The Advantage of Higher-Order Theory of Mind in the Game of Limited Bidding

Harmen de Weerd, Bart Verheij

Institute of Artificial Intelligence, University of Groningen, P.O. Box 407, 9700 AK, Groningen

Abstract. Higher-order theory of mind is the ability to recursively model mental states of other agents. It is known that adults in general can reason adequately at the second order (covering attributions like "Alice knows that Bob knows that she wrote a novel under pseudonym"), but there are cognitive limits on higher-order theory of mind. For example, children under the age of around 6 cannot correctly apply second-order theory of mind, and it seems to be a uniquely human ability. In this paper, we make use of agent-based models to investigate the advantage of applying a higher-order theory of mind among agents with bounded rationality. We present a model of computational agents in the competitive setting of the limited bidding game, and describe how agents achieve theory of mind by simulating the decision making process of their opponent as if it were their own. Based on the results of a tournament held between these agents, we find diminishing returns on making use of increasingly higher orders of theory of mind.

1 Introduction

Humans, in many aspects, are extraordinary within the animal kingdom. They show an impressive ability to reason about the world around them, as well as about unobservable mental content of others, such as others' knowledge, beliefs and plans. This so-called *theory of mind* [1] is said to be unique to humans [2]. Humans use theory of mind beyond its first-order application, concerning other's propositional attitudes with respect to world facts. They take this ability to a second-order theory of mind, in which they reason about the way others reason about mental content. For example, suppose that Alice is throwing Bob a surprise party. Bob engages in second-order theory of mind when he knows about the party, but is playing along with Alice so she won't find out that he already knows; he may make the second-order attribution "Alice doesn't know that I know".

Although the ability to use higher-order (i.e., at least second-order) theory of mind is well established for humans, both through the attribution of second-order false belief [3] as well as in strategic games [4–6], the use of theory of mind of any kind by non-human species is a controversial matter [see for example 2, 7, 8]. Also, research shows that even human adults have difficulty applying higher-order theory of mind correctly [4, 5, 9].

In this paper, we consider agent-based computational models [10, 11] to investigate the advantages of making use of higher-order theory of mind. The use of computational agents allows us to precisely control and monitor the mental content, including application of theory of mind, of our test subjects. This allows us to investigate the conditions under which a theory of mind would present individuals with an evolutionary advantage over individuals without such abilities. Following the Machiavellian intelligence hypothesis [12, 13], the main driving force behind the evolution of social cognition, such as theory of mind, would be the competitive ability within the species (for a discussion of alternative hypotheses, see [9]). We therefore simulate computational agents in a competitive game, and determine the extent to which higher-order theory of mind provides individuals with an advantage over competitors that are more restricted in their use of theory of mind. In particular, we consider whether the ability to use second-order theory of mind provides individuals with advantages beyond the use of first-order theory of mind.

The setting in which we compare the performance of the computational agents is a newly designed competitive game of limited bidding, which is explained in Section 2. Section 3 gives a detailed description of the way agents that are limited in their ability to explicitly represent mental content are implemented for the limited bidding game. These agents are placed in competition with one another, the results of which are presented in Section 4. We compare the advantages of using second-order theory of mind to those obtained using first-order theory of mind. Finally, Section 5 provides discussion and gives directions for future research.

2 Limited bidding

2.1 Game outline

The limited bidding game (adapted from a game in [14]) is a competitive game played by two players. At the start of the game, each player receives an identical set of N tokens, valued 1 to N. Over the course of N rounds, players simultaneously choose one of their own tokens to use as a 'bid' for the round. Once both players have made their choice, the tokens selected by the players are revealed and compared, and the round is won by the player that selected the highest value token. In case of a draw, there are no winners. The object of the game is to win as many rounds as possible while losing as few rounds as possible. However, each token may be used only once per game. This forces players to plan ahead and strategically choose which of the still available tokens to place as the bid. For example, a player that selects the token with the highest value (N) in the first round will ensure that the first round will not result in a win for his opponent. However, this also means that during the remaining $N-1$ rounds, the token with value N will not be available to this player.

After each round in the limited bidding game, the tokens that were played are announced to each player. That is, at the end of each round, every player not only knows who won the round, but also which tokens were used. This allows players to keep track of the tokens that may still be played by their opponent. Since our computational agents are limited in their ability to make use of theory of mind, these agents do not have the capacity to explicitly represent common knowledge. However, we assume that players do not hold beliefs that would be inconsistent with common knowledge of the rules and dynamics of the limited bidding game.

2.2 A scenario for agents with bounded rationality

Under the assumption of common knowledge of rationality, rational agents play the limited bidding game randomly (see Appendix A), such that during each round, a rational agent randomly plays one of the still available tokens. However, experiments with human subjects have shown contexts in which humans regularly fail to behave as predicted by game theory [e.g. 15–19]. In reality, agents may not be fully rational, or consider their opponent to be fully rational. When agents repeatedly interact with the same opponent, they may show patterns of behaviour that deviate from random play, which may be used to their opponent's advantage. In this section, we tentatively describe the process of playing the limited bidding game by agents that are limited in their application of theory of mind. In the remainder, we will speak of a ToM_i agent to indicate an agent that has the ability to use theory of mind up to and including the i-th order. Also, to avoid confusion, we will refer to agents as if they were male, and opponents as if they were female.

Consider the situation in which a ToM_0 agent meets a ToM_1 opponent for the second time in the setting of the limited bidding game. During the first round of the game, suppose that the ToM_0 agent recalls that his opponent played token 1 in the first round of the last game. When deciding what token to play, a ToM_0 agent cannot make use of any theory of mind. In particular, a ToM_0 agent cannot consider the possibility that his opponent has goals that are competitive to his own. The only information available to the agent is that his opponent sometimes plays token 1 in the first round of the game. Against token 1, the best response is token 2, and thus the ToM_0 agent chooses to play token 2.

The ToM_1 opponent, on the other hand, forms beliefs about what the ToM_0 agent believes. She remembers that the last time she played against the ToM_0 agent, she selected token 1 in the first round. She reasons that if the situation were reversed, and she had been in the ToM_0 agent's position, she would conclude that the best response against token 1 is playing token 2. From this, the ToM_1 opponent concludes that the ToM_0 agent will be playing token 2. Against token 2, the best response is token 3, which is the token that the ToM_1 agent will select to play.

In our setup, none of the agents is aware of the abilities of his opponent. Through repeated interaction, a ToM_1 agent may come to believe that his opponent is not a ToM_0 agent, but that she does not have any beliefs at all, and plays according to some unchanging strategy. Based on this belief, a ToM_1 agents can choose to play as if he were a ToM_0 agent himself. Each agent forms and updates his beliefs through repeated interaction, in an attempt to uncover what order of theory of mind he should use to win the game. The need for such learning becomes apparent for agents that make use of higher-order theory of mind. A ToM_2 agent, for example, engages in second-order theory of mind by forming beliefs about what his opponent believes him to believe. The implicit assumption in this modeling is that his opponent is a ToM_1 opponent that is able to form beliefs about what he believes. When in reality she is a ToM_0 opponent, the ToM_2 agent therefore attributes beliefs to his opponent that she cannot represent.

3 A mathematical model of theory of mind agents

In this section, we discuss the implementation of computational agents that are limited in their ability to make use of theory of mind while playing the limited bidding game, similar to the agents described in Section 2.2.

Computational agents in the limited bidding game represent the game situation by its observable features, that is, the set of tokens T that is still available to the agent and the set of tokens S that is available to his opponent. Based on this representation (T, S), an agent has beliefs in the form of a probability distribution $b^{(0)}$, such that $b^{(0)}(s; T, S)$ represents what the agent believes to be the probability that his opponent will play the token with value s in situation (T, S). A ToM_1 agent furthermore attributes beliefs to his opponent in the form of a distribution $b^{(1)}$, such that he believes his opponent to assign probability $b^{(1)}(t; S, T)$ to the event that he will play token t in situation (T, S). A ToM_2 agent maintains an additional belief structure $b^{(2)}$, such that he believes his opponent to believe that he assigns probability $b^{(2)}(s; T, S)$ to the event of her playing token s in situation (T, S).

Since an agent's beliefs $b^{(i)}$ represent probability distributions, we assume that they are non-negative and normalized such that $\sum_{s \in S} b^{(i)}(s; T, S) = 1$ for all $S \neq \emptyset$ and all orders of theory of mind i. Besides these beliefs, agents are governed by their confidence in the predictions based on application of first- and second-order theory of mind, c_1 and c_2 respectively, as well as learning speed λ and discounting rate δ. Unlike the beliefs $b^{(i)}$ and confidences c_i, an agent's learning speed λ and discounting rate δ, to be discussed later in this section, are fixed and agent-specific traits that are beyond the agent's ability to control.

To decide what token to use, agents make use of three basic functionalities: a value function Φ, a decision function t^* and a belief updating function Δ. The value function Φ is used to obtain a measure of the expected outcome of the game when playing token t in situation (T, S). This is achieved through

$$\Phi_{T,S}(t, b^{(i)}) = \begin{cases} \sum_{s \in S} b^{(i)}(s; T, S) \cdot sgn(t - s) & \text{if } |T| = 1 \\ \sum_{s \in S} b^{(i)}(s; T, S) \Big(sgn(t - s) + \delta \max_{t' \in T \setminus \{t\}} \Phi_{T \setminus \{t\}, S \setminus \{s\}}(t', b^{(i)}) \Big) & \text{if } |T| > 1, \end{cases} \tag{1}$$

where sgn is the signum function. Note that the value function Φ makes use of exponential time discounting with parameter $0 \leq \delta \leq 1$ [20, 21]. A higher value of time discounting δ indicates that the agent is more patient, and more willing to lose the next round if it means winning the game.

Agents use the value function Φ to weigh the likelihood of winning the current round by playing token t against the value of the situation that results from losing token t for the remainder of the game. Based on beliefs $b^{(i)}$, agents decide what token to use according to the decision function

$$t^*_{T,S}(b^{(i)}) = \arg\max_{t \in T} \Phi_{T,S}(t, b^{(i)}). \tag{2}$$

Through application of theory of mind, agents come to believe that their opponent will be playing some token $\hat{s}$. The extent to which ith-order theory of mind governs the decisions of the agent's actions is determined by his confidence $0 \leq c_i \leq 1$ that ith-order theory of mind accurately predicts his opponent's behaviour. For every order of theory of mind available to the agent, he therefore adjusts his beliefs using the belief adjustment function Δ, given by

$$\Delta(b^{(i)}, \hat{s}, c_i)(s; T, S) = \begin{cases} (1 - c_i) \cdot b^{(i)}(s; T, S) & \text{if } s \neq \hat{s} \\ c_i + (1 - c_i) \cdot b^{(i)}(s; T, S) & \text{if } s = \hat{s}. \end{cases} \tag{3}$$

The functions Φ, t^* and Δ are shared by all agents, but the type of agent determines how these functions are used. A ToM_0 agent selects what token to play by using Equation (2) directly. That is, given discounting rate δ and zeroth-order beliefs $b^{(0)}$, a ToM_0 agent faced with situation (T, S) will play token $t^*_{T,S}(b^{(0)})$.

In contrast, ToM_1 agents consider the possibility that their opponent is playing as a ToM_0 agent. A ToM_1 agent makes use of this by determining what token he would play if the situation were reversed. To do so, a ToM_1 agent maintains first-order beliefs $b^{(1)}$ that describe what he would believe in his opponent's situation, and thus what he believes his opponent to believe. Using these beliefs, a ToM_1 agent can estimate what token his opponent will believe him to be playing by calculating $\hat{s}^{(1)} = t^*_{S,T}(b^{(1)})$.

Once a ToM_1 agent has derived what token $\hat{s}^{(1)}$ he would play in his opponent's situation, he adjusts his own beliefs $b^{(0)}$ to represent that he believes his opponent to play $\hat{s}^{(1)}$. That is, using the belief adjustment function Δ, a ToM_1 agent decides what token to use by calculating

$$t^*_{T,S}\Big(\Delta\big(b^{(0)}, \hat{s}^{(1)}, c_1\big)\Big) = t^*_{T,S}\left(\Delta\Big(b^{(0)}, t^*_{S,T}\big(b^{(1)}\big), c_1\Big)\right). \tag{4}$$

Note that in this sense, the computational agents described here represent their theory of mind according to *simulation-theory of mind* [22–24]. That is, rather than forming a *theory-theory of mind* [1, 25] that relates observable features of the world to unobservable mental states of their opponent through explicit hypotheses, agents simulate the mental content of their opponent in their own mind. A ToM_1 agent thus considers the mental states of his opponent by considering her viewpoint as if it were his own, implicitly assuming that this accurately describes her thought process. In this particular setting, this means that a ToM_1 agent makes use of his own discounting rate δ in determining $\hat{s}^{(1)}$, and therefore assumes his opponent to have the same rate of impatience he has.

Similar to the way a ToM_1 agent models his opponent as a ToM_0 agent, a ToM_2 agent determines what token he would play if he were in the position of his opponent, playing as a ToM_1 agent. In order to do so, a ToM_2 agent needs to specify his opponent's confidence in first-order theory of mind. In our experiments, we have assumed that all ToM_2 agents use a value of 0.8 to determine their opponent's behaviour playing as a ToM_1 agent, resulting in the estimate $\hat{s}^{(2)} = t^*_{S,T}\left(\Delta\left[b^{(1)}, t^*_{T,S}(b^{(2)}), 0.8\right]\right)$. This estimate is then used to update the ToM_2 agent's beliefs a second time before he makes his choice of what token to use. This choice can therefore be represented as

$$t^*_{T,S}\left(\Delta\left[\Delta\left(b^{(0)}, \underbrace{t^*_{S,T}(b^{(1)})}_{\hat{s}^{(1)}}, c_1\right), \underbrace{t^*_{S,T}\left(\Delta\left[b^{(1)}, t^*_{T,S}(b^{(2)}), 0.8\right]\right)}_{\hat{s}^{(2)}}, c_2\right]\right). \tag{5}$$

To arrive at his decision of what token to play, an agent makes use of beliefs $b^{(i)}$, which are initialized randomly, and confidence levels c_i, which are initialized at zero. After each round, the actual choices of the agent $\tilde{t}$ and his opponent $\tilde{s}$ are revealed. At this moment, an agent updates his confidence in theory of mind based on the accuracy of its predictions. That is, given his agent-specific learning speed $0 \leq \lambda \leq 1$, a ToM_1 agent updates his confidence in first-order theory of mind c_1 according to

$$c_1 := \begin{cases} (1-\lambda)\cdot c_1 & \text{if } \tilde{s} \neq \hat{s}^{(1)} \\ \lambda + (1-\lambda)\cdot c_1 & \text{if } \tilde{s} = \hat{s}^{(1)}. \end{cases} \tag{6}$$

A ToM_1 agent thus increases his confidence in the use of first-order theory of mind if it yields accurate predictions, and lowers his confidence if predictions are inaccurate. A ToM_2 agent additionally adjusts his confidence in the use of second-order theory of mind c_2 according to

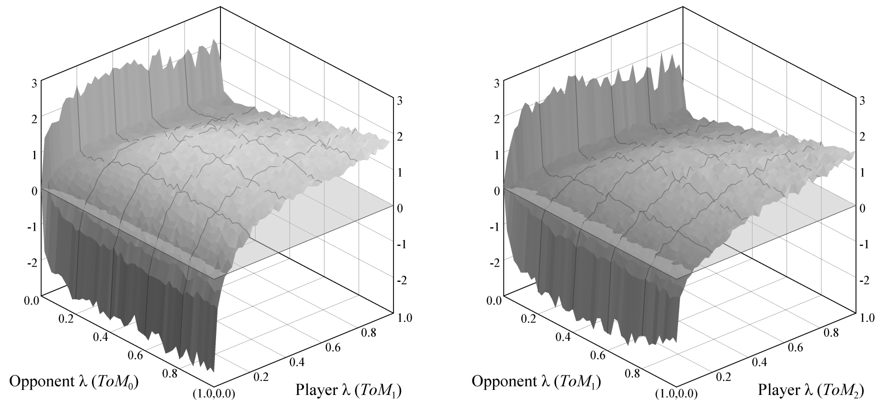
$$c_2 := \begin{cases} (1-\lambda)\cdot c_2 & \text{if } \tilde{s} \neq \hat{s}^{(2)} \\ c_2 & \text{if } \tilde{s} = \hat{s}^{(1)} = \hat{s}^{(2)} \\ \lambda + (1-\lambda)\cdot c_2 & \text{if } \tilde{s} = \hat{s}^{(2)} \neq \hat{s}^{(1)}. \end{cases} \tag{7}$$

This update is similar to the updating of the confidence in first-order theory of mind, except that a ToM_2 agent does not change his confidence in second-order theory of mind when first- and second-order theory of mind both yield correct predictions. That is, a ToM_2 agent only grows more confident in the use of second-order theory of mind when this results in accurate predictions that could not have been made with first-order theory of mind.

Finally, the agent also updates his beliefs $b^{(i)}$. For zeroth- and second-order beliefs $b^{(0)}$ and $b^{(2)}$, an agent updates his beliefs using his opponent's choice $\tilde{s}$, while first-order beliefs $b^{(1)}$ are updated using his own choice $\tilde{t}$, such that

$$b^{(i)}(s;T,S) := \Delta\left(b^{(i)}, \tilde{s}, \lambda\right)(s;T,S) \text{ for } i = 0, 2, \text{ and} \tag{8}$$

$$b^{(1)}(t;S,T) := \Delta\left(b^{(1)}, \tilde{t}, \lambda\right)(t;S,T). \tag{9}$$



(a) Average performance of a focal ToM_1 agent playing against a ToM_0 opponent.

(b) Average performance of a focal ToM_2 agent playing against a ToM_1 opponent.

Fig. 1: Effects of learning speed λ on average performance in a game of 5 tokens. Performance was determined as the average score over 50 trials, for every 0.02 increase of λ in the range $0 \leq \lambda \leq 1$. Discounting rate δ was fixed at 0.9.

The agents described above implicitly assume that their opponents update their beliefs using the same learning speed $0 \leq \lambda \leq 1$ as themselves. Furthermore, equations (8) and (9) maintain the normalization and non-negativity of beliefs, while the confidences c_1 and c_2 remain limited to the range $[0, 1]$. Finally, agents do not update their beliefs and confidence levels after the last round, in which they make the degenerate choice of playing the only token still available to them.

4 Results

The agents described in Section 3 have been implemented in Java and their performance has been tested in competition in a limited bidding game of five tokens. Performance per game was measured as the difference between the number of rounds an agent won and the number of rounds won by his opponent. Note that since it is not possible for an agent to win more than four out of five rounds[1], an agent's game score ranges from -3 to 3. Agents play against each other in trials that consist of 50 consecutive games. An agent's trial score is the average of the agent's game scores over all 50 games in the trial.

Figure 1 shows the advantage of making use of theory of mind as a function of the learning speed of the focal agent (λ_f) and his opponent (λ_o). Higher and lighter areas represent that the focal agent performed better than his opponent, while lower and darker areas show that his opponent obtained a higher average

[1] If an agent wins the first four rounds, the final round will be won by his opponent.

score. To emphasize the shape of the surface, the grid that appears on the bottom plane has been projected onto the surface.

Both figures show that an agent with learning speed $\lambda = 0$ cannot successfully compete with his opponent, and obtains a negative score. Note that in this case, the agent does not learn at all. Instead, he plays according to a fixed strategy, irrespective of his ability to use theory of mind.

Figure 1a shows that ToM_1 agents predominantly obtain a positive score when playing against ToM_0 opponents. The bright area along the line $\lambda_f = \lambda_o$ indicates that this advantage is again particularly high when learning speeds are equal. In this case, the ToM_1 agent's implicit assumption that his opponent has the same learning speed as himself is correct. Surprisingly, Figure 1a shows that even when the ToM_1 agent fails to accurately model his opponent, he will on average obtain a positive score for any learning speed $\lambda_f > 0.08$.

Figure 1b shows that ToM_2 agents obtain an advantage over ToM_1 opponents. However, although Figure 1b shows many of the same features as Figure 1a, such as the brighter area along the line $\lambda_f = \lambda_o$, ToM_2 agents playing against ToM_1 agents obtain a score that is on average 0.5 lower than the score of ToM_1 agents playing against ToM_0 agents. As a result, a ToM_2 agent needs a learning speed of at least $\lambda_f > 0.12$ in order to obtain, on average, a positive score when playing against a ToM_1 agent.

5 Discussion and future research

By making use of agent-based models, we have shown that in the competitive setting of the limited bidding game, the ability to make use of theory of mind presents individuals with an advantage over opponents that lack such an ability. This advantage presents itself even when an agent fails to model his opponent correctly, although an agent that accurately models his opponent obtains more of an advantage than an agent that over- or underestimates the speed at which his opponent learns from past behaviour. In competitive settings like the limited bidding game, there may therefore be an evolutionary incentive that justifies the application of higher-order theory of mind.

Our results also show diminishing returns on higher orders of theory of mind. Concretely, although second-order theory of mind agents outperform first-order theory of mind opponents, the advantage is not as high as for first-order theory of mind agents playing against zeroth-order theory of mind agents. Further evidence suggests that the advantage diminishes quickly for even higher orders of theory of mind (see Appendix B). This could help explain why humans have difficulty applying higher-order theory of mind correctly.

One possible direction for future research presents itself in the form of variable-frame level-n theory [16]. Variable-frame level-n theory expresses theory of mind as levels of bounded rationality, which an agent uses to model the behaviour of his co-player in the setting of a coordination game. An agent makes use of salience to determine what he believes his co-player to believe to be the best course of action, and selects his own action accordingly. In our competitive setting, variable-frame level-n theory could be used to shape an agent's initial beliefs based on the salience of the tokens with the highest and lowest values. This could provide theory of mind agents with additional advantages early in the game.

Although we have shown that the use of theory of mind benefits individuals in the setting of the limited bidding game, in order to represent the beliefs they attribute to others, the higher-order theory of mind agents we describe need additional memory capacity. Based on additional experiments, it seems that the explicit attribution of mental content to competitors presents individuals with advantages beyond those of an increase in memory capacity (see Appendix C). That is, it seems that the advantage obtained by the application of theory of mind cannot be fully explained by an increase in memory capacity.

Acknowledgments

This work was supported by the Netherlands Organisation for Scientific Research (NWO) Vici grant NWO 277-80-001. We would also like to thank Rineke Verbrugge for her valuable comments and advice.

References

[1] Premack, D., Woodruff, G.: Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences **1** (1978) 515–526
[2] Penn, D., Povinelli, D.: On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. Philosophical Transactions of the Royal Society B: Biological Sciences **362** (2007) 731
[3] Perner, J., Wimmer, H.: "John thinks that Mary thinks that...". Attribution of second-order beliefs by 5 to 10 year old children. Journal of Experimental Child Psychology **39** (1985) 437–71
[4] Hedden, T., Zhang, J.: What do you think I think you think?: Strategic reasoning in matrix games. Cognition **85** (2002) 1–36
[5] Flobbe, L., Verbrugge, R., Hendriks, P., Krämer, I.: Children's application of theory of mind in reasoning and language. Journal of Logic, Language and Information **17** (2008) 417–442
[6] Meijering, B., Van Maanen, L., Van Rijn, H., Verbrugge, R.: The facilitative effect of context on second-order social reasoning. In: Proceedings of the 32nd Annual Conference of the Cognitive Science Society, Cognitive Science Society (2010)
[7] Burkart, J., Heschl, A.: Understanding visual access in common marmosets, Callithrix jacchus: Perspective taking or behaviour reading? Animal Behaviour **73** (2007) 457–469
[8] Carruthers, P.: Meta-cognition in animals: A skeptical look. Mind & Language **23** (2008) 58–89
[9] Verbrugge, R.: Logic and social cognition: The facts matter, and so do computational models. Journal of Philosophical Logic **38** (2009) 649–680
[10] Axelrod, R.: The Evolution of Cooperation. Basic Books, New York (1984)

[11] Epstein, J.: Generative Social Science: Studies in Agent-based Computational Modeling. Princeton University Press, Princeton (NJ) (2006)
[12] Byrne, R., Whiten, A.: Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans. Oxford University Press, USA (1988)
[13] Whiten, A., Byrne, R.: Machiavellian Intelligence II: Extensions and Evaluations. Cambridge University Press, Cambridge (1997)
[14] De Bono, E.: Edward de Bono's Super Mind Pack: Expand Your Thinking Powers with Strategic Games & Mental Exercises. Dorling Kindersley Publishers Ltd, London, UK (1998)
[15] McKelvey, R., Palfrey, T.: An experimental study of the centipede game. Econometrica **60** (1992) 803–836
[16] Bacharach, M., Stahl, D.O.: Variable-frame level-n theory. Games and Economic Behavior **32** (2000) 220–246
[17] Stahl, D., Wilson, P.: On players' models of other players: Theory and experimental evidence. Games and Economic Behavior **10** (1995) 218–254
[18] Fehr, E., Gächter, S.: Cooperation and punishment in public goods experiments. American Economic Review (2000) 980–994
[19] Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R.: In search of homo economicus: Behavioral experiments in 15 small-scale societies. The American Economic Review **91** (2001) 73–78
[20] Green, L., Myerson, J.: Exponential versus hyperbolic discounting of delayed outcomes: Risk and waiting time. Integrative and Comparative Biology **36** (1996) 496
[21] Dimitri, N., van Eijck, J.: Time discounting and time consistency. In van Eijck, J., Verbrugge, R., eds.: Games, Actions and Social Software. Texts in Logic and Games (FOLLI subseries of LNCS). Springer Verlag, Berlin (2011)
[22] Davies, M.: The mental simulation debate. Philosophical Issues **5** (1994) 189–218
[23] Nichols, S., Stich, S.: Mindreading: An Integrated Account of Pretence, Self-awareness, and Understanding Other Minds. Oxford University Press, USA (2003)
[24] Hurley, S.: The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading. Behavioral and Brain Sciences **31** (2008) 1–22
[25] Gopnik, A., Wellman, H.: Why the child's theory of mind really is a theory. Mind & Language **7** (1992) 145–171
[26] Osborne, M., Rubinstein, A.: A Course in Game Theory. The MIT press, Cambridge (MA) (1994)
[27] Bicchieri, C.: Common knowledge and backward induction: A solution to the paradox. In: Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge, Morgan Kaufmann Publishers Inc. (1988) 381–393
[28] Von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (NJ) (1944) Commemorative edition 2007.

Appendix A Rational agents in the limited bidding game

In game theory, it is common to make the assumption of common knowledge of rationality [26, 27]. In terms of theory of mind, this means that rational agents possess the ability to make use of theory of mind of any depth or order. In this section, we will explain how rational agents play the limited bidding game under the assumption of common knowledge of rationality.

For simplicity, we consider a limited bidding game of three tokens. In such a game, players decide what token to play at two moments: once at the start of the game, and again once the result of the first round has been announced. Although new information also becomes available after the second round, the choice of which token to play in the third round is a degenerate one; at the start of the third round both players only have one token left. Since both players have the choice of three tokens to play in the first round, there are nine variations of the subgame the agents play at the second round of the game. We first consider what a rational agent will choose to do at the start of the second round.

		Player 2					
		123	132	213	231	312	321
Player 1	123	(0,0)	(0,0)	(0,0)	(-1,1)	(1,-1)	(0,0)
	132	(0,0)	(0,0)	(-1,1)	(0,0)	(0,0)	(1,-1)
	213	(0,0)	(1,-1)	(0,0)	(0,0)	(0,0)	(-1,1)
	231	(1,-1)	(0,0)	(0,0)	(0,0)	(-1,1)	(0,0)
	312	(-1,1)	(0,0)	(0,0)	(1,-1)	(0,0)	(0,0)
	321	(0,0)	(-1,1)	(1,-1)	(0,0)	(0,0)	(0,0)

Table 1: Payoff table for the limited bidding game of three tokens. Each outcome of the game corresponds to a tuple in the table. The first value of the tuple is the payoff for player one, the second is the payoff for player two.

Since every player tries to maximize the number of rounds won and minimize the numbers of rounds lost, at the end of each game, each player receives a payoff equal to the difference between the two. Table 1 lists the payoffs for both players for each possible outcome of the game, where each outcome is represented as the concatenation of the tokens in the order in which the player has played them. Each payoff structure is presented as a tuple (x, y), such that player 1 receives payoff x and player 2 receives payoff y. The subgames that are played at the beginning of the second round are represented as 2-by-2 submatrices, highlighted by alternating background color in Table 1.

Note that whenever the first round of the game ends in a draw, the resulting subgame is a degenerate one. In this case, both players receive zero payoff irrespective of the final outcome. When the first round does not end in a draw, the resulting subgame is a variation on the matching pennies game [28]. This game is known to have no pure-strategy Nash equilibrium. That is, there is no combination of pure strategies such that each player maximizes his payoff given the strategy of its opponent. However, there is a unique mixed-strategy Nash equilibrium in which each player plays each possible strategy with equal probability. If both players play either one of their remaining tokens with 50% probability, neither one of them has an incentive to switch strategies: given that its opponent is playing randomly, a rational agent has no strategy available that will yield a better expected payoff than playing randomly as well.

		Player 2		
		1	2	3
Player 1	1	(0.0,0.0)	(-0.5,0.5)	(0.5,-0.5)
	2	(0.5,-0.5)	(0.0,0.0)	(-0.5,0.5)
	3	(-0.5,0.5)	(0.5,-0.5)	(0.0,0.0)

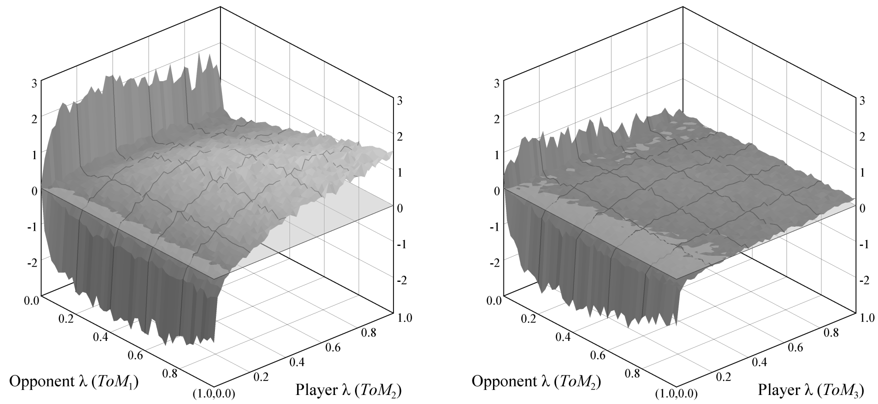
Table 2: Payoff table for the limited bidding game of three tokens once the players have derived that after the first round, both players will play randomly.

Due to the common knowledge of rationality, each player knows that both of them have reached the conclusion that after the first round, they will both play randomly. This means we can rewrite the payoff matrix to reflect the results of each of the subgames, as shown in table 2. Note that this is another variation of the matching pennies game with three strategies, also known as a stone-paper-scissors game [28]. As before, there is no pure-strategy Nash equilibrium, but the unique mixed-strategy Nash equilibrium is reached when both players play each strategy with equal probability. That is, rational agents, under the assumption of common knowledge of rationality, solve the limited bidding game by playing randomly at each round.

This result also holds when the game is played using more than three tokens. That is, to prevent their opponent from taking advantage of any regularity in their strategy, rational agents play the limited bidding game randomly.

Appendix B Limits of the advantage of theory of mind

In Section 4, we have shown that ToM_2 agents can obtain advantages that go beyond those obtained by ToM_1 agents. In this section, we extend the model of Section 3 to allow for ToM_3 agents. These agents possess an additional distribution $b^{(3)}$, such that a ToM_3 agent believes that his opponent believes that he believes her to assign probability $b^{(3)}(t; S, T)$ to him playing token t in situation (T, S). These beliefs are used to determine what token the ToM_3 agent would play if he were in the position of his opponent, and playing as a ToM_2 agent.



(a) Average performance of a focal ToM_2 agent playing against a ToM_1 opponent.

(b) Average performance of a focal ToM_3 agent playing against a ToM_2 opponent.

Fig. 2: Effects of learning speed λ on average performance in a game of 5 tokens. Performance was determined as the average score over 50 trials, for every 0.02 increase of λ in the range $0 \leq \lambda \leq 1$. Discounting rate δ was fixed at 0.9.

The ToM_3 agent considers a 'pure' ToM_2 agent, such that he specifies $c_1 = 0.0$ and $c_2 = 0.8$ for his opponent. The confidence in first-order theory of mind that he believes her to assign to him is $c_1 = 0.8$. This results in the estimate

$$\hat{s}^{(3)} = t^*_{S,T}\Big(\Delta\Big[b^{(1)}, t^*_{T,S}\Big(\Delta\Big[b^{(2)}, t^*_{S,T}(b^{(3)}), 0.8\Big]\Big), 0.8\Big]\Big). \tag{10}$$

This estimate is then used to update the ToM_3 agent's beliefs a third time before he makes his choice of what token to use. This choice therefore is

$$t^*_{T,S}\Big(\Delta\Big\{\Delta\Big[\Delta\Big(b^{(0)}, \underbrace{t^*_{S,T}(b^{(1)})}_{\hat{s}^{(1)}}, c_1\Big), \underbrace{t^*_{S,T}\Big(\Delta\Big[b^{(1)}, t^*_{T,S}(b^{(2)}), 0.8\Big]\Big)}_{\hat{s}^{(2)}}, c_2\Big],$$
$$\underbrace{t^*_{S,T}\Big(\Delta\Big[b^{(1)}, t^*_{T,S}\Big(\Delta\Big[b^{(2)}, t^*_{S,T}(b^{(3)}), 0.8\Big]\Big), 0.8\Big]\Big)}_{\hat{s}^{(3)}}, c_3\Big\}\Big). \tag{11}$$

This agent has been implemented in Java and placed in competition with the ToM_2 agent described in Section 3. The results are shown in Figure 2b. For convenience, the average performance of a ToM_2 agent playing against a ToM_1 opponent has been repeated in Figure 2a. As Figure 2b shows, a ToM_3 agent barely outperforms a ToM_2 agent. The average score only exceeds 0.3 when the ToM_2 opponent has zero learning speed. Although it appears as if a ToM_3 agent can still on average obtain a positive score when his learning speed is at least $\lambda > 0.32$, Figure 2b shows that when the ToM_2 opponent has learning speed $0 < \lambda < 0.1$, performance of the ToM_3 agent may still fall below zero.

Interestingly, the poor performance of ToM_3 agents playing against ToM_2 opponents is partially caused by the model that the ToM_2 opponent holds of the ToM_3 agent. Note that since confidence levels c_i are initialized at zero, all agents start out by playing as ToM_0 agents. When a focal ToM_3 agent is in competition with a ToM_2 opponent, both of them will notice that their predictions based on first-order theory of mind $\hat{s}^{(1)}$ are correct. Through Equation (6), this causes both agents to grow more confident in application of first-order theory of mind. As a result, they both gradually start playing more as a ToM_1 agent. When this happens, predictions based on first-order theory of mind $\hat{s}^{(1)}$ will become less accurate, but predictions based on second-order theory of mind $\hat{s}^{(2)}$ become increasingly accurate, increasing confidence in the application of second-order theory of mind through Equation (7). Both the focal agent and his opponent will therefore start playing as a ToM_2 agent. At this point, the opponent can no longer model the focal agent. That is, she will notice that none of her predictions are correct and start to play as a ToM_0 agent again. However, when the focal agent tries to take advantage of this by playing as a ToM_1 agent, the opponent recognizes this and once again grows more confident in her predictions based on second-order theory of mind. This causes the ToM_2 opponent to constantly keep changing her strategy, which hinders the ToM_3 agent's efforts of trying to model her behaviour.

Appendix C Theory of mind is more than increased memory

The results in Section 4 show that the use of a theory of mind benefits individuals in the setting of the limited bidding game. However, in order to represent the beliefs they attribute to others, the higher-order theory of mind agents we described in Section 3 need additional memory capacity; for every additional order of theory of mind available to the agent, it maintains another belief structure $b^{(i)}$. In this section, we consider the high-memory ToM_0 agent, which has the ability to remember what token $t_{T,S}^{(-1)}$ he played the last time in any game situation (T, S). The high-memory ToM_0 agent makes use of this by representing beliefs of the form $b_{Mem}^{(0)}$, such that he believes that the probability of his opponent playing token s in situation (T, S) is $b_{Mem}^{(0)}(s; T, S, t_{T,S}^{(-1)})$. That is, the high-memory ToM_0 agent has different beliefs concerning what his opponent will play in situation (T, S) based on the last token he played in the same situation.

To determine whether the contribution of theory of mind to an agent's performance can be explained by additional memory alone, we placed the high-memory ToM_0 agent in competition with the ToM_2 agent described in Section 3, both of which have similar demands on memory capacity. The number of game situations in which a player makes a non-trivial choice of what token to play is $\sum_{i=0}^{N-2} \binom{N}{i}^2$. For a game of five tokens, there are 226 such situations. Since a ToM_2 agent needs to maintain three belief structures $b^{(i)}$, a ToM_2 agent needs enough

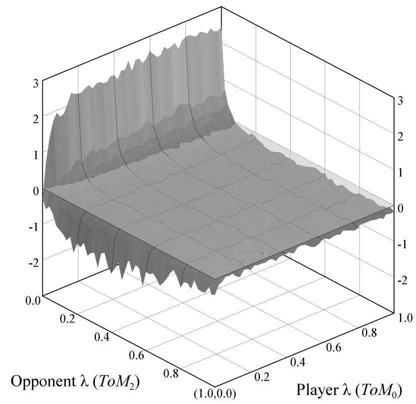


Fig. 3: Effects of learning speed λ on the average performance of a high-memory ToM_0 agent playing against a (low-memory) ToM_2 opponent in a game of 5 tokens. Performance was determined as the average score over 50 trials, for every 0.02 increase of λ in the range $0 \leq \lambda \leq 1$. Discounting rate δ was fixed at 0.9.

memory to represent 678 beliefs. A high-memory ToM_0 agent has a richer representation of the game, which causes him to consider $\sum_{i=0}^{N-2}(N-i)\binom{N}{i}^2$ game situations in which he makes a non-trivial choice of what token to play. In addition to remembering his last choice $t_{T,S}^{(-1)}$ in 226 situations, the high-memory ToM_0 agent therefore needs enough memory to represent 605 beliefs to maintain his belief structure $b_{Mem}^{(0)}$.

Note that the high-memory ToM_0 agent represents an unpredictable opponent for the ToM_2 agent. A ToM_2 agent models the behaviour of his opponent by considering his own actions in her situation. However, the representation of the game situation held by a high-memory ToM_0 agent differs from that of his low-memory ToM_2 opponent. That is, the ToM_2 opponent fails to accurately model the high-memory ToM_0 agent.

Figure 3 show the average performance of a high-memory ToM_0 agent when playing against a low-memory ToM_2 opponent. Surprisingly, even though the ToM_2 opponent is unable to effectively use her theory of mind, she outperforms the high-memory ToM_0 agent whenever her learning speed $\lambda > 0.12$. On average, a high-memory ToM_0 agent scores -0.20 when playing against a ToM_2 opponent.

A possible reason for the negative score of the high-memory ToM_0 agents, even though their ToM_2 opponent is unable to accurately model them, may be the length of the trials. In our setup, trials consist of 50 consecutive games, which may not provide a high-memory ToM_0 agent with sufficient information to gain an advantage over his ToM_2 opponent. In contrast, although the ToM_2 opponent incorrectly models the high-memory ToM_0 agent, her model is accurate enough to obtain a reliable advantage.