



# THEORY OF MIND IN DIRECT COMPETITION WITH MORE CLASSICAL STRATEGIES

Bachelor's Project Thesis

Jeroen Heidstra, s1816292, j.g.k.heidstra@student.rug.nl,

Supervisors: Trudy Buwalda & Burcu Arslan & Harmen de Weerd

**Abstract:** When we humans play games or interact with other people in our daily lives, we often make use of theory of mind. In a previous study we have seen that it seems beneficial to use theory of mind when playing the game of Limited Bidding (De Weerd, Verbrugge, and Verheij, 2013). Since the theory of mind in this previous study made use of a perfect memory, we wanted to see if theory of mind would also be beneficial when using a more human-like memory using computational cognitive models. The use of these computational cognitive models in this research seem to indicate that not only would it be useful to use ToM in a theoretical environment where the models/agents have unlimited memory, but that it would also be useful when limited to a more human-like memory.

## 1 Introduction

When humans play games or interact with other people in their daily lives, we often make use of a phenomenon called *theory of mind* (Premack and Woodruff, 1978), which is the ability to understand that people have mental states, such as desires, beliefs, knowledge and intentions, and to realize that mental states of others might be different from one's own. The lowest level of theory of mind is the zero-order (ToM0). In this case you only look at the behavior of the other person, you don't try to figure out what the other persons intentions are. With the first-order theory of mind (ToM1), you consider the thoughts, beliefs and intentions of the other person. You try to see the world through the other persons eyes. The last level we look at here is the second-order theory of mind (ToM2). Here you try to reason about what the other person believes about what your intentions and beliefs are. In theory you can extend the theory of mind to as high as a level as you want. Previous experiments where adults played games have shown that adults use higher order theory of mind (Meijering, Van Rijn, Taatgen, and Verbrugge, 2011). In children between 5 and 10 there are results that show, under optimal conditions and in the setting of stories, not games, the use of second-order theory of mind (ToM2) (Perner and Wimmer, 1985). Another study conducted with both adults and children between the ages of 8 and

10 showed that both children and adults are having trouble with applying ToM2 (Flobbe, Verbrugge, Hendriks, and Krämer, 2008). They found that children are almost flawless in a second-order false belief task, but a lot less successful when performing a strategic game task. They concluded that second-order ToM is more difficult to apply than first-order ToM. This was found in both children and adults. This is an interesting finding, there could be different reasons for these results. It could be that learning ToM2 in strategic games is more difficult than in false belief tasks, but it could also be the case that it is less useful to use ToM2 in the strategic games.

For example, when we negotiate to get out of a chore at home, we should reason what the other person's desire is (first-order ToM), and further what the other person thinks what we desire (second-order ToM). In understanding human behaviour it is therefore very useful to understand when we use theory of mind (ToM) and at which level it is useful in different situations.

In the following subsections, we first present the previous research about ToM and strategic games. Subsequently we introduce the cognitive architecture PRIMs.

## 1.1 Previous research

To investigate the evolution of higher-order theory of mind, De Weerd et al. (2013) have studied higher-order theory of mind across four different competitive games, including the Limited Bidding game, and determined the advantage of higher-order ToM agents over their lower-order theory of mind opponents.

To determine the effectiveness of ToM, they simulated computational theory of mind agents, playing against one another. They found that first-order ToM agents perform well when they play against zero-order ToM opponents. Furthermore they found that second-order ToM agents have an advantage over first-order ToM agents, although not as big as an advantage as first-order over zero-order. The advantage of third-order over second-order ToM is found to be only marginal. So it seems that agent performance in the Limited Bidding game clearly shows diminishing returns on higher orders of theory of mind. First-order and second-order theory of mind usage will give an advantage over opponents using a lower level of ToM, but higher orders don't show such an advantage. We can see this in figure 1.1. Here the results from their ToM0 against their ToM1 agents and the results from their ToM1 against their ToM2 agents are compared. The learning speed  $\lambda$  indicates how fast a model learns about the behaviour of the opponent. So when the learning speed is low, it takes longer to determine the ToM level of the opponent and vice versa. We can see in part (a) of the figure that the ToM1 agent had a predominantly positive score when playing against the ToM0 agent. In part (b) we also see that the higher-order ToM agent has an advantage over the ToM agent with the lower-order. However the advantage is less big than in figure 1.1(a).

Because the agents in De Weerd et al. (2013) did not have a human-like memory but a perfect memory, we want to see whether the findings still hold when the agents have a more human-like memory by using a cognitive architecture. Since we as humans don't have a perfect memory, it resembles a more realistic view whether ToM would be useful in this game. We want to test whether a human-like model exhibits the same benefits as the agent model, this is a first step in this direction.

The research question we try to answer in this

paper is: do the findings of (De Weerd et al., 2013) still hold when the agents have a more human-like memory by using a cognitive architecture? That is, is it still beneficial to use theory of mind while playing the Limited Bidding game when using a more human-like architecture? Because of the We hypothesize that it is useful against the simple strategies, but that the second-order will be less advantageous against the simple strategies because of what we have seen in figure 1.1. We also hypothesize that it will still be beneficial to use theory of mind, but to a lesser extent, since there is no more perfect memory.

Before explaining the details of the limited bidding game in Section 2.1, we first introduce the cognitive architecture that we used, namely PRIMs, and we explain the details of the human-like memory of PRIMs.

## 1.2 PRIMs

PRIMs (Taatgen, 2013) stands for PRimitive Information processing eleMents. It expands ACT-R (J. Anderson et al., 2007) by considering knowledge of tasks in a broader context. In both ACT-R and PRIMs, the declarative memory represents the factual knowledge in the form of chunks (in the case of our models: "number '2' comes after number '1'"). A model can have these preprogrammed chunks, but it can also add chunks internally as a result of changes in the implemented buffers, or externally when interacting with the outside world. It uses action/perception modules for this. In PRIMs, these actions/perceptions are generated by the models operators.

The main idea for using cognitive architectures like ACT-R and PRIMs is that they have certain general assumptions about human cognition. They have parameters that are set to a default value to mimic human performance as good as possible. For example, it takes 200 milliseconds to press a button on the keyboard once a decision has been made and the finger is ready to press it. These values are based on psychological experiments and are modeled to have approximately the same results as the real human data. PRIMs uses operators to handle input from different modules. For example, it can have an operator that compares visual input from a screen (for example, which chips are still available in the Limited Bidding game explained in

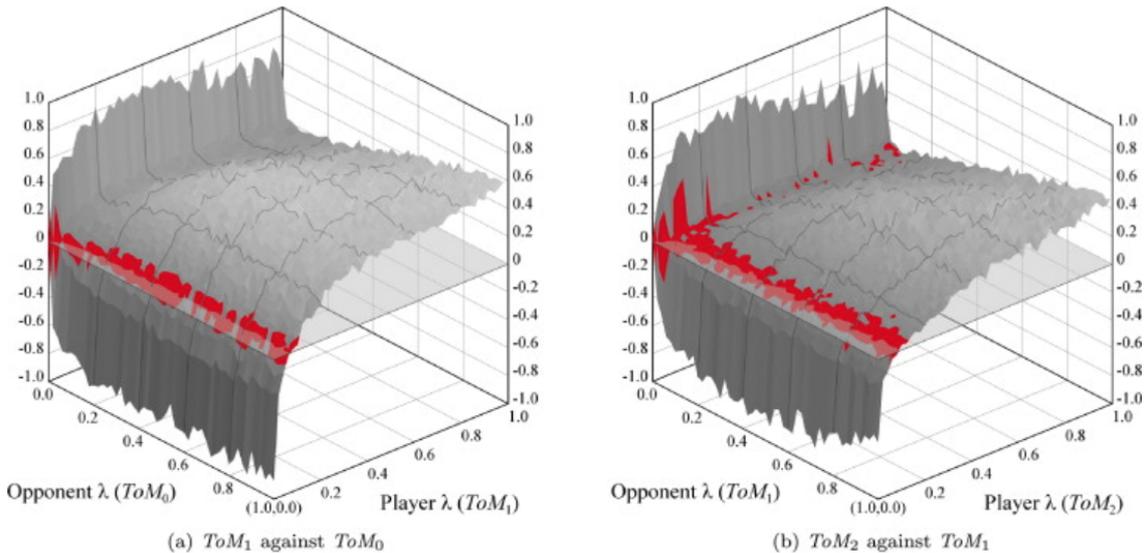


Figure 1.1: Average results of the ToM models against the different opponents. Reprinted with permission from De Weerd et al. (2013).

Section 2.1) to a previous situation in declarative memory (what did I do the last time this situation occurred). Or it can use an operator to use declarative knowledge to count to the next number (if I think the opponent will play a certain chip, what is the chip one higher than that?). To compare the results from (De Weerd et al., 2013), which use models with a perfect memory, this human-like memory is an important feature for choosing to use PRIMs in this research. The models that are used in this paper are explained in Section 2.2.

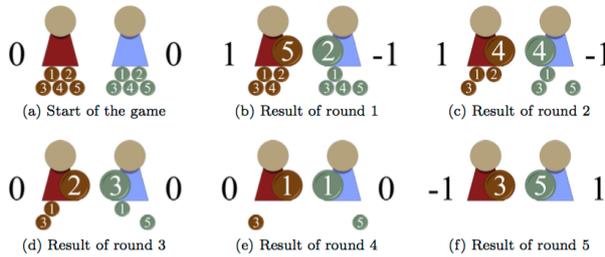
In this paper, a computational cognitive modelling approach is used to test whether the theory of mind models still perform better than simple classical strategies, such as Biased Player, Random Player, Win Stay Lose Shift and Sticky Player, as well as agents using ToM modeled in a conventional programming environment. These strategies are further explained in the Section 2.3. The models were constructed by using a cognitive architecture, PRIMs (Taatgen, 2013). These models are used to show whether there is a benefit of using different levels of theory of mind while playing against other computational models that are using simple classical strategies.

## 2 Methods

### 2.1 Limited Bidding Game

The Limited Bidding game is a two-player game. The game plays over several rounds and is of a competitive nature, in the sense that it is a zero-sum game; there is no possibility for a win-win situation. At the beginning of a game, the players get five chips each. These five chips are numbered one through five. In each round the agent has to play one chip at the same time as your opponent. The chips of both players are revealed and compared. After the round is played, the score is determined and the agent loses the chip he played. The scoring goes as follows: if the agents chip has a higher number than the opponents, the agent wins, when the numbers are the same it's a draw. For winning, the agent gets one point and for losing he loses one point. When there is a draw, no points are awarded. The object of the game is winning as many rounds as possible and lose as few rounds as possible, thus maximizing the amount of points each round. Each chip can be played only once each game, therefore players have to strategically choose which chip it plays in each round. For example, a player selecting a chip with value 1 in the first round will almost always lose this round, but will also lose the lowest

chip and therefore increase the chance of winning in the rounds to come. In the Limited Bidding game it is not possible to win all rounds. A player can at best win four rounds, in which case its opponent wins the remaining round. A maximum winning score is therefore 3 and a maximum losing score -3. Players therefore have to find a good balance between the chip they play in the current round and the chips they will have left in the upcoming rounds. An example of how this game can be played is shown in figure 1. In this example the blue player (the one on the right) has won the game.



**Figure 2.1: An example of the game Limited Bidding. Reprinted with permission from De Weerd et al. (2013).**

## 2.2 The Computational Cognitive Models

We implemented three different levels of ToM models, corresponding to a zero-order, first-order and second-order. These models play the game based on instance-based learning. This means that decisions are made based on the outcomes of previous encounters with the game stored in declarative memory. It is a commonly used mechanism originally developed by (Logan, 1988). Initially, the model has only the numbers in its declarative memory. These chunks look as follows:

```
define facts {
  (next1 next 1 2)
  (next2 next 2 3)
  (next3 next 3 4)
  (next4 next 4 5)
  (next5 next 5 1)
}
```

These fact-types are representing the knowledge the model starts with. They are represented in chunks, for example the fact ‘next1’ has as information that the number ‘2’ follows after the number ‘1’. The model needs this information to figure out what chip it should play if it wants to play one chip higher than ‘1’ for example.

When facing a new situation, an instance-based learning model will look at similar situations it has encountered in the past and attempts to find the best suitable option in the current situation. When using PRIMs to build these models, this means the instance-based learning occurs by adding new chunks to declarative memory. In our example, such a chunk would contain the starting chips, the chip played, and the outcome of the round. At the start of a round the model checks whether he can find the current situation in memory. A chunk will be retrieved if its activation value is higher than the retrieval threshold. When the model can find such a chunk, the new chunk is merged with the identical previous one and their activation values are combined. By updating the contents of declarative memory, the agents can adapt their strategies to suit their opponents.

The zero-order theory of mind (ToM0) plays without thinking about the opponent. When the game begins, there are no previous instances in the memory, so the model plays a random chip. Subsequently, the model saves the instance at the beginning of this round, the chip played and the outcome of the round to its declarative memory. After that, at each start of a round, the model tries to remember a similar situation by comparing the current visual input with a retrieval attempt from the declarative memory where the outcome was a win. If the retrieval is successful, the model plays the chip with the same number as in the retrieval. If the retrieval fails, the model plays a random chip from the available chips.

The first-order theory of mind (ToM1) model was, just like the ToM0 model, build to play the game using its long-term memory to see in each instance what the opponent did in that situation most of the (most recent) times. The main difference between ToM0 and ToM1 models is that the ToM1 model also takes into account the opponent’s decisions and the number of chips, in addition to what it played itself and what chips the model itself still had. In more detail, if a retrieval is successful,

instead of looking at what the model played itself, it looks at what the opponent played and whether the round was lost or not. The model looks for a loss, because this means the opponent won the round and, therefore, it needs to react to the opponents move. Looking at what the opponent played in the retrieved situation, in order to win, the model plays one chip higher than the opponent’s choice in the retrieved situation. However, when the opponent played the chip with five, ToM1 plays the chip with one (chip one is the only chip you can’t win a round with and chip five is the only chip you can’t lose a round with, so playing chip one would be the best case against chip five), so the assumption is that the opponent is ToM0.

The second-order theory of mind (ToM2) model does almost the same as the first order, but instead of playing one higher than the opponent, it tries to figure out what the intentions of the opponent might be. This ToM2 model assumes that it is always playing against a ToM1 opponent, so when the ToM1 model retrieves a situation where it played ‘3’, it assumes it’s opponent knows this and, since the model thinks its opponent is a ToM1 model, plays one higher than this. When the model finished this process of figuring out what the opponent might do in this situation, it can anticipate on that by playing one chip higher than what the model believes his opponent will play.

### 2.3 The opponents

To test how the theory of mind models behave against simple strategies, we have implemented four different of these strategies. ‘Random Player’ this model uses the following strategy: it always plays a random chip. The following opponents use a chance lambda while playing. This chance lambda is the chance that the opponent plays according to its strategy. This lambda is implemented to give a random number from 0 to 1.0. For example for a lambda of 0.6, there is a 60% chance that the opponent plays according to its strategy and a 40% chance that he plays a random chip. The next strategy is ‘Biased Player’, this player is biased to always play the lowest possible chip left. It plays the lowest chip with chance lambda. Then there is the ‘Sticky Player’ he plays the same as his previous play in the same situation with chance lambda. Finally there is the ‘Win Stay Lose Shift’, this opponent plays

the same as the last time in the same situation if it had won. Otherwise he will play another chip.

Beside these simple opponents, we also used the ToM models from (De Weerd et al., 2013) as opponents.

The expectations are that our ToM models will win overall the most games against the simple opponents. For the ToM opponents with the perfect memory from (De Weerd et al., 2013) we expect that our models are not able to win, since they use the same strategy, but the models from (De Weerd et al., 2013) have the perfect memory as advantage.

## 3 Results

The performance of the models described in Section 2.2 has been tested against each of the opponents described in Section 2.3. For each of the combinations model-opponent, there were 50 sets of 300 games each. This represents 50 adults playing 300 games against the same opponent. The score is the combined outcome of all the games together, so  $50 \cdot 300 = 15000$  scores in total. This outcome is the final score of the game, where a positive number represents a win for the ToM model, a score of zero represents a draw and a negative score represents a win for the opponent.

Figure 3.1 shows the average score of the different outcomes when our ToM0 model plays against the simple classical strategies. As expected, the ToM0 model will come out on top in most of the games. It is noticeable that there is a great difference between the distribution between the opponents ‘Biased High’ and ‘Biased Low’ since these are just the opposites of each other. Another thing that stands out is that the opponent ‘Win Stay Lose Shift’ is the only one that defeats the ToM0 model.

Looking at the ToM opponents we can see that for both the ToM1 and ToM2 opponent, the results are that the ToM0 and ToM1 models win most of the time from the ToM opponents and the ToM2 model loses on average from these opponents.

A factorial ANOVA was conducted to compare the main effects of opponent and ToM model and the interaction effect between opponent and ToM model on the outcome of the Limited Bidding game.

A two-way analysis of variance was conducted on the influence of two independent variables (op-

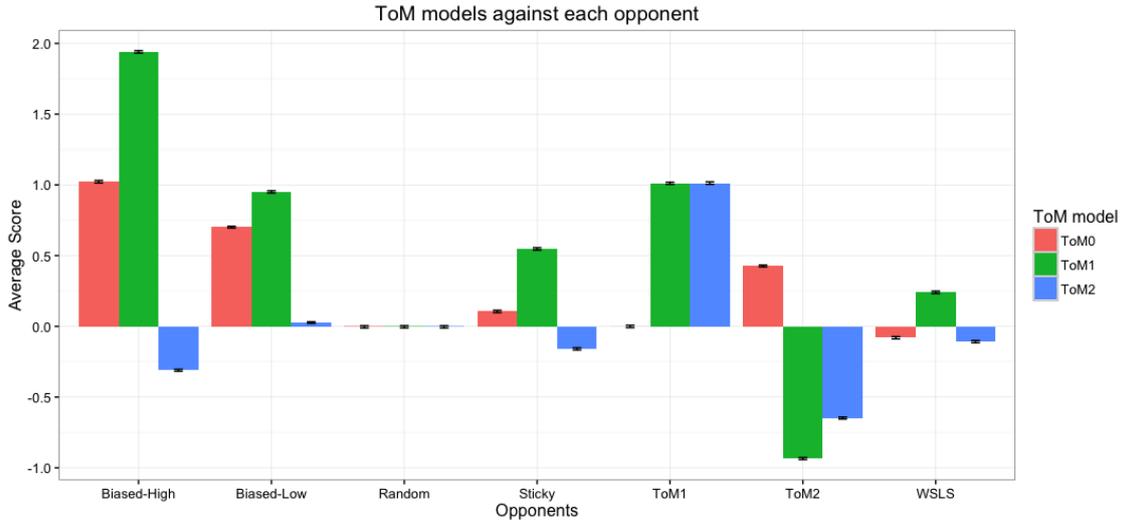


Figure 3.1: Average results of the ToM models against the different opponents.

ponent, ToM level) on the average score in a game of Limited Bidding. Opponent included seven levels (Random, Biased-low, Biased-high, Sticky, Win Stay Lose Shift, ToM1, ToM2) and ToM level consisted of levels (ToM0, ToM1, ToM2). All effects were statistically significant at the .05 significance level. The main effect for ToM level yielded an F ratio of  $F(2, 12) = 8822, p < 2e - 16$ , indicating a significant difference between ToM levels. The main effect for opponent yielded an F ratio of  $F(6, 12) = 9447, p < 2e - 16$ , indicating that the effect for opponent was also significant. The interaction effect was significant,  $F(2, 12) = 5150, p < 2e - 16$ .

## 4 Discussion and future directions

To determine the effectiveness of theory of mind in direct competition with more classical strategies, the ToM models were tested against the simple strategies, namely ‘Random Player’, ‘Biased Player’, ‘Sticky Player’ and ‘Win Stay Lose Shift’, described in Section 2.3. For all of the models, it is the case that the outcomes against the ‘Random’ opponent have an average score around zero, this means they win as much as they lose. This is to be expected, since it is randomly picked whether the model or the opponent wins the round. Therefore

they both win and lose equally likely.

When we look at the ‘Win Stay Lose Shift’ opponent in figure 3.1, we see that this is the only opponent that beat the model in two out of three orders. An explanation for this might be that the ‘Win Stay Lose Shift’ opponent uses less memory than the model, so it will change its strategy, that is ‘stay’ or ‘shift’, immediately if it didn’t result in a win in the previous round. Because the model needs to meet an activation threshold for a memory to be activated, it is possible that it can’t keep up with this fast changing strategy. This could be the case for our ToM0 model, since this plays only with memory and doesn’t look at what the opponent might do. We see that the ToM1 model wins against the ‘Win Stay Lose Shift’. This is what we expected, since the ToM1 tries to reason what the opponent will play, and acts accordingly. Then with our ToM2 model, we see that the ‘Win Stay Lose Shift’ opponent wins this again. This could be because, since the ToM2 model uses only ToM2, the model is thinking one step too much and is therefore playing the wrong chips. All of the other opponents are defeated by the ToM0 model, this is to be expected since the model plays with a memory and these opponents just play according to a fixed strategy. This gives the model the opportunity to figure out the opponents intentions in each round.

Looking at the results in figure 3.1, the ToM1

strategy is a good choice when playing against the extremely simple strategies ‘Biased High’, ‘Biased Low’ and ‘Sticky’. It is clear from looking at the figure that the ToM1 model comes out on top in these cases. It is different when we look at the ‘Win Stay Lose Shift’ outcome. Here again, just like with the results of the ToM0 model, the ‘Win Stay Lose Shift’ is the only opponent that is able to defeat the model. Comparing the results of the ToM models, we can see that, although the ToM1 model still loses, the balance is shifted a bit to the side of the model. So while the opponent still comes out on top, the differences are less big than with the ToM0 model.

Finally, looking at the ToM2 results, we see that this might not be the best strategy to use here. This could be explained by the way the model is implemented. While the higher order ToM models in the case of De Weerd et al. (2013) could also use the lower orders, when appropriate, our model can’t do that. It is set on playing with the ToM2. It could be the case that in this simple game, with limited playing options, the model is ‘overthinking’(where it would be sufficient to think about what the opponent might be thinking about, here the model is thinking about what the opponent might be thinking about what the model might be thinking about), and therefore is not playing the most useful option.

Then, when looking at the ToM opponents, we can see that the ToM0 and ToM1 models win on average from the ToM opponents and that the ToM2 model loses from the ToM opponents. A reason for the poor performance against the ToM2 opponent could be that the ToM2 opponent can switch between whether playing as ToM1 or ToM2 whenever one seems better. Our model doesn’t have this ability and could therefore be at a disadvantage. An indication to support this is the fact that the ToM1 model does win against the ToM opponents. This, however does explain why the ToM1 model wins against a ToM1 opponent, the opponent can’t switch to a lower ToM here, but it doesn’t explain why the ToM1 model wins against a ToM2 opponent. This could be happening because there aren’t enough options in the game to let a ToM2 strategy be useful.

Our main research question was: do the findings of (De Weerd et al., 2013) still hold when the agents have a more human-like memory by using a cogni-

tive architecture? That is, is it still beneficial to use theory of mind while playing the Limited Bidding game when using a more human-like architecture? We hypothesized that it is useful against the simple strategies, but that the second-order would be less advantageous against the simple strategies. So we expected that a ToM strategy would be better at playing the Limited Bidding game than a simple strategy. When looking back at the results, it seems that ToM would overall be a good strategy to use. Although it differs a lot against which opponent the ToM model is playing. We have seen that these results seem to point in the same direction as De Weerd et al. (2013). Looking at figure 1.1, we can see the results from De Weerd et al. (2013). We can see here that the ToM2 against the ToM1 needs a bigger learning speed than the ToM1 against the ToM0. This indicates that, although the ToM2 still has an advantage over ToM1, it is less of an advantage than the ToM1 has over ToM0. Since this study doesn’t use the simple strategies used in this study, we can only compare our ToM models against the ToM opponents we got from De Weerd et al. (2013). When looking at the results in figure 3.1 and the results from De Weerd et al. (2013) in figure 1.1 we can see that in both studies the ToM1 had a clear advantage over the ToM0, in our study the advantage was shown when comparing the results from the average scores of the ToM models against the different opponents (figure 3.1) and in the study from De Weerd et al. (2013) it was shown in direct competition with each other (figure 1.1). The use of computational cognitive models in this research seem to indicate that not only would it be useful to use ToM in a theoretical environment where the models/agents have unlimited memory, but that it would also be useful when limited to a more human-like memory.

For further research, it would be nice to see how the ToM would perform against strategies that are a bit more advanced than the ones used here. Since these are all very simple strategies, apart from the ToM, this maybe wouldn’t be strategies that you would encounter when playing against a human. To further explore the benefits of ToM it would be good to see it play against different, more advanced, opponents. The results seem to suggest that an opponent that is using it’s memory is difficult for the ToM models to play against. Not only could this be a result of the opponent, but also the game that is

used could lead to different results. It would therefore be good to see the ToM used in other games as well. It could very well be that the ToM has more benefits when there are more choices in the game. In this game, after round 3, there are only two choices left. It could be that the ToM2 model has more advantage when the choices aren't this limited. It would also be nice to implement the ToM models further, to make them act even more like ToM. For example it would be nice to add to the ToM2 models the feature that they can also use ToM1 if that is more appropriate for the current opponent or situation. The results now are not in favor of using a ToM2, but adding this feature could mimic the human way of using ToM more. It would be interesting to see if the ToM2 model then would perform better. Also, comparing these results with behavioral data would be nice to validate these results.

D. Premack and G. Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(04):515–526, 1978.

N. Taatgen. The nature and transfer of cognitive skills. *Psychological Review*, 120(3):439, 2013.

## References

H. De Weerd, R. Verbrugge, and B. Verheij. How much does it help to know what she knows you know? an agent-based simulation study. *Artificial Intelligence*, 199:67–92, 2013.

L. Flobbe, R. Verbrugge, P. Hendriks, and I. Krämer. Childrens application of theory of mind in reasoning and language. *Journal of Logic, Language and Information*, 17(4):417–442, 2008.

R. King J. Anderson, Mellon et al. *How can the human mind occur in the physical universe?* Oxford University Press, USA, 2007.

G. Logan. Toward an instance theory of automatization. *Psychological review*, 95(4):492, 1988.

B. Meijering, H. Van Rijn, N. Taatgen, and R. Verbrugge. I do know what you think i think: Second-order theory of mind in strategic games is not that difficult. In *Proceedings of the 33rd annual conference of the cognitive science society*, pages 2486–2491, 2011.

J. Perner and 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*, 39(3):437–471, 1985.