

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.

If You Know What I Mean

*Agent-Based Models for Understanding
The Function of Higher-Order Theory of Mind*

Harmen de Weerd

Printing: Ridderprint BV.

ISBN printed version: 978-90-367-8128-2

ISBN electronic version: 978-90-367-8127-5

© Harmen de Weerd, Groningen, the Netherlands, 2015.



university of
 groningen

If You Know What I Mean

Agent-Based Models for Understanding
The Function of Higher-Order Theory of Mind

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 2 October 2015 at 14.30 hours

by

Hermanes Albertus de Weerd

born on 22 January 1981
in Raalte

Supervisor

Prof. dr. L.C. Verbrugge

Co-supervisor

Dr. H.B. Verheij

Assessment committee

Prof. dr. N.A. Taatgen

Prof. dr. S. Kraus

Prof. dr. P. Rosenbloom

Contents

1	Introduction	1
1.1	What is theory of mind?	2
1.2	Perspectives on theory of mind	4
1.3	Theory of mind in humans	7
1.3.1	Development of theory of mind reasoning	7
1.3.2	Adult theory of mind reasoning	9
1.3.3	Theory of mind in strategic games	11
1.3.4	Computational cognitive models of human theory of mind	13
1.4	Theory of mind in non-human species	15
1.4.1	Primates	16
1.4.2	Corvids	19
1.4.3	Computational models of non-human theory of mind	20
1.5	The function of theory of mind and its higher orders	20
1.5.1	The Machiavellian intelligence hypothesis	21
1.5.2	The Vygotskian intelligence hypothesis	22
1.5.3	The mixed-motive interaction hypothesis	23
1.5.4	Summary: Hypotheses for higher-order theory of mind	24
1.6	Agent-based computational models of interaction	25
1.7	Outline of the thesis	26
 Part I. Testing the Machiavellian intelligence hypothesis		29
2	How much does it help to know what she knows you know?	31
2.1	Introduction	32
2.1.1	Theory of mind abilities in humans and animals	32
2.1.2	Evolution of theory of mind	33
2.1.3	Agent-based modeling	34
2.2	Game settings	35
2.2.1	Rock-paper-scissors variations	35
2.2.2	Limited Bidding	39
2.2.3	Rational players	40

2.2.4	Hypotheses about the effectiveness of theory of mind	42
2.3	Playing the games using simulation-theory of mind	43
2.3.1	Zero-order theory of mind	43
2.3.2	First-order theory of mind	44
2.3.3	Second-order theory of mind	45
2.4	Model	45
2.4.1	Representation of the games	45
2.4.2	Zero-order theory of mind agents	47
2.4.3	First-order theory of mind agents	49
2.4.4	Second-order theory of mind agents	52
2.4.5	Higher orders of theory of mind agents	56
2.4.6	Belief adjustment and learning speed	56
2.5	Results	60
2.5.1	Rock-paper-scissors	61
2.5.2	Elemental rock-paper-scissors	64
2.5.3	Rock-paper-scissors-lizard-Spock	66
2.5.4	Limited Bidding	69
2.5.5	Summary of results	71
2.6	Discussion and conclusion	72
3	Theory of mind in the Mod game	75
3.1	Introduction	76
3.2	Mod game	77
3.3	Theory of mind agents in the Mod game	79
3.3.1	Zero-order theory of mind	79
3.3.2	First-order theory of mind	80
3.3.3	Higher orders of theory of mind	80
3.4	Simulation results	81
3.5	Discussion	84
	Part II. Testing the Vygotskian intelligence hypothesis	87
4	Higher-order theory of mind in the Tacit Communication Game	89
4.1	Introduction	90
4.2	Theory of mind in communication	91
4.3	Game setting	92
4.4	Theory of mind in the Tacit Communication Game	95
4.4.1	Zero-order theory of mind	95
4.4.2	First-order theory of mind	97
4.4.3	Higher orders theory of mind	99
4.5	Mathematical model of theory of mind	100

4.5.1	Zero-order theory of mind model	101
4.5.2	Theory of mind model	102
4.6	Results	105
4.7	Discussion	109
 Part III. Testing the mixed-motive interaction hypothesis		111
 5 The adaptive advantage of reasoning about other minds when competition and cooperation are mixed		113
5.1	Introduction	114
5.1.1	Three hypotheses for the emergence of higher-order theory of mind	114
5.1.2	Contribution and structure of this paper	116
5.2	Game setting	117
5.3	Theory of mind in Colored Trails	121
5.3.1	Zero-order theory of mind allocator	121
5.3.2	First-order theory of mind allocator	122
5.3.3	Higher orders of theory of mind agent	123
5.4	Mathematical model of theory of mind	123
5.4.1	Model of zero-order theory of mind	124
5.4.2	Model of first-order theory of mind	126
5.4.3	Model of higher-order theory of mind	130
5.4.4	Learning across games	131
5.5	Results of simulation experiments	132
5.5.1	Individual performance	133
5.5.2	Social welfare results	136
5.6	The role of zero-order theory of mind	137
5.6.1	The role of agent adaptivity	137
5.6.2	The role of the learning model	140
5.7	Related work on models of theory of mind	142
5.8	Conclusion	144
 6 Negotiating with other minds		147
6.1	Introduction	148
6.2	Colored Trails	151
6.3	Theory of mind in Colored Trails	153
6.3.1	Zero-order theory of mind agent	154
6.3.2	First-order theory of mind agent	154
6.3.3	Higher orders of theory of mind agent	155
6.4	Mathematical model of theory of mind	156
6.4.1	Model of zero-order theory of mind	157

6.4.2	Model of first-order theory of mind	158
6.4.3	Higher-order theory of mind agents	160
6.4.4	Learning from observations	161
6.4.5	Learning across games	164
6.5	Simulation results	165
6.5.1	Individual performance results	166
6.5.2	Social welfare results	171
6.6	Summary and discussion	174
6.7	Conclusion	177
7	Savvy software agents can encourage the use of second-order theory of mind by negotiators	179
7.1	Introduction	180
7.2	Colored Trails	181
7.3	Theory of mind software agents	183
7.3.1	Zero-order theory of mind	183
7.3.2	First-order theory of mind	183
7.3.3	Second-order theory of mind	184
7.4	Methods	185
7.4.1	Participants	185
7.4.2	Materials	185
7.4.3	Design and procedure	186
7.5	Results	187
7.6	Discussion and conclusion	189
7.7	Acknowledgments	191
	Part IV. Putting the pieces together	193
8	Related work	195
8.1	Models of recursive reasoning	195
8.2	Models of human behavior	199
9	Discussion and conclusion	201
9.1	Summary	201
9.1.1	The Machiavellian intelligence hypothesis	202
9.1.2	The Vygotskian intelligence hypothesis	203
9.1.3	The mixed-motive interaction hypothesis	204
9.2	Conclusion	206
	References	211
	Nederlandse samenvatting	237

CONTENTS

ix

Acknowledgments

247

List of publications

249

