

ESTIMATING THE USE OF THEORY OF MIND USING AGENT-BASED MODELING

Bachelorproject

Denny Diepgrond, d.diepgrond@student.rug.nl

Abstract: Decisions made in a social context can lead people to attribute mental states to others. For example assuming an approaching road user to hit the brakes so you can drive through. This phenomenon is called theory of mind (ToM). Engaging in social interactions can lead to recursive thinking of the sort “I think that you think that I think”. The depth of this recursion reflects the ToM-sophistication of a person. In this study we will use agents based on reinforcement learning to estimate the ToM-behavior of Bayesian agents and human participants in the zero-sum game of hide and seek. Our method will be tested on data of hide and seek games between participants and Bayesian agents. Results prove our estimation method to be effective on the Bayesian agents. Their ToM-behavior could be modeled accurately by our estimation method. The behavior of human participants was best modeled by zero- and second-order ToM. They showed a striking lack of first-order ToM-use and limited evidence of opponent modeling. The inability to use first-order ToM could indicate the use of a shortcut in the reasoning process of people in simple social interactions that skips first-order theory ToM.

1. Introduction

People make predictions of possible outcomes before making a decision. These predictions involve their beliefs, emotions and intentions. When a decision must be made in a social context, they also take the beliefs, emotions and intentions of other people into account. Attributing such mental states to others is called *theory of mind* or *mentalizing* (Premack & Woodruff, 1978). Theory of mind (ToM) allows people to understand that mental states can be used to predict and explain the behavior of other people they encounter in social interactions. In this paper we will use agent-based modeling to estimate the ToM-behavior of human participants. More information on the focus of this study can be found in Section 1.4.

1.1. Development of theory of mind

The ability to attribute mental states to others gradually develops during childhood. The first indications of theory of mind understanding are found at children between the age of three and five. (Frith & Frith, 2003). At this age, children develop the ability to attribute false beliefs to others (Wimmer & Perner, 1983; Perner & Lang, 1999). This means that children know that other people can have beliefs about the world that are different from their own. This means they must understand how knowledge is formed, that

knowledge influences beliefs, that other people’s behavior can be predicted by their mental states and that mental states can differ from what is really happening in the world. Thinking about other people’s mental states is known as first-order theory of mind.

Failure in learning to attribute mental states to others is associated with several neuropsychiatric disorders (American Psychiatric Association, 2013). This especially holds for autism (Baron-Cohen *et al.*, 1985), schizophrenia (Bora *et al.*, 2009) and attention deficit hyperactivity disorder (Korkmaz, 2011). Learning about the characteristics of theory of mind can therefore help us understand the minds of the people suffering from those mental disorders.

Although the human ability for theory of mind is well-established, the use of theory of mind by non-human animals is controversial. There are studies that propose non-human animals are able to attribute mental states to other animals (Schmelz *et al.*, 2011). Proving theory of mind in non-human animals is however difficult. Both studies that prove non-human animals use theory of mind and studies that prove they do not use theory of mind are criticized. The first because they would not rely on mental state attribution (Penn & Povinelli, 2007) and the second because they are too complex (Hare *et al.*, 2001). Proving the

effectiveness of the estimation of ToM-properties could stimulate new ways of research on non-human ToM-abilities.

1.2. Theory of mind in games

Games are a great way to simulate several social real-life aspects like negotiation and cooperation. In zero-sum games, not every player can win. The total value of each game will be zero because for every win of an agent stands a loss for an opponent. In this study we will use a zero-sum game in its simplest form, hide and seek. In this game, a player has two hiding options. The second player needs to pick one of the hiding options to search for the first player.

When people engage in games in which they interact with others, they use recursive thinking of the sort “I think that you think that I think” (Hedden & Zhang, 2002; Meijering *et al.*, 2011; Nagel, 1995). Thinking about the fact that other people also think about your mental states is second-order ToM. This behavior assumes that the opponent itself is also a rational player with a possible theory of mind. However, attributing ToM to others is not an automatic process (Devaine *et al.*, 2014a; Meijering *et al.*, 2010). This is shown in a study in which participants do two exactly the same experiments except in one case it is framed as if they were playing against a mentalizing agent whereas in the other it is framed as if they were playing a slot machine (Devaine *et al.*, 2014a). It was shown that people obtain better results in the social framing and that they were not able to decipher intentional behavior without a priori attributing mental states to others.

The recursive thought processes during games can become labyrinthine very quickly. Although people can gain an advantage over an opponent by using a higher-order theory of mind, meaning more sophisticated recursive thinking, it is shown that people do not use ToM of arbitrary depth (Kinderman *et al.*, 1998; Stiller & Dunbar, 2007). This could possibly be explained by the advantage being rather limited (de Weerd *et al.*, 2013). Alternatively, the information cost for using higher-order theories of mind may be larger than the benefits. Since there is no consensus on to what extent people use ToM in games like hide and seek, we will try

to gain new insights by estimating ToM-properties.

1.3. Agent-based modeling

Through simulation it is possible to model agents that interact with an opponent (Epstein, 2006). Agents that model the opponent as an opponent-modeling agent itself allow complicated recursive models. The sophistication in these is measured in the amount of iterative steps an agent can consider during decision making. This corresponds with ToM-behavior of different orders. The main advantages of using agent-based modeling is that the mental states are controllable and known.

In this paper we consider the agents based on reinforcement learning (from now on: *RL agents*) as used in the studies of de Weerd *et al.* (2013). These agents will be used to model the ToM use of Bayesian agents from a study of Devaine *et al.* (2014a). The main difference between the two types of agents is the relative simplicity of the RL agents.

Both types of agents are designed to play a computational version of hide and seek. Devaine *et al.* used Bayesian agents of different sophistication to play the hide and seek game with 26 human participants. Participants played a game of sixty trials of hide and seek against eight opponents divided over two different framings. The first framing is social and in this framing participants play hide and seek against another player that also needs to achieve its own goals. The second framing is non-social and the participants play against a slot machine like the ones in a casino. Both framings include the same four opponents so despite the different framings, opponent behavior is similar. The used models will be described in detail in Section 2.

1.4. Estimating theory of mind behavior

In this study a ToM estimation method based on the RL agents of de Weerd *et al.* (2013) will be used to model the ToM-behavior of Bayesian agents from Devaine *et al.* (2014a). Accurate modeling of the Bayesian agents with the RL based agents will confirm the effectiveness of our estimation method.

The accuracy of the estimation can be verified because the ToM-orders of the Bayesian agents are known. When the estimation method

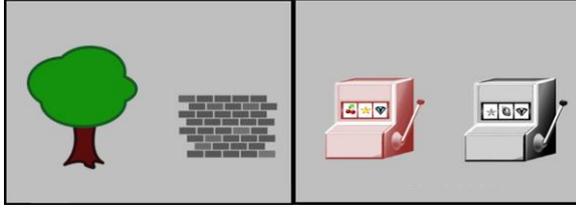


Figure 2.1: The stimuli for the main task of the Devaine *et al.* (2014a) study. The social framing can be seen on the left whilst the non-social framing can be seen on the right.

proves to be accurate, we have shown that the simpler reinforcement learning algorithm can model the more complex Bayesian agents in the hide and seek game. This would also show that we can use RL agents to estimate ToM-levels. Because the Bayesian agents and the RL agents are both based on the same recursive ToM-principle, we expect the estimations of the RL agents to have a high similarity with the actual behavior of the Bayesian agents.

When the modeling of the Bayesian agents proved our estimation method to be effective, we will use it to estimate the ToM use of human participants in hide and seek. This could give us more insight in social cognition during simple social interactions. We will also use the estimation method to see if participants engage in opponent modeling. That is, to see if participants change their ToM-level according to the behavior of their opponent.

2. Methods

In the remainder of this paper, k-order theory of mind is sometimes abbreviated as ToM-k. For example ToM-1 is the abbreviation of first-order theory of mind.

2.1. Data

In this study the data of the Devaine *et al.* (2014a) study will be used as input for the RL algorithm that estimates ToM-orders (from now on *estimator*). The data is divided in two parts, one for the games with the social framing and the other for the non-social framing games. In the social framing games, the participants received a stimulus which displayed a tree and a brick wall. They had to choose one of those objects to search behind. In the non-social framing, the participants received a stimulus which displayed two slot machines. They had to choose one of the

Table 2.1: The hide and seek payoff table as used in the Devaine *et al.* (2014a) study. The payoffs are a function of the action of the seeking player a^{self} and the action of the hiding opponent a^{opp} . Every cell has the following structure: [payoff_{self}, payoff_{opp}]. Choosing the left option corresponds with $a = 1$ while choosing the right option corresponds with $a = 0$.

Utilities	$a^{\text{opp}} = 1$	$a^{\text{opp}} = 0$
$a^{\text{self}} = 1$	[1,0]	[0,1]
$a^{\text{self}} = 0$	[0,1]	[1,0]

slot machines to play. Both stimuli are displayed in Figure 2.1. In the study, participants played these two games against all four types of opponent agents. The first opponent agent played according to a random sequence with a 65% bias for one option that was counterbalanced between framings within participants. The other three opponent agents were ToM-0, ToM-1 and ToM-2 agents. ToM-0 agents are learning agents as they adapt to the participant's behavior. ToM-1 and ToM-2 agents also have this property but in addition they also use an artificial form of mentalizing.

All 26 participants played sixty trials per opponent per game in which they had to choose one of two options in less than 1300 milliseconds. When both players picked the same option, the participant received a payoff of 1 and the agent received a payoff of 0. When the two players picked different options, the payoffs were the other way around. After every trial, they received feedback. More information on their study can be found in the paper of Devaine *et al.* (2014a).

2.2. Model

The hide and seek game is modeled by a payoff table as a function of the action a^{self} of the player and the action a^{opp} of its opponent. The table shows the payoff the player gets when choosing action a^{self} while the other player chooses a^{opp} . a^{self} and a^{opp} take binary values encoding the two possible hide and seek options. The payoff table of the hide and seek game used in this study is shown in Table 2.1. The context is competitive as the payoffs of the winning agent and the losing agent are balanced. Every agent tries to maximize the expected payoff by using its predictions of the next action of the opponent.

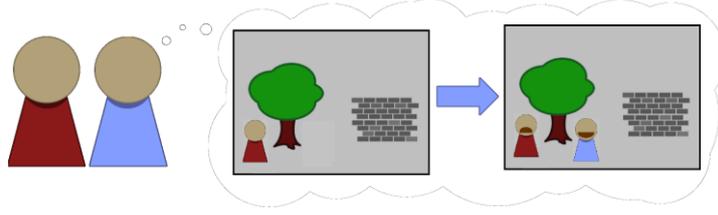


Figure 2.2: An example of a thought process of a zero-order theory of mind agent playing hide and seek. The blue agent is the seeker while the red agent is the hider. Figures of Devaine *et al.* (2014a) and de Weerd *et al.* (2013) were used during the creation of this figure.

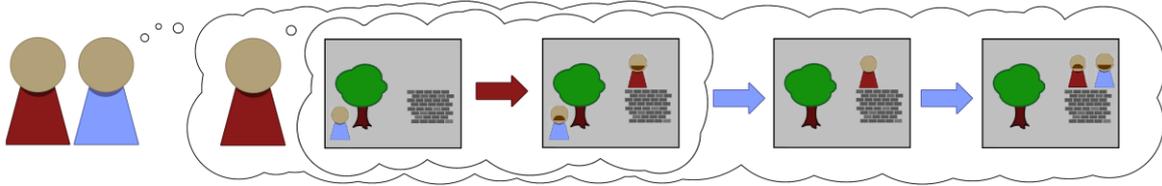


Figure 2.3: An example of a thought process of a first-order theory of mind agent playing hide and seek. The blue agent is the seeker while the red agent is the hider. Figures of Devaine *et al.* (2014a) and de Weerd *et al.* (2013) were used during the creation of this figure.

The agents imitate the human strategy of simulation-theory of mind. This means they take the perspective of their opponent and they determine what they would do in that position (Nichols & Stich, 2003). Thus, they assume that their opponent uses the same thought-process as they do. The next subsections explain how these agents of different ToM-orders come to their decisions using the uncertain predictions of their opponent's next action. It is important to remember that in all examples the agent is the seeker whilst the opponent is the hider.

2.2.1 Zero-order theory of mind

A zero-order theory of mind agent is unable to engage in mentalizing. It does not expect the opponent to have a goal different from its own. By memorizing the actions the opponent performs, a probability distribution is computed and constantly updated after every action. When an agent remembers its opponent chose option 1 most often, it will choose option 1 as its next action. An example of a thought process of a ToM-0 agent can be seen in Figure 2.2.

Let us assume that the ToM-0 agent's beliefs $b^{(0)}$ of the opponent's actions are: $b^{(0)}(a^{opp}=0) = 0.4$ and $b^{(0)}(a^{opp}=1) = 0.6$. That is, the agent believes that there is 60% probability that its opponent will perform action 1. These beliefs are based on the probability distribution of the actions the opponent played in previous encounters. $a^{opp}=1$ here resembles the opponent picking option 1 in the hide and seek game. To compute the value of

a possible action, the agent multiplies the probabilities of the opponent's actions with their corresponding payoffs (Table 2.1). These probability-payoff pairs are added up, resulting in the value of that action. The expected value (EV) for playing action 1 based on the zero-order beliefs in the example is: $EV^{(0)}(a^{self}=1) = 0.6 \cdot 1 + 0.4 \cdot 0 = 0.6$. Hence, the value for choosing option 0 is $EV^{(0)}(a^{self}=0) = 0.6 \cdot 0 + 0.4 \cdot 1 = 0.4$. The agent always chooses the action with the maximum expected value so in this state action 1 will be selected.

2.2.2 First-order theory of mind

A first-order theory of mind agent realizes that its opponent could also be a ToM-agent, meaning that the opposing agent will have its own goals and desires that could be different from the agent's goals. The ToM-1 agent assumes the opponent is a ToM-0 agent. It uses this assumption to form beliefs about the thought process of its opponent on which the agent can anticipate.

The agent has information on its own probability distribution which it uses to form beliefs the same way a ToM-0 agent would do. These beliefs form a prediction of which action the opponent will choose. A thought process of this kind is displayed in Figure 2.3. When an agent remembers it played option 1 most often, it will predict the opponent to play option 0. Therefore the agent should choose option 0.

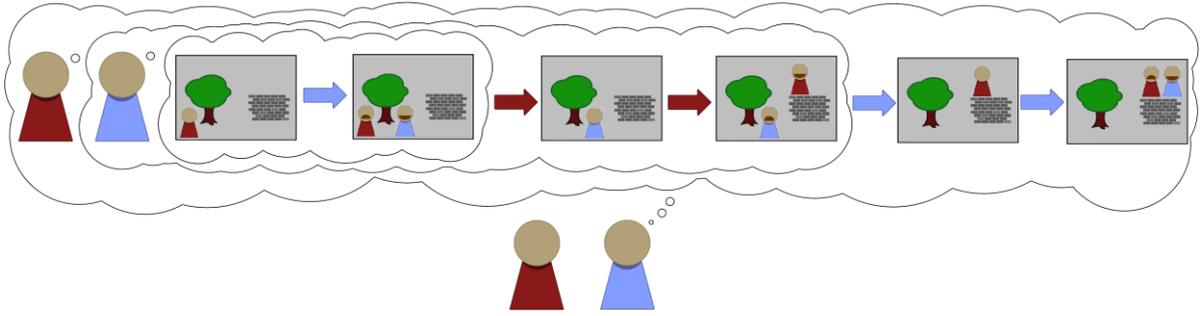


Figure 2.4: An example of a thought process of a second-order theory of mind agent playing hide and seek. The blue agent is the seeker while the red agent is the hider. Figures of Devaine *et al.* (2014a) and de Weerd *et al.* (2013) were used during the creation of this figure.

Let us assume that the agent's zero-order beliefs of the opponent's actions are the same as for the ToM-0 example: $b^{(0)}(a^{opp}=0) = 0.4$ and $b^{(0)}(a^{opp}=1) = 0.6$. Using the probability distribution of its own actions, it predicts the beliefs its opponent has. We call this the agent's first-order beliefs $b^{(1)}$. Let us assume that the first-order beliefs in this example are as follows: $b^{(1)}(a^{self}=0) = 0.3$ and $b^{(1)}(a^{self}=1) = 0.7$. The ToM-1 agent now uses the ToM-0 calculation of the optimal action to determine which action its opponent is most likely to play. The value for choosing option 1 is $EV^{(1)}(a^{opp}=1) = 0.7 \cdot 0 + 0.3 \cdot 1 = 0.3$. The value for choosing option 0 is $EV^{(1)}(a^{opp}=0) = 0.7 \cdot 1 + 0.3 \cdot 0 = 0.7$. The agent therefore predicts that its opponent is most likely to play action 0. The agent is a seeker so it should play action 0 as well according to its first-order beliefs.

The model allows the ToM-1 agent to deviate from the assumption that the opponent is a ToM-0 agent. The agent will adjust its behavior when its assumptions of the ToM-order of the opponent do not accurately predict actual opponent behavior. The ToM-1 agent does so by integrating its zero-order beliefs and the prediction it made using first-order beliefs. For the integration, the agent uses the confidence c_1 that the first-order theory of mind accurately predicts the behavior of its opponent. When the confidence is low, the agent will play more like a ToM-0 agent.

The agent computes the integrated belief for every possible action. It determines the values for playing the actions like a ToM-0 agent does but instead of zero-order beliefs it uses its integrated beliefs. More information on the integration of beliefs can be found in the paper of de Weerd *et al.* (2013). Confidence alterations and

additional parameters will be discussed in Section 2.3.1.

2.2.3 Higher-order theories of mind

A second-order theory of mind agent assumes its opponent is a ToM-1 agent. This means that the ToM-2 agent assumes that its opponent considers it is playing against a ToM-0 agent. An example of a thought process of a ToM-2 agent is shown in Figure 2.4. The agent assumes that its opponent knows the agent's probability distribution. The opponent assumes the agent to play option 1 because the opponent chose option 1 most often. Therefore, the opponent decides to choose option zero. The agent ultimately has predicted this so it chooses option 0 as well. An agent that uses third-order theory of mind will predict this behavior and so every additional theory of mind-order leads to deeper recursion of this mental state attribution. Keep in mind that every ToM-agent can adjust its beliefs about the ToM-order of the opponent using the confidence if it does not resemble the opponent's behavior.

Every ToM-k agent (where $k > 1$) has a belief structure $b^{(k)}$ to model the thought process of a ToM-(k-1) opponent. An agent starts with integrating its zero-order beliefs with its first-order prediction. Then it integrates this integration with its second-order prediction. This recursive integrating continues until the k-order prediction ultimately is included in the integration.

2.3. Design

2.3.1 ToM-estimation

The data described in Section 2.1 serves as input to an estimator that tests which ToM-order resembles the behavior displayed in a game (sixty trials) most accurately. The data is preprocessed by omitting the first five trials of every game. For these trials, it is unlikely to have a reasonably accurate probability distribution. Comparing the random choices the players had to make during these trials with estimates based on various ToM-orders could lead to infecting the data with noise.

The estimator is an algorithm based on the agents described in Section 2.2. Remember that a ToM-k agent assigns a value to every possible action it can play, $EV = p_1 \cdot \text{payoff}_1 + p_0 \cdot \text{payoff}_0$. The ToM-k agent will perform the action with the highest value. The estimator uses the same principle but instead of playing the game, it reports the value for playing action 1 and then proceeds with the next trial. The estimations of used ToM-orders are therefore based on the similarity between the computed value of the estimator and the value of the action actually performed by the player. A given ToM-order will be more likely to be used by the player if the expected value of that order is high when the player chose action 1 and when the expected value of that order is low when the player chose action 0. It is important to point out that this estimator models itself as the actual player it tries to determine the used ToM-order of. This is contrary to the study of de Weerd *et al.* (2015), in which the estimator modeled itself as the opponent of the player of which it tried to determine the used ToM-order.

The estimator computes the values for ToM-0, ToM-1, ToM-2, ToM-3 and ToM-4. Higher-order ToM-estimations are not used because recent studies have suggested diminishing returns on higher orders of theory of mind in games (de Weerd *et al.*, 2013).

2.3.2 Beliefs and parameters

The model of Section 2.2 uses a couple of parameters that need some explanation. As stated before, a ToM-k agent has beliefs concerning the actions of its opponent for each possible ToM-order available to the agent. The

beliefs are randomly initialized when an agent meets an unfamiliar opponent. The decisions of the agent are based on these beliefs in combination with its confidence in the use of the different ToM-orders. Every ToM-k agent adjusts its beliefs and confidence in the different ToM-orders according to the decisions made by itself and its opponent and the corresponding outcome of the game.

The beliefs are adjusted to reflect the new probability distributions of chosen actions after each game of hide and seek. There is an additional variable, the learning speed λ , which controls the influence of new observations on the beliefs of the agent. A higher learning speed means the beliefs are more strongly influenced by the most current information while a lower learning speed makes the agent take information from the past into account as well.

The learning speed also controls the influence of new information on the adjustment of the agent's confidence in the use of the different ToM-orders. Each order of theory of mind makes a prediction of the next choice of the opponent. When the ToM-k prediction corresponds with the actual choice the opponent made, the confidence in the use of that order increases: $c_k = \lambda + (1-\lambda) \cdot c_k$. This only happens when there is no ToM-order n with $n < k$ that also predicted the action of the opponent correctly. In that case the confidence in ToM-k will remain unchanged. When the prediction of a certain ToM-order was incorrect, the confidence in that order will decrease: $c_k = (1-\lambda) \cdot c_k$. These update rules make sure the agent does not overestimate the ToM-order of its opponent because it only uses a higher ToM-order when the behavior of its opponent cannot be explained by a lower ToM-order.

In this study, the estimator will use the optimal learning speed for every different ToM-order per participant. These learning speeds are calculated by running the estimator on the data for every learning speed between 0 and 1 with a step size of 0.02. The learning speed that best fits the player data will be used in further analyses. The confidences were all initialized at $c=1$.

2.4. Analysis

After the estimator is applied to the data, the expected values will be analyzed. The data will

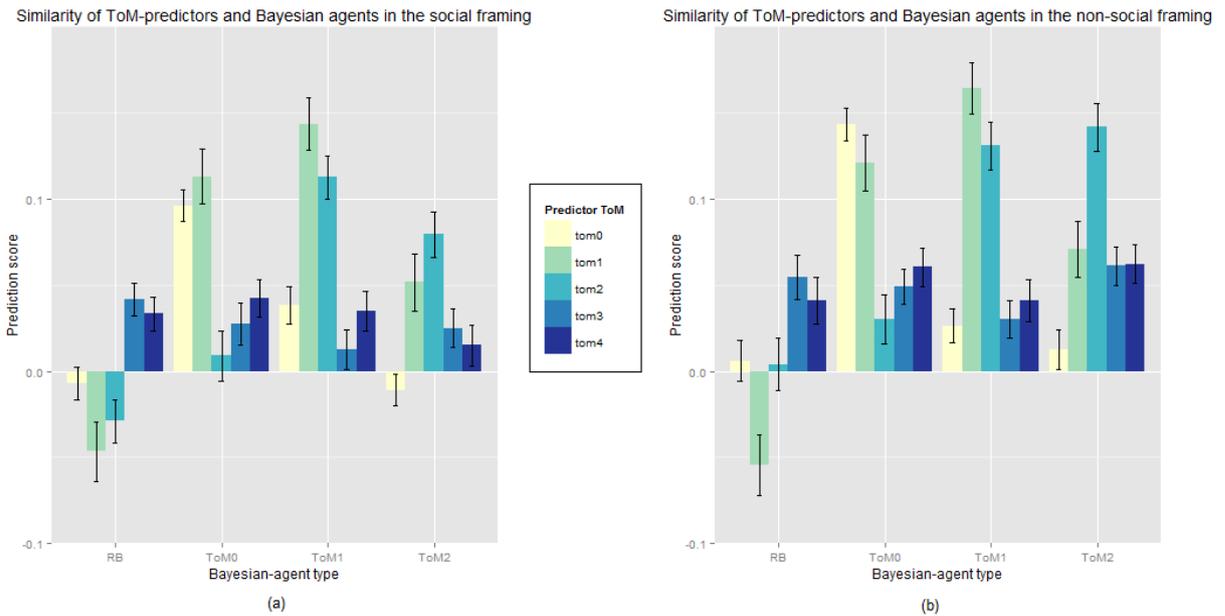


Figure 3.1: The estimation scores for the different ToM-orders of the estimator given the different Bayesian agent types (random biased agent RB, ToM-0 agent, ToM-1 agent, ToM-2 agent) in the social (a) and the non-social (b) framing. The estimation score consists of the difference between the mean expected value of the trials in which the Bayesian agent chose 1 and the mean expected value of the trials in which the Bayesian agent chose zero. Error bars depict one standard error.

be divided in four subgroups: Bayesian agents in the social framing (1a), Bayesian agents in the non-social framing (1b), participants in the social framing (2a) and participants in the non-social framing (2b). For every subgroup, estimation scores will be plotted for every ToM-order of the estimator against the Bayesian agent type. The *estimation score* is the difference between the mean expected value of the trials in which the player chose 1 and the mean expected value of the trials in which the player chose zero.

A logistic regression model will be created in the statistical programming language R (Team, 2000) in which the actions the player actually performed will be modelled by the interactions between the opponent agent type and the expected values for all different ToM-orders. The model will be tested for every subgroup. The fact that a coefficient is negative or positive in combination with its significance could give information regarding the ToM use of the hide and seek player.

3. Results

Prior to the analysis of the results, a model comparison was performed using AIC values

(Akaike, 1974). The full model, with all ToM-estimation orders included, was tested against nested models from which the highest ToM-estimation orders were omitted. The full model appeared to be the model with the minimum AIC value, hence being the preferred model.

The estimator based on the model from Section 2.2 has been applied to the data from Devaine *et al.* (2014) described in Section 2.1. This means estimations of the used ToM-orders are made for all subgroups of the data.

3.1. Agent data

Remember that no ToM-3 and ToM-4 agents were used in the Devaine *et al.* (2014a) study so for the estimations of the used ToM-order of the Bayesian agents, these expected values serve as a control condition. The results for group 1a and 1b are shown in Figures 3.1a and 3.1b.

In these figures, the estimation scores are plotted for every ToM-order of the estimator against the Bayesian-agent type. The general characteristics are the same for both plots. This suggests that the Bayesian agents display similar behavior in both framings. Remember that the underlying algorithms were the same for both framings in the study of Devaine *et al.* (2014a) so

Table 3.1: Coefficients with their significance codes between brackets, extracted from the logistic regression models showing the interaction between the expected values of the different ToM-estimators and the Bayesian agent type.

Significance codes: 0 < '***' < 0.001 < '**' < 0.01.

Agent model		<i>Social</i>	<i>Non-social</i>
<i>Agent type</i>	<i>Estimator</i>	β	β
<i>RB</i>	<i>ToM-0</i>	-0.350	-0.108
	<i>ToM-1</i>	-0.107	-0.579 (*)
	<i>ToM-2</i>	-0.873 (*)	0.018
	<i>ToM-3</i>	1.523 (**)	0.768 (*)
	<i>ToM-4</i>	0.323	0.499
<i>ToM-0</i>	<i>ToM-0</i>	3.094 (**)	4.310 (**)
	<i>ToM-1</i>	1.279 (**)	1.043 (**)
	<i>ToM-2</i>	-0.090	-0.385
	<i>ToM-3</i>	0.647	0.269
	<i>ToM-4</i>	0.332	0.187
<i>ToM-1</i>	<i>ToM-0</i>	0.747	0.166
	<i>ToM-1</i>	1.196 (**)	1.500 (**)
	<i>ToM-2</i>	1.583 (**)	1.125 (**)
	<i>ToM-3</i>	-0.361	-0.021
	<i>ToM-4</i>	0.725 (*)	0.405
<i>ToM-2</i>	<i>ToM-0</i>	-0.276	-0.090
	<i>ToM-1</i>	0.198	0.059
	<i>ToM-2</i>	1.147 (**)	1.737 (**)
	<i>ToM-3</i>	0.169	0.591
	<i>ToM-4</i>	0.112	0.486

no differences between the framings are expected in this study as well. For the ToM-0 Bayesian agent, the estimation scores of the zero- and first-order ToM-estimators stand out. The same holds for the estimation scores of the first- and second-order ToM-estimator for the ToM-1 Bayesian agent and for the estimation score of the second-order ToM-estimator for the ToM-2 Bayesian agent. Keeping in mind that a ToM-k agent can also play a ToM-(k-1) strategy, these results are not unexpected.

Remember that the learning speed that best fits the player data was used in the experiment. The mean ideal learning speed for the ToM-3 (M= 0.133, SD= 0.111) estimator proves to be significantly lower than the ideal learning speeds for the ToM-1 (M= 0.252, SD= 0.158), (t(91.6)=4.44, p < 0.001) and ToM-2 (M= 0.196, SD= 0.147), (t(95.0) = 2.46, p = 0.016) estimators. This explains why for the ToM-2 opponents, the

ToM-3 estimator apparently does not play like a ToM-2 agent. The low learning speed prevents it from basing its decision mainly on new information. The other ToM-agents have a higher learning speed which makes them more dynamic in changing their ToM-behavior.

The behavior of the random biased agent is hard to estimate using a ToM-agent so the results for this Bayesian agent type are very different from the Bayesian agent types that actually do have ToM-properties.

To see which ToM-estimation orders best model the behavior of the Bayesian agent, a logistic regression model is created which is tested for every Bayesian agent type. The combined results for all Bayesian agent types and both framings are displayed in Table 3.1. Since a large number of coefficients is estimated, we used a significance level of 0.01. To make sure that the behavior of different ToM-estimators was sufficiently different, we tested for the presence of *multicollinearity* (Farrar & Glauber, 1967). Correlations between ToM-estimator variables were well below 0.8 which means no evidence of multicollinearity has been found.

The results of the logistic regression support the findings in Figure 3.1. The only additional significant coefficient is found for the fourth-order estimation of the Bayesian ToM-1 agent in the social framing. This is striking because the Bayesian agents cannot reason using fourth-order ToM. The coefficient is however lower than the other significant coefficients.

The only ToM-estimator that estimates the RB agent's behavior in both framings is the third-order estimator. In both framings, one other ToM-estimation is negatively correlated with the RB agent behavior. Meaning that the RB agent plays hide and seek the exact opposite of that ToM-order. For the social framing this is ToM-2 and for the non-social framing this is ToM-1.

The difference between the estimates for both framings were also tested in a logistic regression model. For all ToM-orders, no significant interaction between the ToM-estimation and the framing could be found. Meaning that, as expected, we find no significant influence of framing on the behavior of the agent according to our ToM-estimator. This

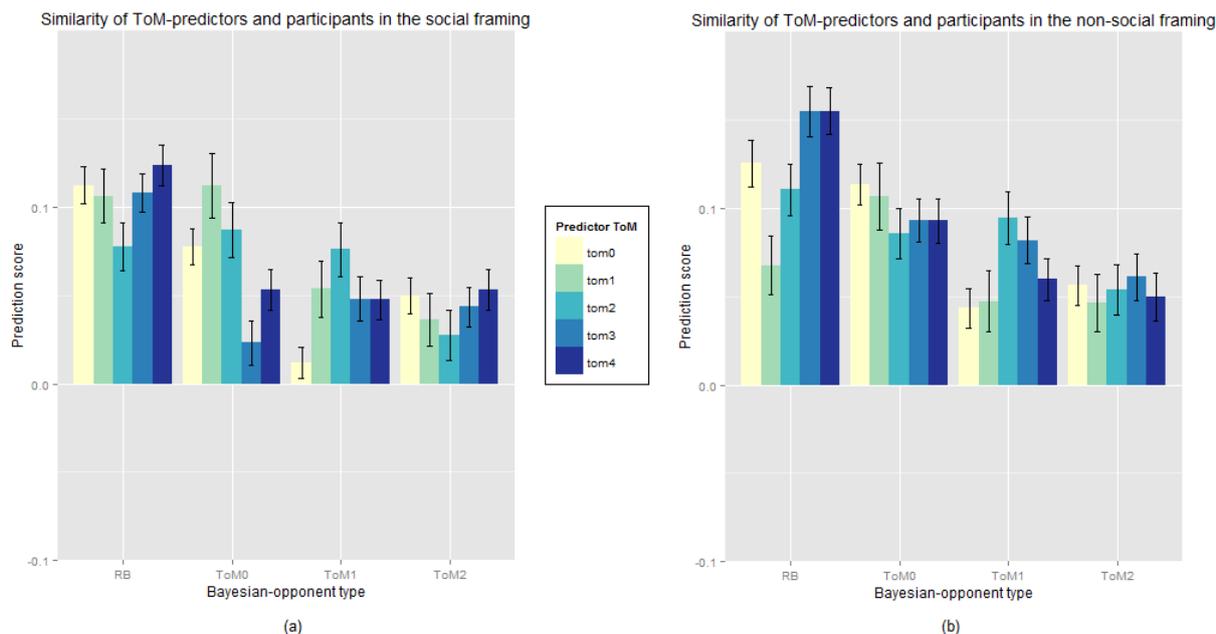


Figure 3.2: The estimation scores for the different ToM-orders of the estimator given the different Bayesian opponent types (random biased opponent RB, ToM-0 opponent, ToM-1 opponent, ToM-2 opponent) in the social (a) and the non-social (b) framing. The estimation score consists of the difference between the mean expected value of the trials in which the participant chose 1 and the mean expected value of the trials in which the participant chose zero. Error bars depict one standard error.

corresponds with the fact that the participants played against the same agents in both framings.

With the exception of the interaction between the fourth-order estimation and the Bayesian ToM-1 agent behavior in the social framing, the method used for estimation seems to work. Meaning that estimator agents can successfully determine ToM-level of Bayesian agents. For every type of Bayesian ToM-agent holds that the estimator with the same ToM-order is positively correlated with it. The tendency that the ToM-(k+1) estimation is positively correlated with the ToM-k Bayesian agent can also be extracted from the results. A ToM-2 agent can also choose to play like a ToM-1 agent when its own behavior is proved ineffective. Therefore it is not alarming that the ToM-(k+1) estimations are positively correlated with the ToM-k Bayesian agent behavior. For the ToM-2 Bayesian agent this is not the case. Probably because people are not likely to use ToM-3 or higher in negotiation games like hide and seek and rock-paper-scissors (Devaine *et al.*, 2014a; de Weerd *et al.*, 2013). The combination of correctly estimated ToM-orders and absence of differences between both framings yields enough trust in our estimation method to apply it to the participant data.

3.2. Participant data

The ToM-estimates for the participants cannot be verified because there is no information on their used ToM-orders. By computing the expected values for the participants' choices, the ToM-order can be found that correlates most with the participants' behavior. The results for group 2a and 2b are visualized in Figures 3.2a and 3.2b.

When looking at this figure, it is important to keep in mind the fact that the x-axis displays the opponent type. Therefore a high estimation score for a certain ToM-order means the participants are more likely to have played that ToM-order against that opponent. An optimal strategy would imply the use of ToM-k against a ToM-(k-1) opponent (Yoshida *et al.*, 2008; Devaine *et al.*, 2014a; de Weerd *et al.*, 2013). An optimal strategy against the RB opponent would be ToM-0 behavior with a low learning speed. This agent would use the probability distribution of actions made by its opponent over a long time as an indication of the next action. Meaning the ToM-0 agent would use the bias of the RB agent in its own advantage.

When looking at Figure 3.2, we see that the difference between the estimation scores of the

different ToM-estimators is less obvious than for the Bayesian agents. The difference between the social and the non-social framing seems to be bigger for the participants' behavior than for the agents' behavior. This suspicion is confirmed by the results of the logistic regression model. Using a confidence interval of 0.05, this model gives significant interactions of framing with the fourth-order ToM-estimation ($Z=2.20$, $p=0.028$) and of framing with the first-order ToM-estimation ($Z=2.78$, $p=0.005$). The observations regarding Figure 3.2 are clarified by the logistic regression model for the participants' data. For this data, the correlations between ToM-estimator variables were well below 0.8 as well. Therefore there is no evidence for multicollinearity. Again, we used a significance level of 0.01 because a large number of coefficients is estimated. The results of this model are shown in Table 3.2.

The model tells us that there are definitely some significant correlations between estimations and participant behavior. Against a random biased opponent, multiple estimations are significant. The optimal response, playing ToM-0, has the highest coefficient in both framings but with so much significant correlations it is obscure to speak of optimal behavior against a RB opponent.

For the ToM-0 opponent, the zero-order estimations have a significant correlation with the participant behavior in both framings whilst the second-order estimation only significantly correlates with participant behavior in the social framing. Note the absence of the ToM-1 estimation between the significant correlations. This is contrary to the optimal behavior against a ToM-0 opponent.

For the ToM-1 opponent, we see indications of optimal behavior in the social framing. Only the second-order ToM-estimation has a significant correlation with the participant behavior in this case. In the non-social framing we see a significant correlation between the behavior of participants with the zero-order estimation, the second-order estimation and the third-order estimation. In this case, playing ToM-0 would mean the opponent models the participant's behavior correctly so this is considered a bad strategy.

Table 3.2: Coefficients with their significance codes between brackets, extracted from the logistic regression models showing the interaction between the expected values of the different ToM-estimators for the participants' choice and the Bayesian opponent type.

Significance codes: $0 < *** < 0.001 < ** < 0.01$.

Participant model		<i>Social</i>	<i>Non-social</i>
<i>Opponent</i>	<i>Estimator</i>	β	β
<i>RB</i>	<i>ToM-0</i>	1.972 (**)	1.361 (**)
	<i>ToM-1</i>	0.471	-0.124
	<i>ToM-2</i>	0.799 (*)	1.018 (**)
	<i>ToM-3</i>	1.101	0.817 (*)
	<i>ToM-4</i>	0.835	1.174 (**)
<i>ToM-0</i>	<i>ToM-0</i>	1.891 (**)	2.238 (**)
	<i>ToM-1</i>	0.252	-0.030
	<i>ToM-2</i>	0.761 (**)	0.762
	<i>ToM-3</i>	-0.185	0.673
	<i>ToM-4</i>	0.650	0.282
<i>ToM-1</i>	<i>ToM-0</i>	0.676	1.556 (**)
	<i>ToM-1</i>	0.357	0.162
	<i>ToM-2</i>	0.603 (*)	0.879 (**)
	<i>ToM-3</i>	0.624	1.135 (*)
	<i>ToM-4</i>	0.510	-0.182
<i>ToM-2</i>	<i>ToM-0</i>	1.019 (*)	1.549 (**)
	<i>ToM-1</i>	0.355	-0.018
	<i>ToM-2</i>	0.235	0.727 (*)
	<i>ToM-3</i>	0.710	0.698
	<i>ToM-4</i>	0.492	-0.050

For the ToM-2 opponent, the zero-order estimations have a significant correlation with the participant behavior in both framings. In the non-social framing we also see a significant correlation between the behavior of participants and the second-order estimation. Therefore, against a ToM-2 opponent, participants do not play the theoretically optimal ToM-3 strategy. This is line with the evidence Devaine *et al.* (2014a) found, suggesting that human ToM-capabilities during competitive games are possibly limited at ToM-2 (Devaine *et al.*, 2014b). Assuming they will not play ToM-3, participants should at least avoid playing ToM-1 because this is the worst-case strategy, which they seem to do.

To test the opponent modelling properties of the participants, a logistic regression model was constructed to show interactions between the use of ToM of various orders and the type of

opponent. The model was created with a significance level of 0.01 because of the many variables that are tested for. For the social framing, results suggest that participants are able to control their ToM-0 behavior. The estimation score of ToM-0 is significantly higher against a RB opponent than against a ToM-1 opponent ($Z=2.63$, $p=0.009$). In the non-social framing, the fourth-order estimation score is higher against RB opponent than against ToM-1 ($Z=2.93$, $p=0.003$) and ToM-2 ($Z=2.89$, $p=0.004$) opponents. The interaction we found in the social framing suggests opponent modelling because ToM-0 is the optimal strategy against a RB opponent and the worst strategy against a ToM-1 opponent. However, the lack of significant interactions for the other ToM-orders attenuates the evidence of participants making use of optimal opponent modelling.

4. Discussion

In this study we have estimated the used ToM-orders of Bayesian agents and participants with an estimation method based on reinforcement learning agents. To test the estimating abilities of the estimator, it was first applied to the data of the Bayesian agents. The actual ToM-orders of the agents were known so performance could be tested. After this, the estimator was applied to the participant data to see which ToM-order human participants use against different opponents in the hide and seek game. We were also interested in the opponent modeling abilities of the participants.

Results have shown the effectiveness of our ToM-estimation method on Bayesian agents. The ToM-orders of the Bayesian agents all had a high and significant estimation score. Besides that, there was no difference between the social and the non-social framing. That is, our method successfully recovered the ToM-level of the Bayesian agents. Estimations of the used ToM-order of the participants were also made. Their behavior seems to consist of a combination of ToM-0 and ToM-2 strategies with a striking lack of ToM-1. Participants show little evidence of opponent modelling although they seem to be able to control their zero-order ToM use. We found no evidence of strategic use of ToM-1 or ToM-2. For the participants there is actually a significant difference between the social and the

non-social framing which is consistent with the results of Devaine *et al.* (2014a).

One of the goals of this study was confirming the ability of the estimator based on RL agents to correctly estimate the used ToM-order of Bayesian agents. Results have shown the method to be effective. Meaning that the Bayesian agents of Devaine *et al.* (2014a) could be modelled by the simpler reinforcement learning based agents of de Weerd *et al.* (2013). The estimation scores of the actual used ToM-order were significantly high for all Bayesian ToM-agents. This also holds for the estimation scores for the ToM-order that was one order above the actually used ToM-order of the Bayesian agents. This is probably caused by the fact that the RL agents can adjust their ToM-behavior to a lower order when the behavior of their opponent does not correspond with the prediction the agents made using their current beliefs. This does not happen for the ToM-3 estimator when the Bayesian agent actually used ToM-2. The mean ideal learning speed for the ToM-estimator is however, significantly lower than the ideal learning speeds for the ToM-1 and ToM-2 estimators. Therefore the ToM-3 estimator is less sensitive to new information.

The ToM-3 and ToM-4 estimators served as control conditions because the Bayesian agents were not able to use these ToM-orders. This makes the significant correlation of the fourth-order estimation of the ToM-1 Bayesian agent notable. This estimation is however only significant for the social framing. The game of hide and seek has a very small action space of two choices which can cause the behavior of various ToM-orders to appear to be the same. This could be a possible explanation of the significant correlation between the fourth-order estimation and the Bayesian ToM-1 agent in the social framing. The estimation scores for the third- and fourth-order being insignificant for the other Bayesian ToM-agents confirms the effectiveness of the method. Our findings hold for the game of hide and seek, which has simple rules and actions. Forthcoming experiments need to address the generalization of these findings to more complex situations.

The other focus of this study was estimating the ToM-behavior of participants in the hide and seek game. The behavior of human participants

was most consistent with zero- and second-order theory of mind. Perhaps the most striking result of this study is the lack of first-order theory of mind use by the participants. People are assumed to be perfectly able to think about what other people think so the lack of first-order ToM-use cannot be accounted to the fact that they are unable to perform first-order recursive thoughts. Failure in the reliable use of their ToM-abilities is something that is observed before (Keysar *et al.*, 2003). The shown use of ToM-2 makes it however, very hard to believe they fail to use their ToM-1 abilities. The fact that we found a gap between the zero- and second-order ToM-use could however point towards a possible shortcut people take during simple ToM-situations. Human players are perfectly able to think about what action they themselves should perform. Adding to this the possibility that the opponent knows what the player is about to do results in thinking about what the opponent is assuming the player to think. This is known as second-order theory of mind without the intervening step of first-order theory of mind.

The participants have shown some evidence of opponent modeling abilities. The ToM-0 use seemed to be controllable. Playing a ToM-1 opponent led to a decrease in ToM-0 behavior while playing a RB opponent led to increasing ToM-0 behavior. This interaction corresponds to optimal ToM-behavior. However, the absence of other opponent modeling abilities combined with the lack of first-order ToM use leads to the conclusion that participants do not show optimal ToM-behavior in the hide and seek game.

Why do participants seem to be unable to use an optimal ToM-strategy in such a simple environment? Maybe this phenomenon is actually caused by the environment itself. A frequently repeated game in which a decision has to be made between just two options could be not enough to trigger ToM-behavior in participants. Hide and seek is a game that is purely competitive, the gain of the winner is exactly balanced by the loss of the loser. Ethological debates highlight the importance of competitive versus cooperative social interactions in the evolution of theory of mind (Moll & Tomasello, 2007). The inability to use optimal ToM-strategies could be related to the missing cooperating factor in hide and seek. This

is in line with the theory of Verbrugge (2009) which states that mixed-motive social interactions, with partially shared, partially competing interests are needed to evolve higher-order social cognition.

The inability of the participants to display optimal ToM-behavior could also be explained by the setup of the experiment of Devaine *et al.* (2014a). Participants could only choose an action when they decided within 1300ms. This relatively short reaction time could have caused agents to be unable to think enough about the optimal action. Another important aspect could be the length of the experiment. Every participant played 480 trials of hide and seek, admittedly in two different framings but still the influence of fatigue cannot be ruled out.

Despite these points of criticism, this study proved the usefulness of ToM-estimation based on the principle of reenacting social interactions. Using this method, opponent modeling properties of human participants have been identified. These properties are not always similar to optimal ToM-behavior but this could possibly point towards a shortcut in the human reasoning system.

References

- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6), 716-723.
- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders, (DSM-5®). *American Psychiatric Pub.*
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1), 37-46.
- Bora, E., Yucel, M., & Pantelis, C. (2009). Theory of mind impairment in schizophrenia: Meta-analysis. *Schizophrenia research*, 109(1), 1-9.
- Devaine, M., Hollard, G., & Daunizeau, J. (2014a). The social bayesian brain: Does mentalizing make a difference when we learn? *PLoS computational biology*, 10(12), e1003992.
- Devaine, M., Hollard, G., & Daunizeau, J. (2014b). Theory of mind: Did evolution fool us? *PloS one*, 9(2), e87619.

- Epstein, J. M. (2006). Generative social science: Studies in agent-based computational modeling. *Princeton University Press*.
- Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economic and Statistics*, 92-107.
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431), 459-473.
- Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, 61(1), 139-151.
- Hedden, T., & Zhang, J. (2002). What do you think I think you think? Strategic reasoning in matrix games. *Cognition*, 85(1), 1-36.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 89(1), 25-41.
- Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, 89(2), 191-204.
- Korkmaz, B. (2011). Theory of mind and neurodevelopmental disorders of childhood. *Pediatric research*, 69, 101R-108R.
- Meijering, B., Van Maanen, L., Van Rijn, H., & Verbrugge, R. (2010). The facilitative effect of context on second-order social reasoning. In *Proceedings of the 32nd annual conference of the Cognitive Science Society* (pp. 1423-1428). Cognitive Science Society.
- Meijering, B., Van Rijn, H., Taatgen, N., & Verbrugge, R. (2011). I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. *CogSci, Cognitive Science Society*, 2486-2491.
- Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 639-648.
- Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, 1313-1326.
- Nichols, S., & Stich, S. P. (2003). Mindreading: An integrated account of pretence, self-awareness, and understanding other minds. *Clarendon Press/Oxford University Press*.
- Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 362(1480), 731-744.
- Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in cognitive sciences*, 3(9), 337-344.
- Premack D, & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515-526.
- Schmelz, M., Call, J., & Tomasello, M. (2011). Chimpanzees know that others make inferences. *Proceedings of the National Academy of Sciences*, 108(7), 3077-3079.
- Stiller, J., & Dunbar, R. I. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, 29(1), 93-104.
- Team, R. C. (2000). R Language Definition.
- Verbrugge, R. (2009). Logic and social cognition. *Journal of Philosophical Logic*, 38(6), 649-680.
- de Weerd, H., Broers, E. & Verbrugge, R. (2015). Savvy software agents can encourage the use of second-order theory of mind by negotiators. *Proceedings of the 37th Annual Cognitive Science Society Meeting*.
- de Weerd, H., Verbrugge, R., & Verheij, B. (2013). How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence*, 199, 67-92.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128.
- Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game theory of mind. *PLoS Computational Biology*, 4(12), e1000254-e1000254.