

University of Groningen

If you know what I mean

de Weerd, Hermanes Albertus

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

de Weerd, H. A. (2015). *If you know what I mean: agent-based models for understanding the function of higher-order theory of mind*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Nederlandse samenvatting

In het dagelijks leven denken mensen vaak na over wat andere mensen willen, geloven, of van plan zijn. Het op deze manier toekennen van interne mentale toestanden aan anderen wordt *theory of mind* of *mentaliseren* genoemd. Door te mentaliseren kan de lezer van een tekst zich bijvoorbeeld afvragen wat de schrijver bedoeld heeft met een bepaalde zin. Mensen mentaliseren niet alleen door na te denken over wat anderen willen en geloven, maar ook om na te denken over hoe andere mensen mentaliseren. Een schrijver past dit soort *hogere-orde mentaliseren* bijvoorbeeld toe wanneer hij zich afvraagt of de lezer zal begrijpen wat de schrijver bedoelt met een bepaalde zin.

Mensen zijn verrassend goed in hogere-orde mentaliseren. Zo hebben de meeste volwassenen een goed idee wat het betekent als ‘Anke denkt dat Bram weet dat Carolien van plan is een verrassingsfeestje voor hem te organiseren’. Maar hogere-orde mentaliseren lijkt een vaardigheid te zijn die alleen mensen bezitten. Hoewel van diersoorten als chimpansees en verschillende kraaisoorten wordt gedacht dat ze mogelijk gebruik maken van mentaliseren, is het controversieel of deze diersoorten het menselijke niveau van eerste-orde mentaliseren halen. Hogere-orde mentaliseren is in geen enkele andere soort dan de mens aangetoond. Dit roept de vraag op wat er toe heeft geleid dat mensen deze vaardigheid wél bezitten.

In dit proefschrift wordt de vraag omtrent de oorsprong van menselijk hogere-orde mentaliseren onderzocht met behulp van agentmodellen. Deze agentmodellen stellen ons ertoe in staat om te definiëren wat mentaliseren is, om precies te bepalen wanneer een agent gebruik maakt van mentaliseren en in hoeverre een agent door te mentaliseren geholpen is in het behalen van zijn doel. Door het voordeel van hogere-orde mentaliseren in verschillende omgevingen te vergelijken, kunnen we vervolgens bepalen hoe waarschijnlijk het is dat een bepaalde soort omgeving aanleiding heeft gegeven voor de emergentie van hogere-orde mentaliseren.

Verschillende theorieën wijzen naar verschillende omgevingen die de emergentie van hogere-orde mentaliseren zouden kunnen verklaren. In dit proefschrift onderzoek ik drie van deze theorieën met behulp van agentmodellen. Volgens de *Machiavelliaanse-intelligentiehypothese* (Byrne & Whiten, 1988; Whiten & Byrne, 1997) helpt hogere-orde mentaliseren mensen om tegenstanders te slim af te zijn in competitieve situaties. De *Vygotskiaanse-intelligentiehypothese* (Vygotsky, 1978) daarentegen beweert dat mensen voordeel behalen van hogere-orde mentaliseren doordat het helpt om beter en efficiënter samen te werken. Tenslotte beweert de *gemengde-doelenhypothese* (Verbrugge, 2009) dat alleen samenwerking of al-

	Hypotheses		
	Machiavelliaans	Vygotskiaans	Gemengde doelen
Extra voordeel van mentaliseren			
Eerste-orde	++	++	++
Tweede-orde	++	+	++
Derde-orde	+	-	-
Vierde-orde	-	-	+

Tabel 1: Samenvatting van de hoofdbevindingen in dit proefschrift omtrent de Machiavelliaanse-intelligentiehypothese, de Vygotskiaanse-intelligentiehypothese en de gemengde-doelenhypothese. Voor elk van deze hypothesen geeft de tabel het voordeel weer van k de-orde mentaliseren ten opzichte van $(k - 1)$ ste-orde mentaliseren.

leen competitie er niet voor hebben gezorgd dat mensen hogere-orde mentaliseren hebben ontwikkeld, maar dat juist situaties waarin zowel competitie als samenwerking een rol spelen belangrijk zijn. In deze situaties zou hogere-orde mentaliseren helpen om de juiste balans te vinden tussen competitie en samenwerking.

De hoofdbevindingen van dit proefschrift zijn samengevat in Tabel 1, die het voordeel van hogere-orde mentaliseren beschrijft voor de Machiavelliaanse-intelligentiehypothese, de Vygotskiaanse-intelligentiehypothese, en de gemengde-doelenhypothese. In de volgende secties worden deze resultaten in meer detail behandeld.

Machiavelliaanse-intelligentiehypothese

Deel 1 van dit proefschrift (Hoofdstukken 2 en 3) gaat over de Machiavelliaanse-intelligentiehypothese. Volgens deze hypothese zijn individuen die gebruik kunnen maken van hogere ordes mentaliseren in staat om meer tegenstanders te slim af te zijn. We testen deze hypothese door agenten te modelleren die verschillen in hun mentaliseervaardigheden en deze tegen elkaar te laten spelen in verschillende competitieve spellen. Daarbij onderzoeken we onder andere het voordeel van hogere-orde mentaliseren in het spel *steen, papier, schaar*. In dit spel kiezen twee spelers tegelijk één van de drie mogelijke acties ‘steen’, ‘papier’, of ‘schaar’. Wanneer de spelers dezelfde actie kiezen eindigt het spel in gelijkspel. In andere gevallen wint ‘steen’ van ‘schaar’, ‘schaar’ wint van ‘papier’, en ‘papier’ wint van ‘steen’.

Volgens de speltheorie zouden volledig rationele individuen uiteindelijk een strategie hanteren waarbij ze elk van de drie mogelijke acties met dezelfde kans spelen. Voor elke andere strategie s is er namelijk een tegenstrategie s' die wint

van strategie s . Mensen blijken deze rationele strategie echter niet te hanteren. Dit zorgt ervoor dat mentaliseren in dit geval een voordeel zou kunnen opleveren.

Hoofdstuk 2 introduceert het agentmodel voor agenten die verschillen in hun mentaliseervaardigheden. Een nulde-orde mentaliseeragent (ToM_0 agent) kan niet nadenken over wat zijn tegenstander wil of gelooft. In plaats daarvan vertrouwt een ToM_0 agent op eerdere ervaringen in het spel. Wanneer de tegenstander bijvoorbeeld vaak ‘steen’ heeft gespeeld denkt de ToM_0 agent dat ‘papier’ spelen de hoogste kans heeft om te winnen.

Een eerste-orde mentaliseeragent (ToM_1 agent) denkt dat zijn tegenstander mogelijkwerwijs een doel wil behalen in het spel. Door zichzelf te verplaatsen in zijn tegenstander probeert de ToM_1 agent te achterhalen wat hijzelf zou doen als hij de tegenstander zou zijn, en reageert op basis van deze informatie. Een ToM_1 agent die zich verplaatst in zijn tegenstander zou zich bijvoorbeeld kunnen realiseren dat hijzelf vaak ‘steen’ heeft gespeeld en dat de tegenstander daarom zou kunnen denken dat ‘papier’ spelen de hoogste kans heeft om te winnen. Hieruit concludeert de ToM_1 agent dat hij zelf ‘schaar’ zou moeten spelen.

Agenten die dieper kunnen mentaliseren, zoals een k de-orde mentaliseeragent (ToM_k agent), geloven dat hun tegenstander een mentaliseeragent zou kunnen zijn van een lagere orde. Tijdens het spelen van het spel probeert een ToM_k agent te bepalen welke orde mentaliseren het best past bij zijn tegenstander, en gebruikt het bijbehorende model om het toekomstige gedrag van de tegenstander te voorspellen.

In Hoofdstuk 2 onderzoeken we naast drie varianten van steen, schaar, papier ook een spel genaamd *Limited Bidding*, waarin meerdere beslissingen per spel moeten worden gemaakt. In elk van deze vier spellen zien we hetzelfde patroon in de resultaten. Het voordeel van mentaliseren in competitieve situaties blijkt uit het feit dat ToM_1 agenten winnen van ToM_0 tegenstanders en dat ToM_2 agenten winnen van ToM_1 tegenstanders. Daarnaast zien we een patroon van afnemende meeropbrengsten voor hogere-orde mentaliseren. Dat wil zeggen, hoewel ToM_3 agenten winnen van ToM_2 tegenstanders, behalen ze een relatief laag voordeel bovenop het voordeel dat ToM_2 agenten behalen. Nog dieper mentaliseren blijkt nauwelijks nog meer voordeel op te leveren.

De simulaties met agentmodellen laten zien dat er competitieve situaties zijn waarin hogere-orde mentaliseren een voordeel met zich meebrengt. Daarnaast laat Hoofdstuk 3 van dit proefschrift zien dat het gedrag van proefpersonen verrassende overeenkomsten vertoont met hogere-orde mentaliseeragenten in het Mod spel. Het Mod spel is een variant van steen, schaar, papier waarin n spelers tegelijk een getal tussen 1 en 24 kiezen. Spelers krijgen één punt per tegenstander die het getal heeft gekozen dat precies 1 lager is dan het getal dat zij zelf hebben gekozen. Spelers die hebben gekozen voor het getal 1 krijgen een punt per tegenstander die het getal 24 heeft gekozen. Frey & Goldstone (2013) laten zien dat proefpersonen dit spel niet spelen volgens de voorspelling van de speltheorie. Uit Hoofdstuk 3

blijkt bovendien dat hogere-orde mentaliseeragenten het gedrag van proefpersonen beter beschrijven dan nulde-orde en eerste-orde mentaliseeragenten.

Vygotskiaanse-intelligentiehypothese

Volgens de Vygotskiaanse-intelligentiehypothese is niet concurrentie, maar samenwerking de belangrijkste drijfveer voor het gebruik van hogere-orde mentaliseren. Volgens deze hypothese kunnen mensen beter samenwerken met anderen wanneer ze gebruik maken van hogere-orde mentaliseren. Dit komt omdat mensen die mentaliseren kunnen nadenken over de doelen en de bedoelingen van anderen. Hogere-orde mentaliseren is bijvoorbeeld nodig om te begrijpen dat alle leden van een groep weten dat iedereen in de groep hetzelfde doel heeft.

In Deel 2 (Hoofdstuk 4) stellen we de Vygotskiaanse-intelligentiehypothese op de proef in het spel genaamd *Tacit Communication Game* (De Ruiter et al., 2007; Newman-Norlund et al., 2009; Blokpoel et al., 2012). Dit spel wordt gespeeld door twee spelers op een spelbord dat verdeeld is in vakjes. Beide spelers hebben een spelfiguur die ze kunnen verplaatsen over het spelbord. Het doel van het spel is voor beide spelers hetzelfde: beide spelers willen dat de twee spelfiguren naar bepaalde vakjes van het spelbord worden verplaatst. Het probleem is dat de eerste speler (de *Zender*) weet wat deze doelvakjes zijn, terwijl de tweede speler (de *Ontvanger*) dit niet weet.

Het spel wordt gespeeld in twee delen. In het eerste deel beweegt alleen de Zender zijn eigen spelfiguur over het bord. In het tweede deel van het spel mag alleen de Ontvanger haar spelfiguur verplaatsen. Wanneer het spel eindigt horen de Zender en de Ontvanger of beide spelfiguren op de juiste vakjes staan. Als dit niet het geval is krijgt de Ontvanger echter niet te horen welk spelfiguur op de verkeerde plaats staat, en op welke plaats deze figuur wel had moeten staan. De Zender heeft dus twee doelen om te bereiken. Niet alleen wil hij ervoor zorgen dat zijn spelfiguur op het juiste vakje staat aan het einde van zijn beurt, maar de Zender wil ook dat de Ontvanger aan het einde van zijn beurt weet waar haar spelfiguur moet komen te staan. De Ontvanger heeft op haar beurt de taak om de acties van de Zender te interpreteren en te achterhalen waar ze haar spelfiguur moet plaatsen.

Om te onderzoeken of hogere-orde mentaliseren individuen kan helpen om beter samen te werken simuleren we interacties tussen mentaliseeragenten in de Tacit Communication Game. Deze agenten werken op basis van hetzelfde principe als de agenten in de vorige sectie. De ToM_0 Zender en ToM_0 Ontvanger kunnen niet bevatten dat hun partner hetzelfde doel heeft als zij zelf. In plaats daarvan baseren ze hun keuzes puur op eerder geobserveerd gedrag.

Door zichzelf te verplaatsen in de positie van de ander realiseren de ToM_1 Zender en ToM_1 Ontvanger dat de twee spelers hetzelfde doel hebben. Bovendien

kan een ToM_1 Ontvanger onbekend gedrag van de Zender al gedeeltelijk interpreteren. Stel bijvoorbeeld dat Zender door gedrag g te vertonen er voor heeft gezorgd dat Ontvanger haar spelfiguur op het juiste doelvakje v heeft geplaatst. Dat wil zeggen, Zender en Ontvanger hebben doelvakje v al eens succesvol gecommuniceerd. Wanneer dit het geval is, dan is de Ontvanger ervan overtuigd dat onbekend gedrag van de Zender betekent dat haar doelvakje niet vakje v is. Immers, zo redeneert de Ontvanger, wanneer de Zender tot doel had om de Ontvanger te laten kiezen voor vakje v , dan had hij wel gedrag g vertoond. Aangezien hij niet gedrag g vertoont, heeft de Zender aan ander doel voor ogen, en moet de Ontvanger dus niet kiezen voor vakje v .

Simulatieresultaten met mentaliseeragenten laten zien dat mentaliseren agenten helpt om sneller te leren communiceren met elkaar. Vooral de combinatie van een ToM_0 Zender met een ToM_1 Ontvanger is erg efficiënt in het tot stand brengen van samenwerking. De resultaten laten ook zien dat de combinatie van twee agenten die gebruik maken van dezelfde orde van mentaliseren over het algemeen minder goede resultaten behalen dan wanneer de Ontvanger één orde hoger mentaliseert dan de Zender. Wanneer agenten samenwerken kan te veel mentaliseren dus leiden tot slechtere resultaten.

Deze simulaties laten zien dat mentaliseren agenten kan helpen om beter samen te werken. Echter, de resultaten laten ook zien dat mentaliseeragenten anders communiceren dan mensen. Waar menselijke Zenders verrassend veel overeenkomsten vertonen in de manier waarop ze doelvakjes naar de Ontvanger communiceren (zie [De Ruiter et al., 2007](#)), vertoont het gedrag van mentaliseeragenten enorme variatie. Dit betekent dat menselijk gedrag in de Tacit Communication Game niet volledig wordt verklaard door mentaliseren alleen. Aanvullende simulaties laten een interessante rol voor voorkeuren zien. Wanneer Zenders bijvoorbeeld een voorkeur hebben om weinig zetten te maken, maar wel hun eigen spelfiguur het doelvakje van de Ontvanger willen laten bezoeken, is er een extra voordeel voor hogere-orde mentaliseren. Door rekening te houden met de voorkeuren van Zenders kiezen ToM_2 Zenders hun gedrag zodanig dat dit gedrag sneller door ToM_1 Ontvangers wordt begrepen. Daarnaast lijkt het gedrag van mentaliseeragenten met voorkeuren meer op menselijk gedrag.

Gemengde-doelenhypothese

In Deel 3 van dit proefschrift (Hoofdstukken 4, 5 en 6) komt de gemengde-doelenhypothese aan bod. Volgens deze hypothese is het niet alleen concurrentie of alleen samenwerking dat het menselijk vermogen tot hogere-orde mentaliseren verklaart, maar juist situaties waarin zowel concurrentie als samenwerking een rol spelen. Volgens deze theorie zou hogere-orde mentaliseren individuen kunnen helpen om deze twee aspecten beter te balanceren.

Om de gemengde doelen hypothese te testen simuleren we interacties tussen mentaliseeragenten in een onderhandelings spel genaamd *Colored Trails* (Lin et al., 2008; Gal et al., 2010; De Jong et al., 2011; Van Wissen et al., 2012). In *Colored Trails* onderhandelen twee of meer spelers over de verdeling van verschillend gekleurde fiches die de spelers kunnen gebruiken om zich te verplaatsen over een speelbord. Omdat spelers verschillende doelen hebben, verschillen de spelers in de relatieve waarde die ze hechten aan fiches. Zo zou de ene speler vooral rode fiches nodig kunnen hebben om zijn doel te bereiken, terwijl de andere speler liever blauwe fiches heeft. In dit spel is het daarom vaak mogelijk om fiches te ruilen op een manier die voor beide partijen gunstig uitpakt. Alle fiches zijn echter waardevol voor elke speler. Spelers hebben dus doelen die gedeeltelijk overlappen. Hoewel spelers kunnen samenwerken, door fiches zodanig te ruilen op dat ze er beide beter van worden, is het uiteindelijke doel van elke speler om een zo hoog mogelijke score te behalen door zo veel mogelijk fiches te krijgen.

In Deel 3 van dit proefschrift beschouwen we twee manieren van onderhandelingen. Voor beide onderhandelingsvormen simuleren we interacties tussen mentaliseeragenten om te bepalen in hoeverre hogere-orde mentaliseren leidt tot betere individuele resultaten, maar ook tot betere resultaten van de groep. De agenten werken op dezelfde manier als in de vorige secties. Voor een ToM_0 agent is het niet mogelijk om na te denken over het doel van onderhandelingspartners. In plaats daarvan onderhandelt de ToM_0 agent op basis van wat hij zelf wil en wat hij op basis van eerdere ervaringen denkt dat zijn onderhandelingspartner zal gaan accepteren. Een ToM_1 agent realiseert zich daarentegen dat zijn onderhandelingspartner een doel heeft en probeert te achterhalen wat dit doel is. Hogere-orde mentaliseeragenten zoals ToM_2 agenten houden er ook rekening mee dat onderhandelingspartners mogelijk proberen te achterhalen wat hun doel is.

In Hoofdstuk 5 onderzoeken we *single-shot* onderhandelingen, waarin twee spelers (de *Delers*) tegelijk een voorstel doen om een deel van hun eigen fiches te ruilen tegen een deel van de fiches van een derde speler (de *Kiezer*). De Kiezer kiest daarop het voorstel dat haar het meest oplevert. Wanneer geen van beide voorstellen voordelig is voor de Kiezer kan ze er ook voor kiezen om beide voorstellen af te wijzen. De Delers concurreren dus voor de kans om te kunnen onderhandelen met de Kiezer. In deze opzet blijkt zowel eerste-orde als twee-orde mentaliseren tot betere uitkomsten te leiden voor Delers, zowel in termen van individuele score als in termen van de gezamenlijke score van alle drie de spelers. Hoewel ToM_3 Delers geen hogere score halen dan ToM_2 Delers, blijken ToM_4 Delers wel een hogere individuele score te kunnen behalen dan lagere-orde mentaliseeragenten.

De simulaties van onderhandelingen in Hoofdstuk 5 laten ook het belang van adaptief gedrag zien. Het model van ToM_0 agenten dat wordt gepresenteerd in dit proefschrift neemt aan dat ToM_0 agenten hun gedrag aanpassen op basis van geobserveerd gedrag van anderen. Wanneer ToM_0 Delers in plaats daarvan een statische strategie gebruiken die het gedrag van andere Delers negeert, zien we

een verschuiving in de resultaten. In dit geval blijken ToM_4 Delers niet beter te scoren dan lagere-orde mentaliseeragenten, maar zien we dat het voordeel van vierde-orde mentaliseren doorschuift naar ToM_5 Delers.

In plaats van single-shot onderhandelingen kijken we in Hoofdstuk 6 naar een dynamischer vorm van onderhandelen waarin twee spelers om de beurt een voorstel doen voor de onderlinge verdeling van fiches. Het spel eindigt wanneer een speler het voorstel van zijn partner accepteert, of wanneer één van beide spelers besluit niet verder te willen onderhandelen. De resultaten van simulaties met agenten laten zien dat wanneer agenten de mogelijkheid krijgen om gedurende meerdere rondes van onderhandelingen zich aan te passen aan het gedrag van hun partner, ToM_0 agenten in principe dezelfde score kunnen behalen als hogere-orde mentaliseeragenten. Echter, ToM_0 agenten die minder toegeven aan hun partner behalen een hogere score dan ToM_0 agenten die hun eigen voorstellen meer aanpassen in de richting van de voorstellen van hun partner. Natuurlijke selectie zou er in dit geval dus voor zorgen dat ToM_0 agenten uiteindelijk niet meer onderhandelen, maar aandringen op de uitkomst die hun eigen score maximaliseert.

De simulaties in Hoofdstuk 6 laten zien dat het gebruik van eerste-orde mentaliseren een compleet verval van onderhandelen kan voorkomen, maar ook dat ToM_1 agenten nog steeds de aandrang hebben om te proberen hun eigen score te verhogen ten koste van de score van de ander. Dit zorgt ervoor dat, hoewel ToM_1 agenten de fiches uiteindelijk zodanig ruilen dat beide spelers er beter van worden, ToM_1 agenten niet de hoogst mogelijke gezamenlijke score behalen. Het gebruik van tweede-orde mentaliseren blijkt agenten beter te helpen om samenwerking en competitie te balanceren. Wanneer twee ToM_2 agenten met elkaar onderhandelen komen ze over het algemeen uit bij een uitkomst waarbij beide spelers een vergelijkbare score behalen.

In Hoofdstuk 7 onderhandelen proefpersonen rechtstreeks met mentaliseeragenten in deze dynamische onderhandelingsituatie. Uit de resultaten blijkt dat proefpersonen in deze omstandigheden gedrag vertonen dat meer overeenkomsten heeft met eerste-orde en tweede-orde mentaliseren dan met nulde-orde mentaliseren. Daarnaast blijkt dat, wanneer proefpersonen onderhandelen met ToM_2 agenten, het gedrag van proefpersonen ook meer overeenkomstig is met tweede-orde mentaliseren. Met andere woorden, proefpersonen lijken te worden aangeemoedigd tot het gebruik van hogere-orde mentaliseren door het gedrag van een onderhandelingspartner.

Conclusie

De agentsimulaties in dit proefschrift laten zien dat het gebruik van hogere-orde mentaliseren in verschillende situaties voordelen met zich mee kan brengen. Zoals de Machiavelliaanse-intelligentiehypothese voorspelt helpt hogere-orde mentalise-

	Hypotheses		
	Machiavelliaans	Vygotskiaans	Gemengde doelen
Overeenkomst met empirische resultaten			
Menselijk mentaliseren	hoog	laag	hoog
Mentaliseren andere soorten	laag	laag	hoog

Tabel 2: Samenvatting van de overeenkomsten van menselijke en niet-menselijke mentaliseervaardigheden met de voorspellingen van de Machiavelliaanse-intelligentiehypothese, de Vygotskiaanse-intelligentiehypothese, en de gemengde-doelenhypothese.

ren agenten om hun tegenstander te slim af te zijn in een competitieve situatie. Bovendien blijkt dat het gedrag van mensen in deze situaties ook meer lijkt op het gebruik van hogere-orde mentaliseren dan op het gebruik van nulde-orde en eerste-orde mentaliseren. Het is echter niet waarschijnlijk dat alleen competitieve situaties tot hogere-orde mentaliseren in mensen heeft geleid. Vergeleken met andere soorten zoals bijvoorbeeld chimpansees worden mensen vaak als een coöperatieve soort beschreven (Tomasello, 2009). Als concurrentie de belangrijkste drijfveer zou zijn voor de emergentie van hogere-orde mentaliseren is, dan zou de verwachting zijn dat chimpansees meer baat hebben bij hogere-orde mentaliseren dan mensen. Uit experimenten blijkt echter dat mensen gebruik maken van hogere-orde mentaliseren, terwijl chimpansees deze vaardigheid niet blijken te bezitten.

Deel 2 van dit proefschrift laat zien hoe agenten kunnen profiteren van hogere-orde mentaliseren bij het opzetten van een nieuwe manier van communiceren. Hoewel deze voorspellingen van de Vygotskiaanse intelligentie hypothese bevestigd, is het niet waarschijnlijk dat het nut van menselijk hogere-orde mentaliseren verklaard kan worden puur op het feit dat het samenwerking bevordert. Ten eerste laten onze simulaties alleen zien dat hogere-orde mentaliseren helpt bij het opbouwen van samenwerking. Zodra agenten weten hoe ze samen kunnen werken valt de noodzaak voor hogere-orde mentaliseren weg. Bovendien hebben eerdere agentsimulaties ook aangetoond dat samenwerkingsverbanden kunnen ontstaan zonder gebruik te maken van mentaliseren (Boyd & Richerson, 1992; Boyd et al., 2003; Nowak, 2006; Sigmund, 2010; De Weerd & Verbrugge, 2011). Dit laat zien dat mentaliseren niet noodzakelijk is voor samenwerking.

Van de drie hypothesen voor de emergentie van hogere-orde mentaliseren die we in dit proefschrift hebben bekeken, lijkt de gemengde doelen hypothese het meest waarschijnlijkst. Agentsimulaties in onderhandelingsituaties laten zien dat hogere-orde mentaliseren kan leiden tot hogere individuele scores, maar ook betere uitkomsten voor de onderhandeling in het algemeen. Daarnaast kan tweede-orde mentaliseren helpen om competitieve en coöperatieve doelen beter te balanceren

en een verval van onderhandelen kan voorkomen. Dit laat zien dat hogere-orde mentaliseren niet alleen voordelig kan zijn in dit soort situaties, maar mogelijk zelfs noodzakelijk.

De resultaten in dit proefschrift laten echter ook zien dat de voordelen voor hogere-orde mentaliseren verschillend is in de verschillende situaties. In competitieve situaties kunnen agenten voordeel behalen uit mentaliseren door zich te realiseren dat er geen evenwichtssituatie bereikt is. In een dergelijke situatie zou een tegenstander zijn gedrag kunnen aanpassen om een hogere score te bereiken. Mentaliseren helpt agenten om dit te anticiperen. Zolang het evenwicht nog niet is bereikt kunnen mentaliseeragenten blijven profiteren.

Agenten in coöperatieve situaties kunnen daarentegen baat hebben bij hogere-orde mentaliseren door sneller naar een evenwichtssituatie te komen. Door na te denken over hoe lagere-orde mentaliseeragenten leren van observaties kunnen hogere-orde mentaliseeragenten de uitkomst van dit leerproces voorspellen en daarmee aansturen op een stabiele samenwerking. Mentaliseren verliest echter zijn nut wanneer dit evenwicht uiteindelijk bereikt is.

In situaties waarin concurrentie als samenwerking beide een rol spelen kan hogere-orde mentaliseren agenten helpen om ongewenste evenwichtssituaties te vermijden. Nulde-orde mentaliseeragenten hebben alleen oog voor hun eigen doelen, wat ertoe kan leiden dat oplossingen die voor beide partijen gunstig zijn over het hoofd worden gezien. Hogere-orde mentaliseeragenten kunnen redeneren hoe ze hun eigen doel relevant kunnen maken voor de ander, bijvoorbeeld op de manier waarop de *tit-for-tat* strategie samenwerking bevordert in het *prisoner's dilemma*.

De resultaten in dit proefschrift laten zien dat agenten op verschillende manieren kunnen profiteren van hogere-orde mentaliseren, en dat het voordeel van hogere-orde mentaliseren zich niet beperkt tot één type situatie. Sterker nog, door te mentaliseren kunnen agenten leren over het gedrag van anderen op een manier die onafhankelijk is van het feit of de situatie competitief, coöperatief, of een combinatie van beide is. De belangrijkste drijfveer achter de emergentie van hogere-orde mentaliseren zou daarom ook algemener kunnen zijn dan een specifieke situatie. Hogere-orde mentaliseren zou mensen er toe in staat kunnen stellen om te gaan met complexe sociale interacties die voortkomen uit een moeilijk voorspelbare omgeving, waarin agenten zich in situaties bevinden die soms coöperatief, soms competitief, en soms een combinatie van beide zijn.

