeling and Retrieval. Sense of Sounds, 4th International Symposium, CMMR 2007, Copenhagen, Denmark, August 27-31, 2007. Revised Papers*, pages 351-359.