

Denken over denken over denken: logica en sociale cognitie

*Rineke Verbrugge*¹

Mijnheer de Rector Magnificus, dames en heren,

Ik denk dat u denkt dat ik denk dat ik de betekenis van de titel van deze oratie het best aan de hand van een voorbeeld kan uitleggen. Bij deze dus.

Hogere-orde sociale cognitie

Wat maakt ons mensen nu eigenlijk menselijk? Dat is een van de belangrijkste vragen in de wetenschap. Het onderzoek waarover ik u vandaag vertel, draagt bij aan een antwoord op die vraag. Denken over denken over denken zou weleens een belangrijke onderscheidende eigenschap van mensen kunnen zijn.

Deze oratie gaat over sociale cognitie, in het bijzonder over de ‘theory of mind’ van mensen. Wat is dat nu precies? Mensen redeneren over wat anderen weten, geloven en willen, om zo het gedrag van anderen te begrijpen, en soms zelfs om het gedrag van anderen te voorspellen.

Ik zal in deze oratie speciaal aandacht geven aan *hogere-orde* sociale cognitie: wij denken over andermans denken over ons eigen denken, of we denken over andermans denken over het denken van *nog* weer anderen. Zoals ik bijvoorbeeld dacht over uw denken over mijn denken over het voorbeeld in het begin. Zulk denken kan doorgaan tot in oneindig veel lagen. Voor de informatici en wiskundigen onder u: mensen lezen recursief elkaars gedachten, elkaars mentale toestanden, zo goed en zo kwaad als dat gaat.

Wij volwassenen kunnen hogere-orde sociaal redeneren, ten minste, in principe. In deze oratie wil ik een paar intrigerende vragen over deze vorm van sociale cognitie met u bespreken: zijn volwassenen de enigen die kunnen denken over denken over denken? Of kunnen kinderen het ook? En dieren dan? Is het nodig om te overleven? Hoe werkt hogere-orde sociale cognitie nu

precies? Lijkt het alleen zo ingewikkeld, of is het dat ook echt? Kunnen mensen in het dagelijks leven net zo goed denken over denken over denken als logici dat kunnen op papier? En kun je mensen helpen er beter in te worden?

Eerst schets ik drie intrigerende resultaten over sociale cognitie uit het verleden. Uit de biologie, de psychologie en de logica. Daarna komen het heden en de toekomst aan bod. Welke spannende vragen zijn er nog over en met welke methoden kun je die beantwoorden?

Echt en vermeend teamwork bij mensen en mensapen

Biologen hebben onderzocht of mensapen ook theory of mind hebben. Denken ze over elkaars denken? Mensen en chimpansees zijn evolutionair gezien een soort neven en nichten. Toch is er een groot verschil: mensen kunnen *echt* als team met elkaar samenwerken, terwijl chimpansees dat alleen *lijken* te doen. Bij als team samenwerken hoort een *collectieve intentie*, met andere woorden: een gezamenlijk doel, waarvan alle teamleden op de hoogte zijn. Daarop gebaseerd is er een *gezamenlijk plan* waarin de teamleden ieder hun eigen rol hebben. Als teamlid let je erop of de anderen hun rol goed vervullen en of ze misschien wat extra hulp kunnen gebruiken. Denk bijvoorbeeld aan een strijkkwartet waarin iedereen zich op de anderen afstemt om samen mooie muziek te maken. Of denk aan het verfijnde teamwork dat nodig is als je met zijn tweeën een grote, zware kast een trap op moet sjuwen².

Soms *lijken* chimpansees wel als team samen te werken en dus over hun gezamenlijke doel en elkaars plannen na te denken, bijvoorbeeld als ze met zijn allen tegelijk achter een prooi aangaan. De leider achtervolgt de prooi in de bomen, terwijl de anderen de vluchtwegen afsluiten. Maar in werkelijkheid zijn ze in zo'n situatie alleen maar met hun eigen doel bezig. Ze werken niet *echt* samen, zo blijkt uit de experimenten van Michael Tomasello en zijn collega's in Leipzig. Zij toonden aan dat chimpansees hun acties wel kunnen synchroniseren met die van een ander om een probleem op te lossen. Maar als die ander plotseling ophoudt met zijn aandeel in het werk, doen de chimpansees niets om hun partner weer bij het probleem te betrekken. Er is dus geen sprake van een collectieve intentie, en niet van *echte* samenwerking³. We kunnen dus

concluderen dat chimpansees niet de hogere-orde sociale cognitie hebben die nodig is voor teamwork.

Klassieke experimenten met kinderen over eerste-orde sociale cognitie

Een tweede intrigerende vraag is of kinderen ook denken over denken over denken.

Ontwikkelingspsychologen hebben ontdekt dat deze vaardigheid zich stapsgewijs ontwikkelt.

Een klassiek experiment waarmee je kunt onderzoeken of een kind correct over andermans mentale toestanden redeneert, heet de *false belief task*. Dit experiment gaat over situaties waarin iemand anders een incorrecte overtuiging heeft, terwijl het kind zelf wél weet wat de waarheid is. Ik neem nu het klassieke experiment van Wimmer en Perner met u door⁴. Een onderzoeker vertelt aan een kind een verhaal. Hij doet dat met behulp van twee poppen - Rosa en Tom -, een doos, een vaas en een knikker.

“Op plaatje 1 ziet u Rosa en Tom, en een doos, en een vaas. Nu heeft Rosa een knikker, en ze stopt haar knikker in de doos, op plaatje 2. Dan gaat Rosa buiten spelen, zoals u ziet op plaatje 3. Maar wat gebeurt er terwijl Rosa weg is? Tom gaat naar de doos, hij pakt Rosa’s knikker eruit, op plaatje 4, en doet hem in de vaas, dat ziet u op plaatje 5. Vervolgens gaat Tom ook buiten spelen, op een andere plek dan Rosa. Even later, en dat ziet u op plaatje 6, komt Rosa terug in de kamer.”

De onderzoeker vraagt nu aan het kind: “Waar gaat Rosa als eerste haar knikker zoeken?” Het grappige is: bijna alle kinderen van drie jaar beantwoorden deze vraag fout. Ze zeggen dat Rosa de knikker in de vaas gaat zoeken, waar Tom hem buiten haar gezichtsveld heeft verstopt.

Bijzonder is dat kinderen van vier jaar een sprong in inzicht maken. Zij beantwoorden de vraag over de knikker bijna allemaal goed. Zij zeggen dat Rosa eerst in de doos zal zoeken, omdat ze de knikker daar zelf in heeft gedaan. Natuurlijk neemt Rosa aan dat hij daar is gebleven. Tussen drie en vier jaar gebeurt er dus iets bijzonders: opeens hebben kinderen door dat een ander kind wel eens iets anders kan denken dan zichzelf⁵.

Als u zelf kinderen hebt, of u bent oom of tante, dan hebt u vast iets van deze bijzondere omslag rond de vierde verjaardag van een kind gemerkt. Een kind van drie dat verstoppertje speelt, roept luidkeels vanachter het gordijn: “Papa, je kunt nu gaan zoeken!” Zo’n jong kind kan moeilijk informatie achterhouden, ook als dat haar zou helpen om te winnen bij een spelletje. Kinderen van drie jaar oud begrijpen ook nog niet dat hun ouders ongerust zijn als ze opeens verdwijnen in een winkelcentrum. Ze hebben niet de mentale capaciteit om zich te realiseren dat hun ouders niet weten waar ze zijn.

Tussen dit moment en de volwassenheid leren mensen nog heel wat bij aan sociale intelligentie. Rond hun zevende leren kinderen omgaan met tweede-orde sociale cognitie. Ze begrijpen dan bijvoorbeeld wat er aan de hand is als Jans moeder niet weet dat Jan stiekem al weet wat hij morgen voor zijn verjaardag krijgt⁶.

Klassieke resultaten over sociaal redeneren bij volwassenen

Nog oudere kinderen begrijpen de situatie waarin iedereen in het gezin een bepaald geheim eigenlijk wel weet, maar ze weten dat niet van elkaar. Er is dan in vaktermen nog geen *common knowledge* van het geheim. Als je er wel met zijn allen over spreekt, ontstaat er juist wel *common knowledge*; iedereen weet dan dat iedereen het weet, en ook *dat* weet weer iedereen, enzovoorts, tot in het oneindige⁷. Begrip van zulke subtiele verschillen vergroot de sociale intelligentie van kinderen. Dit maakt het hen mogelijk om steeds verfijnder rekening met anderen te houden, maar ook om anderen steeds slimmer te manipuleren.

Pas als mensen volwassen én verstandig zijn, weten ze precies welke informatie ze op welk moment aan wie moeten melden. Stel, je gaat een huis kopen. Daar komt vaak tweede-orde sociaal redeneren bij te pas. Heb je bijvoorbeeld via via gehoord dat de verkoper al een ander huis heeft gekocht en daardoor haast heeft met de verkoop, dan is het verstandig om hem niet te laten weten dat je het weet.

Echt goed hogere-orde redeneren is vreselijk complex. Derde-orde redeneringen komen we in de gewone wereld nauwelijks tegen, behalve dan misschien bij experts zoals beroepsonderhandelaars of meester-bridgespelers, en onder logici natuurlijk, hoewel het de vraag is of die in de gewone wereld leven.

Uit experimenteel onderzoek blijkt dat bijna iedereen derde-orde sociale cognitie te moeilijk vindt: je hoofd gaat ervan tollen. Maar ook valt het mensen niet mee om in ongewone situaties tweede- en zelfs eerste-orde sociale cognitie foutloos en soepel toe te passen⁸. Gelukkig levert dat in het dagelijks leven niet altijd problemen op. Als je de ander vertrouwt, maar je kunt zijn gedachten niet raden, dan kun je er altijd nog naar vragen, of misschien corrigeert hij jou uit zichzelf als jij onterecht dacht dat hij wel wist wat jij wilde. Dan is er ook geen man overboord.

Ideaal sociaal redeneren met epistemische logica

U hebt nu een beeld van enkele belangrijke inzichten uit de biologie en de psychologie over hogere-orde sociaal redeneren. Maar hoe verloopt sociaal redeneren vanuit het gezichtspunt van de logica? Ik vraag u nu zich te verdiepen in een raadsel waarvoor derde-orde sociaal redeneren vereist is⁹.

Stelt u zich voor dat Anja en Bart perfecte logici zijn. Dat weten ze ook van elkaar, want ze kennen elkaar heel goed. Beiden krijgen een sticker met een natuurlijk getal op hun voorhoofd van de spelleidster. Hun eigen getal kunnen Anja en Bart dus niet zien, maar dat van de ander wel. De spelleidster zegt erbij: “Ik heb jullie elk een natuurlijk getal op je voorhoofd geplakt. Dat kan dus zijn 0, 1, 2, 3, enzovoorts. De twee getallen verschillen precies 1 van elkaar”.

De één heeft dus het getal n op het voorhoofd, en de ander het getal $n+1$. Bijvoorbeeld, Anja heeft 8 en Bart heeft 9, maar dat kunt u meteen weer vergeten. Ik nodig u graag uit om met Anja en Bart mee te denken.

Nu stelt de spelleidster de eerste vraag: “Als je weet welk getal je hebt, stap dan naar voren.” Noch Anja, noch Bart verroert een vin. Daarna stelt de spelleidster de tweede vraag: “Als je weet

welk getal je hebt, stap dan naar voren.” Weer blijven Anja en Bart allebei onbeweeglijk. Voor de derde keer stelt de spelleidster de vraag: “Als je weet welk getal je hebt, stap dan naar voren.” Op dat moment stapt Anja naar voren en noemt, correct, het getal dat op haar voorhoofd staat.

Als u inderdaad, net als Anja en Bart, perfect kunt redeneren, weet u misschien welk natuurlijk getal Anja heeft genoemd. Ik geef u een moment de tijd... Het getal 3 is het juiste antwoord. Anja moet 3 op haar voorhoofd hebben staan, en Bart 2. Hoe weet ik dat nu? Tijd voor een minicursus recursief gedachtelezen, ofwel epistemische logica, dat is de logica van kennis.

In de epistemische logica kunnen we verschillende mogelijke situaties weergeven in een zogenaamd mogelijke-werelden-model. Een model van de beginsituatie van de puzzel staat op de dia. Zoals u ziet zijn er in eerste instantie oneindig veel mogelijke combinaties van getallen, zowel van de vorm (A_n, B_{n+1}) als (B_n, A_{n+1}) . Elke mogelijke combinatie noemen we een ‘mogelijke wereld’.

Als Anja twee situaties op grond van haar informatie niet van elkaar kan onderscheiden, tekenen we tussen die twee mogelijke werelden een rode verbinding. Dus bijvoorbeeld, stel Anja ziet een 1 bij Bart, dus B_1 , dan kan de situatie voor haar net zo goed (A_0, B_1) als (A_2, B_1) zijn. Ze kan immers het getal op haar eigen voorhoofd niet zien. Daarom staat er een rode verbinding tussen de werelden (A_0, B_1) en (A_2, B_1) : ze moet 1 kleiner of 1 groter dan Bart hebben. Een ander voorbeeld: stel dat Bart een 3 bij Anja ziet, A_3 , dan kan de situatie voor hem net zo goed (A_3, B_2) als (A_3, B_4) zijn. Daarom staat er tussen de werelden (A_3, B_2) en (A_3, B_4) een blauwe verbinding.

Terug naar de puzzel. De spelleidster stelt de eerste vraag: “Wie het weet, stapt naar voren.” Niemand doet iets. Wat kunnen wij als toeschouwers nu concluderen? Wij kunnen nu weten dat de twee meest linkse situaties wegvallen. Immers, als (A_0, B_1) de werkelijke situatie was geweest, had Bart een 0 gezien bij Anja en had hij meteen geweten dat hij zelf een 1 had – hij redeneert immers perfect. En spiegelbeeldig, als (A_1, B_0) de werkelijke situatie was geweest, had Anja de juiste conclusie getrokken uit de 0 op Bart’s voorhoofd. Onze redeneringen kunnen Anja en Bart ook prima zelf uitvoeren, en dat weten ze ook van elkaar, het is zelfs common knowledge. Dus we blijven over met het volgende model.

En dan stelt de spelleidster de vraag voor de tweede keer: “Wie het weet, stapt naar voren.” En weer gebeurt er niets. Wat kunnen we daar nu uit afleiden? De twee situaties die nu het meest links staan, vallen weg. Immers, als (A_2, B_1) de werkelijke situatie was, had Anja een 1 bij Bart gezien. Dan had ze na de eerste vraag uitgesloten dat ze zelf een 0 had. Vervolgens had zij met haar perfecte redeneerkunst het antwoord geweten. En weer, symmetrisch, als (A_1, B_2) de werkelijke situatie was geweest, dan zou Bart nu het antwoord hebben geweten na de tweede vraag van de spelleidster. Deze redeneringen kunnen Anja en Bart net als wij prima uitvoeren en dat weten ze ook van elkaar. Dus we knippen beide mogelijkheden weg en gaan over op het volgende model.

Nu stelt de spelleidster haar vraag voor de derde keer en Anja stapt naar voren. Zij weet het nu. Weten wij het daardoor nu ook? In welk van de mogelijke werelden kan Anja onderscheiden wat de echte situatie is? Inderdaad, in de wereld linksonder, (A_3, B_2) , ziet Anja een 2 bij Bart en heeft ze al uitgesloten dat ze zelf een 1 heeft, anders had Bart het immers al eerder geweten. Dus heeft ze zelf een 3 en het spel is uit¹⁰. In de volgende dia ziet u de redenering van Anja en haar gedachten over de mogelijke gedachten van Bart over haar redenering nog eens in één plaatje samengevat.

Wat blijkt uit deze dia's is dat de mogelijke-werelden-modellen uit de epistemische logica ons kunnen helpen bij hogere-orde redeneren. Een complexe redenering kunnen we zo onderverdelen in verschillende stappen die elk voor zich veel minder complex zijn. Lastig is dat logici op deze manier de mens idealiseren. Zij construeren de modellen vanuit de aanname dat iedereen altijd perfect logisch redeneert. Bovendien gaan zij ervan uit dat die perfecte rationaliteit common knowledge is. Misschien heeft u net gemerkt dat dit wel erg idealistisch is.

We hebben nu dus gezien dat dieren niet echt denken over elkaars intenties. We hebben ook gezien dat kinderen stapsgewijs leren redeneren over andermans mentale toestanden, en we hebben gezien dat logici het met behulp van modellen erg goed kunnen. De stand van zaken in het onderzoeksgebied heb ik hiermee in grote lijnen geschetst. Nu richten we onze blik op de toekomst. Ik schets voor u welke thema's mijn onderzoeksgroep in de komende jaren rondom het VICI-programma 'Cognitive Systems in Interaction' bestudeert. De eerste intrigerende vraag luidt: hoe zijn eerste- en hogere-orde sociale cognitie in de evolutie nu eigenlijk ontstaan?

De evolutie van eerste-orde sociale cognitie

Daarvoor kijken we eerst naar het dierenrijk. Beroemde biologen zoals Michael Tomasello en Nicola Clayton stellen dat sommige sociaal slimme dieren – mensapen en, verrassend genoeg, kraaien – redeneren over elkaars kennis. Clayton en haar groep in Cambridge laten zien dat kraaiachtigen heel slim hun voedsel beschermen. Ze kijken erg goed uit als ze een lekkere worm verstoppen voor later: heeft een collega-kraai hen gezien? Dan gaan ze de worm stiekem ergens anders verstoppen, als de concurrent zijn hielen gelicht heeft¹¹.

Op het eerste gezicht lijkt het alsof dit gedrag bewijst dat de kraaien gebruik maken van eerste-orde sociale cognitie. “Als hij niet weet dat ik de worm ergens verstop, kan ik de worm daar veilig laten liggen. Als mijn collega-kraai wel weet waar ik de worm verstop, moet ik die worm opnieuw elders verstoppen.”

Kan dit gedrag misschien anders verklaard worden dan door eerste-orde sociale cognitie te veronderstellen? De meningen hierover zijn zeer verdeeld. Het zou ook kunnen dat de vogels door hun ervaring eenvoudige gedragsregels leren over veilige verstopplekken.

Daar komt geen redenering over andermans kennis of plannen in voor. Elske van der Vaart heeft kort geleden aangetoond dat een computerkraai met dergelijke eenvoudige gedragsregels precies hetzelfde verstopgedrag vertoont als echte kraaien¹². Dit toont aan dat een andere verklaring dan eerste-orde sociale cognitie bij kraaien mogelijk is. Nieuwe experimenten en computermodellen zijn nodig om deze controverse te beslechten.

Hogere-orde sociale cognitie bij onze voorouders

Als we kijken naar hogere-orde sociale cognitie is er een heel duidelijke, veel minder controversiële sprong in de evolutie van de mens: geen enkel dier vertoont hogere-orde sociale cognitie, de mens wel. Die sprong tussen dieren en mensen is intrigerend. Helaas kunnen we onze voorouders, zoals de Neanderthalers, die rond één miljoen jaar geleden de aarde bevolkten, niet

meer direct observeren in gedragsexperimenten. Daarom namen onderzoekers tot nu toe vaak genoegen met gevolgtrekkingen gebaseerd op archeologische evidentie. De resten uit de prehistorie, zoals wapens en tekeningen, dienden als aanwijzing voor bijvoorbeeld samenwerking.

Wel kunnen we de Neanderthalers - net als de kraaien - simuleren met cognitieve computermodellen, zoals u op het plaatje ziet; daar zijn ze voorgesteld als blauwe punten met een blikveld. Grote groepen 'virtuele mensen', stellen we door simulatie aan verschillende moeilijke situaties bloot. Zo onderzoeken we welke redeneerregels hun overlevings- en voortplantingskans vergroten. De vraag is dan: zijn virtuele mensen met hogere-orde sociale cognitie evolutionair succesvoller dan anderen met simpele associatieve gedragsregels? En onder welk soort evolutionaire druk gebeurt dat dan? Deze agentgebaseerde simulatiemethode heeft in onderzoek naar taalevolutie al tot overtuigende resultaten geleid¹³.

Nieuw aan ons onderzoek is dat we de *vorm* van de redeneerregels laten evolueren via een soort genetisch programmeren. Iedereen weet dat de genen van een kind ontstaan door een combinatie van de stukjes van de genen van beide ouders. Mutaties zijn daarbij mogelijk. Analoog daaraan ontstaat in onze simulaties een nieuwe virtuele mens door delen van de gedragsregels van de virtuele ouders op een logische manier met elkaar samen te voegen¹⁴.

Hoe werkt hogere-orde sociale cognitie nu precies?

Een tweede vraag in ons onderzoeksprogramma gaat over de volwassene van tegenwoordig. Wij willen weten hoe volwassenen nu precies redeneren over mentale toestanden van anderen.

Waarom vinden ze het vaak moeilijk om tweede-orde sociale cognitie correct en snel toe te passen?

Om volwassenen te bestuderen, ontwikkelden wij een nieuw spel, de knikkerbaan. Stel, u gaat het spel nu spelen. Er zijn twee buizen. Het doel van het spel is dat u de knikker terecht laat komen in een buis met een zo donkeroranje mogelijke knikker. Om te bepalen of de knikker valt in de linkerbuis of doorgaat naar rechts, heeft u controle over de oranje luiken; dat is uw zet.

Er bestaat ook een ingewikkelder versie van spel, met drie buizen. Daarbij bedient u nog steeds de oranje luiken in de eerste zet. Maar nu heeft u een tegenstander, een superintelligent computerprogramma. Als u de knikker in uw zet verder naar rechts laat rollen, heeft uw tegenstander daarna bij de tweede zet controle over de blauwe luiken. U wilt nog steeds dat de knikker valt in een buis met een zo donkeroranje mogelijke knikker. Uw tegenstander wil juist dat de knikker valt in een buis met een zo donkerblauw mogelijke knikker. Wat denkt u dat hij daar tijdens zijn beurt zou doen als de knikker bij de blauwe luiken terecht was gekomen? Inderdaad, hij zou het rechterluik openen. En wat moet u daarom in uw beurt doen?

Of nog ingewikkelder, met vier buizen, en hoogstens drie zetten: eerst u bij de oranje luiken links, dan hij bij de blauwe luiken in het midden, en dan weer u bij de oranje luiken rechts. Wat zou u hier het best kunnen doen bij uw eerste zet bij de linker oranje luiken om uiteindelijk bij een zo donkeroranje mogelijke knikker terecht te komen? De tegenstander weet dat u bij een eventuele derde beurt het meest rechtse oranje luik zou openen. Om dat te voorkomen zou hij bij een eventuele tweede zet juist het linker blauwe luik openen. In die tweede buis ligt een donkerder oranje knikker dan in de eerste buis, dus is het voor u gunstiger de knikker naar rechts te laten gaan, dus nu opent u bij de eerste zet het oranje luik dat daarvoor zorgt.

Met een beetje training bedenken eerstejaars psychologiestudenten bij dit spel snel de beste strategie. Hierbij voorspellen ze dus correct wat de tegenstander over hun eigen plannen denkt. Inderdaad, dit is dus tweede-orde redeneren: denken over het denken van de ander over je eigen denken. Het was heel verrassend voor ons dat de psychologiestudenten dit zo goed konden toepassen¹⁵. Eerdere onderzoeken leidden namelijk tot de conclusie dat volwassenen in een spelsituatie niet uitblinken in tweede-orde redeneren. In de dia ziet u de andere spellen genoemd - een matrixspel en een autowegspel - en de resultaten van het onderzoek hiermee¹⁶. Het frappante is dat alle drie de spellen speltheoretisch gezien volkomen equivalent zijn. Blijkbaar lijkt de knikkerbaansituatie meer op dagelijkse redeneringen van mensen, dan de andere twee spellen. Die zijn abstracter en onnatuurlijker.

Maar hoe redeneren onze eerstejaars psychologiestudenten nu precies? Op dit moment onderzoeken we dat door met een eye-tracker hun oogbewegingen te volgen. Onze resultaten zijn

weer verrassend: proefpersonen volgen namelijk niet precies de ‘achterwaartse inductie’-strategie uit de speltheorie. Ze redeneren niet van rechts naar links terug, dus van de laatste beurt naar de eerste. In werkelijkheid nemen ze eerst even de situatie als geheel in zich op, zoekend naar gemakkelijke manieren om te winnen, en pas daarna redeneren ze zo nodig terug.

Waarom is hogere-orde redeneren nu precies zo moeilijk voor mensen? Onze hypothese is dat de moeilijkheid zit in het opslaan van zogenaamde ‘tussentoestanden’¹⁷. Om dit te testen ontwikkelen we cognitieve computermodellen voor de redeneringen over de knikkerbaan. Op basis daarvan doen we voorspellingen over de reactietijden en de actieve hersengebieden van een proefpersoon. Dan doen we nieuwe experimenten om onze hypothese te testen en uit te vinden waar de moeilijkheid precies zit. Want als we dat weten, kunnen we mensen ook beter ondersteunen bij hun complexe redeneertaken. De logica en speltheorie helpen ons goede modellen te ontwikkelen van mogelijke redeneerstrategieën.

Ondersteuning van mensen en teams bij hogere-orde sociale cognitie

Hoe kunnen we mensen beter ondersteunen met taken die hogere-orde sociale cognitie vereisen? Ik vertel u over de twee manieren die wij hebben gekozen. Ten eerste willen we computersystemen maken die mensen trainen bij het ‘recursief gedachtelezen’. Denkt u hierbij aan een virtuele ‘leraar’, die u kan helpen bij een taak als onderhandelen. Deze leraar zal gebruik maken van nieuwe in het project te ontwikkelen logica’s, die meer cognitief plausibel zijn dan de epistemische logica die ik eerder schetste. Zulke plausibele logica’s zijn niet meer gebaseerd op de vooronderstelling dat iedereen perfect logisch redeneert. Ze houden rekening met verschillende deelnemers die uiteenlopende redeneerstrategieën hanteren¹⁸.

Naast virtuele training, gaan we ook intelligente softwareprogramma’s ontwikkelen die kunnen functioneren als coördinerend en ondersteunend lid van een team. Stel dat er bijvoorbeeld een grote ecologische ramp is gebeurd. Een softwareprogramma, mensen en robots kunnen dan samenwerken als team om de schade te beperken. Deze zeer verschillende deelnemers hebben elk hun sterke kanten. Zo zijn de robots heel geschikt om onder gevaarlijk puin te kruipen, maar

minder om een gewonde te troosten. Het deelnemende softwareprogramma kan hen helpen goed te plannen, met elkaar te onderhandelen over de taakverdeling en elkaar voldoende op de hoogte te houden¹⁹.

Ook in mijn onderzoeksprogramma speelt teamwork tussen heel verschillende partners een belangrijke rol, ook al zult u bij ons voorlopig nog geen robots aan de lunchtafel aantreffen. Onderzoekers uit de kunstmatige intelligentie, de logica, de psychologie en de biologie reiken elkaar de hand en versterken elkaar. Samen streven we naar meer helderheid over hogere-orde sociale cognitie. Ik ervaar deze interdisciplinaire samenwerking in het onderzoeksteam nu al als heel inspirerend.

Het wordt tijd om af te ronden. Achter mij ziet u een deel van de wandschildering “De boom der kennis” van Matthijs Röling en Wout Muller. U heeft in het afgelopen halfuur veel gehoord over kennis, met name de kennis over andermans mentale toestanden. De titel van de wandschildering verwijst natuurlijk naar het scheppingsverhaal, de boom der kennis van goed en kwaad. Daarom rond ik deze oratie vanaf dit spreekgestoelte af met een moraal. We kunnen mensen met behulp van logica en computersystemen helpen om steeds intelligenter te denken over denken over denken. Ik denk dat u misschien denkt dat ik denk dat de wetenschap alle problemen van de wereld *alleen* kan oplossen. Maar dat denk ik zeker niet. Ik hoop van harte dat ons onderzoek uiteindelijk bijdraagt aan goede toepassingen, zoals een doorbraak in vastgelopen internationale onderhandelingen, rondom de klimaatproblematiek of het Midden-Oosten. Maar voor zo’n doorbraak heeft de mensheid veel meer nodig dan intelligentie alleen. Minstens even belangrijk zijn moed, begrip en wijsheid.

Ik heb gezegd.

Dankwoord

Volgens het protocol voor oraties aan de Rijksuniversiteit Groningen is het helaas niet meer gewenst om hardop een dankwoord uit te spreken. Graag wil ik hier toch enige groepen en personen noemen, in het volle besef van de onvolledigheid van de lijst namen. Ik dank allereerst alle toehoorders bij mijn oratie voor hun aanwezigheid: u maakte deze gelegenheid voor mij heel feestelijk.

Dan ga ik eerst terug in de tijd, naar 1993, om mijn promotores hartelijk dank te zeggen voor hun waardevolle leerschool: Anne Troelstra, Dick de Jongh en Albert Visser. Ook aan de andere docenten aan de opleiding Wiskunde aan de Universiteit van Amsterdam denk ik met plezier terug, zoals aan Kees Doets en Frank Veltman.

Sommige van mijn mentoren zijn helaas al overleden, en dan gaan mijn gedachten uit naar Jan Stevens, mijn wiskundeleraar aan de middelbare school, en George Boolos, die mij in 1995 voor een geweldig half jaar naar het MIT in Cambridge, Massachusetts, uitnodigde.

Ik ben blij dat ik vanaf 1997 heb mogen werken te midden van mijn huidige en vroegere collega's en studenten bij de inspirerende werkeenheid Kunstmatige Intelligentie, destijds Technische Cognitiewetenschap. Ik ben iedereen dankbaar, maar noem hier met name: Tinie Alma, Tjeerd Andringa, Egon van Baars, Bart de Boer, Fokie Cnossen, Hans van Ditmarsch, Liesbeth Flobbe, Sujata Ghosh, Margriet Halbersma, Maaïke Harbers, Petra Hendriks, Anja Lobanova, Nancy Lokai, Leendert van Maanen, Ben Meijering, Lisette Mol, Ben Rodenhäuser, Hedderik van Rijn, Lambert Schomaker, Elina Sietsema, Sonja Smets, Jennifer Spenader, Freek Stulp, Niels Taatgen, Wouter Teepe, Elske van der Vaart, Gert van Valkenhoef, Ingrid Veltman, Bart Verheij, Harmen Wassenaar, Esther Wiersinga, Karin Zondervan, Ronald Zwaagstra en Frans Zwarts.

Verder dank ik mijn gewaardeerde collega's in de Faculteit der Wiskunde en Natuurwetenschappen, en dan noem ik met name Piter Dykstra, Charlotte Hemelrijk, Gerard Renardel de Lavalette en Harry Trentelman.

Daarnaast zijn er vruchtbare contacten met de collega's in de Faculteit der Wijsbegeerte, waarvan ik met name noem: Barteld Kooi, Erik Krabbe, Theo Kuipers en Jeanne Peijnenburg.

Misschien gaat er niets boven Groningen, maar ik ervaar de samenwerking met collega's van elders, bijvoorbeeld van het ILLC en het CWI in Amsterdam, van de Universiteit Utrecht en van allerlei instituten in het buitenland, als zeer inspirerend. Ik dank hier al mijn gewaardeerde leermeesters en co-auteurs van elders, met name Johan van Benthem, Frank Dignum, Barbara Dunin-Kępicz, Jan van Eijck, Wiebe van der Hoek en Irene Krämer.

Ik dank al mijn vrienden en vriendinnen van buiten de universitaire wereld, hier met name Marjanne de Jong en Astrid Rijkens.

Ik dank graag mijn ouders Piet en Eef Verbrugge en mijn schoonouders Jan en Mary Baars voor alle steun. Mijn nichtjes Evelien Vink en Rachel, Evita en Loeka Lutters dank ik voor de vele mooie voorbeelden van sociale cognitie in ontwikkeling.

Tot slot ben ik iedere dag verheugd te mogen leven met mijn lieve, grappige en wijze vrouw, Nicole Baars.

Voetnoten

¹ Dit document bevat de tekst van mijn inaugurele rede, uitgesproken op 25 mei 2010, bij de aanvaarding van het ambt van hoogleraar in de Logica en Cognitie. Deze leerstoel is onderdeel van het instituut ALICE (Artificial Intelligence and Cognitive Engineering) bij de Faculteit der Wiskunde en Natuurwetenschappen aan de Rijksuniversiteit Groningen. De oratie en de bijbehorende dia's zijn downloadbaar vanaf <http://www.ai.rug.nl/~rineke>.

² Meer over de formalisering en het belang voor teamwork van gezamenlijke doelen, collectieve intenties, gezamenlijke plannen en collectieve commitments is te vinden in B. Dunin-Kępicz en R. Verbrugge, *Teamwork in Multi-Agent Systems: A Formal Approach*. Wiley and Sons, 2010. Het gaat daar met name om samenwerking tussen 'software agents', maar de theorie is ook toepasbaar op menselijke teams. Voor een meer filosofische aanpak, zie onder meer: M. Bratman, Shared cooperative activity. *Philosophical Review*, **101** (1992) 327-341; R. Tuomela, *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford University Press, Stanford, 1995; M. Gilbert, Shared intentions and personal intentions, *Philosophical Studies* **144** (2009) 1167-1187.

³ Voor een beschrijving en interpretatie van het gezamenlijk jagen door chimpansees op colobusapen, zie C. Boesch, Joint cooperative hunting among wild chimpanzees: Taking natural observations seriously. *Behavioral and Brain Sciences* **28**(5) (2005) 692-693.

Voor een kritische interpretatie van de daarbij voorkomende samenwerking, zie M. Tomasello, *Why We Cooperate*. MIT Press, Cambridge (MA), 2009.

De experimentele resultaten over het gebrek aan echt teamwork bij chimpansees staan beschreven in F. Warneken en M. Tomasello, Altruistic helping in children and young chimpanzees. *Science* **311** (2006) 1301-1303.

⁴ Het klassieke *false belief*-experiment, aan de hand van de zogenaamde 'Maxi'-taak, wordt beschreven in H. Wimmer en J. Perner, Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **13**(1) (1983) 103-128. In de oratie gebruik ik een variatie op een latere versie van de taak, met hoofdpersonen Sally en Tom, weergegeven in M. Doherty, *Theory of Mind: How Children Understand Others' Thoughts and Feelings*. Psychology Press, New York, 2009. Dit overzichtboek bevat ook veel literatuurverwijzingen naar de overvloed aan experimenteel onderzoek betreffende de sprong naar eerste-orde sociale cognitie bij kinderen.

⁵ De precieze leeftijdsgrens voor de *false belief task* is controversieel, zie bijvoorbeeld K.H. Onishi en R. Baillargeon, Do 15-month-old infants understand false beliefs? *Science* **308** (2005) 255-258; en als reactie daarop J. Perner en T. Ruffman, Infants' insight into the mind: how deep? *Science* **308** (2005) 214-216.

⁶ De meeste kinderen van acht jaar brengen een tweede-orde *false belief task* tot een goed einde, zie L. Flobbe, R. Verbrugge, P. Hendriks en I. Krämer, Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information* **17**(4) (2008) 417-442. Zie ook J. Perner en H. Wimmer, "John thinks that Mary thinks that...": Attribution of second-order beliefs by 5- to 10-year old children. *Journal of Experimental Child Psychology* **5** (1985) 125-137.

⁷ Kinderen van rond negen jaar oud begrijpen vaak nog niet het verschil tussen een *belofte* van een ander en het enkel *noemen van een plan* door die ander; soms is dat onbegrip aanleiding tot grote teleurstelling. Experimentele voorbeelden van dit fenomeen zijn te vinden in C.M. Mant en J. Perner, The child's understanding of commitment. *Developmental Psychology* **24** (1988) 343-351. Een interpretatie van deze experimenten in termen van het nog niet begrijpen van *common belief* en *common knowledge* vindt u in R. Verbrugge, Logic and social cognition: The facts matter, and so do computational models. *Journal of Philosophical Logic* **38** (2009) 649-680.

⁸ Volwassenen hebben moeite om niet-coöperatief te communiceren, bijvoorbeeld als ze in een competitief spel moeten redeneren over andermans kennis over hun kennis. Zie de uitkomsten van experimenten met een symmetrische versie van Mastermind in R. Verbrugge en L. Mol, Learning to apply theory of mind. *Journal of Logic, Language and Information* **17**(4) (2008) 489-511. Over problemen met het correct en snel toepassen van eerste-orde sociale cognitie gaat B. Keysar, S. Lin en D. J. Barr, Limits on theory of mind use in adults. *Cognition* **89** (2003) 25-41.

⁹ Een heldere uitleg van de algemene versie van dit *consecutive numbers problem* vindt u in B.P. Kooi, H.P. van Ditmarsch en W. van der Hoek, *Dynamic Epistemic Logic*. Springer, Berlijn, 2007.

¹⁰ Soortgelijke raadsels, waarin de deelnemers meer te weten komen juist doordat ze merken dat de ander 'het niet weet', komen veelvuldig voor in de literatuur over (dynamisch-)epistemische logica. De mogelijk eerste analyse van zo'n raadsel met modale logica is te vinden in J. Groenendijk, M. Stokhof en P. van Emde Boas, The Conway paradox. In: J. Groenendijk, T. Janssen en M. Stokhof (eds.), *Formal Methods in the Study of Language*. Vol. I, Mathematical Centre, Amsterdam, 1998, pp. 87-111.

Voor beschrijvingen in de epistemische logica van het beroemde raadsel van de 'modderige kinderen', zie bijvoorbeeld W. van der Hoek en R. Verbrugge, Epistemic logic: a survey. In: V. Mazalov and L. Petrosjan (eds.), *Game Theory and Applications*, vol. 8, Nova Science Publishers, New York, 2002, pp. 53-94. Zie ook de klassieker R. Fagin, J. Halpern, Y. Moses en M. Vardi, *Reasoning about Knowledge*, MIT Press, Cambridge (MA), 1995.

Het wiskundig complexere 'som- en productraadsel' van Hans Freudenthal wordt dynamisch-epistemisch geanalyseerd in H. van Ditmarsch, J. van Eijck and R. Verbrugge, Publieke werken: breinbrouwsels over sommen en producten. *Nieuw Archief voor Wiskunde*, serie 5, **10**(2) (2009) 126-131. Daarbij wordt gebruik gemaakt van de automatische stellingbewijzer DEMO van Jan van Eijck, waarmee de modellen van opeenvolgende situaties in dergelijke raadsels eenvoudig berekend kunnen worden.

¹¹ N.J. Emery, J. Dally, J & N.S. Clayton. Western scrub-jays (*Aphelocoma californica*) use cognitive strategies to protect their caches from thieving conspecifics. *Animal Cognition*, **7** (2004) 37-43.

¹² Dit model heeft Elske van der Vaart gepresenteerd in haar lezing 'Social cognition in corvids: a computational model', op het internationale symposium *Social cognition in animals: is it as smart as it looks?* Haren, oktober 2009. Het model bouwt voort op onze eerdere computationele cognitieve modellen van het geheugen van kraai-achtigen bij het verstoppen en zoeken van voedsel. Zie E. van der Vaart, C.K. Hemelrijk en R. Verbrugge, Learning where (not) to cache: A cognitive model for corvids. In: N.A. Taatgen en R. van Rijn (eds.) *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, 2009, 2420-2425; zie ook E. van der Vaart,

R. Verbrugge en C.K. Hemelrijk, Memory effects in the Clark's nutcracker: A cognitive model for corvids. In: A. Howes, D. Peebles en R. Cooper (eds.) *Proceedings of the 9th International Conference on Cognitive Modeling*, Manchester, UK, 2009.

¹³ Agent-gebaseerde modellen van de evolutie van taal worden onder meer beschreven in L. Steels, The synthetic modeling of language origins. *Evolution of Communication* **1**(1) (1997) 1-34; voor recent onderzoek zie B. de Boer en W. Zuidema, Multi-agent simulations of the evolution of combinatorial phonology. *Adaptive Behavior* **18**(2) (2010) 141-154.

¹⁴ Een voorstudie voor dergelijk onderzoek is te vinden in E. van der Vaart en R. Verbrugge, Agent-based models for animal cognition: A proposal and a prototype. In: D. Parkes en L. Padgham (eds.), *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2008, pp. 1145-1152. In dat artikel is al wel sprake van de evolutie van de vorm van gedragsregels, maar nog niet van expliciet gerepresenteerde *theory of mind*.

¹⁵ B. Meijering, L. van Maanen, H. van Rijn en R. Verbrugge, The facilitative effect of context on second-order social reasoning. In: *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, te verschijnen in 2010.

¹⁶ Experimenten met het matrixspel worden beschreven in T. Hedden en J. Zhang, What do you think I think you think? *Cognition* **85** (2002) 1-36. Experimenten met het wegspeel, een variant op het *centipede*-spel uit de speltheorie, worden beschreven in L. Flobbe, R. Verbrugge, P. Hendriks en I. Krämer, Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information* **17**(4) (2008) 417-442.

¹⁷ Zie J.P. Borst, N.A. Taatgen en H. van Rijn, The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **36**(2) (2010) 363-382; D.D. Salvucci en N.A. Taatgen, Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review* **115** (2008) 101-130.

¹⁸ Op het gebied van logica's die passen bij cognitieve beperkingen is al interessant onderzoek gedaan, bijvoorbeeld H.P. van Ditmarsch en W.A. Labuschagne, My beliefs about your beliefs – a case study in theory of mind and epistemic logic. *Synthese (Knowledge, Rationality & Action)* **155** (2007) 191-209; K. Stenning en M. van Lambalgen, Logic in the study of psychiatric disorders: executive function and rule-following. *Topoi* **26**(1) (2007) 97-114.

Over de gespannen verhouding tussen de logica en het psychologisch laboratorium vindt u meer in J.F.A.K. van Benthem, Logic and reasoning: Do the facts matter? *Studia Logica* **88**(1) (2008) 67-84; R. Verbrugge, Logic and social cognition: The facts matter, and so do computational models. *Journal of Philosophical Logic* **38** (2009) 649-680.

¹⁹ Een studie van teamwork bij ecologische rampen door een gemengd team met robots, onbemande vliegtuigjes, een menselijke helicopterpiloot en een software agent als teamleider, vindt u in B. Dunin-Kępicz, R. Verbrugge en M. Slizak, TeamLog in action: a case study in teamwork. *Computer Science and Information Systems*, **7** (2), 2010. In dat artikel wordt het niveau van te bereiken kennis aangepast aan de rol van de deelnemers en aan de situatie, maar nog niet aan individuele redeneerstrategieën en cognitieve beperkingen.