



# How much does it help to know what she knows you know? An agent-based simulation study



Harmen de Weerd\*, Rineke Verbrugge, Bart Verheij

*Institute of Artificial Intelligence, Faculty of Mathematics and Natural Sciences, University of Groningen, PO Box 407, 9700 AK Groningen, The Netherlands*

## ARTICLE INFO

### Article history:

Received 2 October 2012

Received in revised form 8 May 2013

Accepted 10 May 2013

Available online 14 May 2013

### Keywords:

Agent-based models

Evolution of theory of mind

## ABSTRACT

In everyday life, people make use of theory of mind by explicitly attributing unobservable mental content such as beliefs, desires, and intentions to others. Humans are known to be able to use this ability recursively. That is, they engage in higher-order theory of mind, and consider what others believe about their own beliefs. In this paper, we use agent-based computational models to investigate the evolution of higher-order theory of mind. We consider higher-order theory of mind across four different competitive games, including repeated single-shot and repeated extensive form games, and determine the advantage of higher-order theory of mind agents over their lower-order theory of mind opponents. Across these four games, we find a common pattern in which first-order and second-order theory of mind agents clearly outperform opponents that are more limited in their ability to make use of theory of mind, while the advantage for deeper recursion to third-order theory of mind is limited in comparison.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

In everyday life, we regularly make use of theory of mind, by reasoning about what other people know and believe. For example, we identify with characters in literature and movies, and accept that they may have beliefs and intentions different from our own. When telling a joke, a speaker engages in higher-order theory of mind, by believing that the hearer knows that the speaker does not intend to convey an actual fact or opinion. In this paper, we make use of agent-based computational models to explain the evolution of our ability to reason about mental content of others.<sup>1</sup>

In settings where humans and computational agents perform actions that influence each other's decision-making process, for example in automated negotiation [3,4], it is necessary to accurately predict the behaviour of others in order to respond appropriately. In artificial intelligence, modeling an opponent explicitly can be achieved through formal approaches such as for example dynamic epistemic logic [5,6], recursive opponent modeling [7], interactive POMDPs [8], networks of influence diagrams [9], game theory of mind [10], or iterated best-response models such as cognitive hierarchy models [11] and level-*n* theory [12,13]. These models allow for recursive modeling of an opponent, by modeling the opponent as an opponent-modeling agent itself, creating increasingly complicated models to predict the actions of increasingly sophisticated opponents. For cognitive agents that are meant to interact with humans, it is important to know whether these formal models of cognition allow for accurate modeling of human reasoning, or whether other models better capture the type of *bounded rationality* exhibited by humans [14,15].

\* Corresponding author. Tel.: +31 50 363 4114; fax: +31 50 363 6687.

E-mail addresses: hdeweerd@ai.rug.nl (H. de Weerd), rineke@ai.rug.nl (R. Verbrugge), b.verheij@ai.rug.nl (B. Verheij).

<sup>1</sup> This research is a continuation of [1,2].

### 1.1. Theory of mind abilities in humans and animals

In humans, the ability to predict the actions of others by explicitly attributing to them unobservable mental content, such as beliefs, desires, and intentions, is known in psychology as *theory of mind* [16]. Experiments in which humans play games show evidence that humans use theory of mind recursively in their decision-making process [17–20]. 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, when asked to search for a hidden object in one of four boxes, participants tend to ignore the most salient box, using their nested belief that a hider would believe that a seeker would consider the most obvious place to search for a hidden object to be a box that stands out [21].

The use of higher-order (i.e. at least second-order) theory of mind allows individuals to make a second-order attribution such as “Alice doesn’t know that Bob knows that she is throwing him a surprise party”. The human ability for higher-order theory of mind is well-established, both through false belief tasks [17,22,23] and strategic games [18–20,24]. However, the use of theory of mind of any kind by non-human species is a controversial matter. Primates [25,26], monkeys [27], but also goats [28], dogs [29] and corvids [30,31] have been proposed to be able to take the mental content of others into account. However, experiments in which animals behave in a way that is consistent with them having a theory of mind are criticized for not being able to distinguish between theory of mind and strategies that do not rely on mental state attribution [32,33]. Opponents of attributing theory of mind to animals posit that the animal could have learned the behaviour through previous experiences, combined with simple mechanisms such as stress [34,35]. Likewise, experiments in which animals fail to show an ability to attribute mental states to others are criticized as well, either for being too complex or ecologically not meaningful [36].

### 1.2. Evolution of theory of mind

The differences in the ability to make use of theory of mind between humans and other animals raise the issue of the reason for the evolution of a system that allows humans to make use of theory of mind recursively, and use higher-order theory of mind to reason about what other people understand about mental content, while other animals, including chimpanzees and other primates, do not appear to have this ability. Furthermore, whereas recursive opponent modeling could continue indefinitely, humans appear to use higher-order theory of mind only up to a certain point [18,19,37]. In an evolutionary sense, the costs of using higher orders of theory of mind may therefore outweigh the benefits.

One of the hypotheses that explain the emergence of social cognition is the Machiavellian intelligence hypothesis<sup>2</sup> [38]. According to the Machiavellian intelligence hypothesis, social cognition allows individuals to make use of deception and social manipulation to obtain an evolutionary advantage over others. If a parallel can be drawn to higher-order theory of mind, the evolution of a higher-order theory of mind would then be favored by giving individuals a competitive advantage over others. This way, the ability to make use of higher-order theory of mind would both be beneficial to individuals that have this trait, as well as detrimental to individuals without such abilities.

In this paper, we aim to test the Machiavellian intelligence hypothesis by making use of agent-based modeling in an attempt to show that there are reasonably natural competitive settings in which higher-order theory of mind is advantageous for agents.

### 1.3. Agent-based modeling

Agent-based modeling is a simulation technique in which individual agents act and interact based on their own perception of their local situation. By explicitly modeling heterogeneity among individual agents, agent-based models can represent systems that are too complex to capture through equation-based modeling approaches. This technique has proven its usefulness as a research tool to investigate how behavioral patterns may emerge from the interactions between individuals (cf. [39,40]). Among others, agent-based models have been used to explain fighting in crowds [41], trust in negotiations [42], the evolution of agriculture [43], the evolution of cooperation and punishment [44–46], and the evolution of language [37,47–49]. In this paper, we consider agent-based computational models to investigate the advantages of making use of higher-order theory of mind. The use of agent-based models allows us to precisely control and monitor the mental content, including application of theory of mind, of our test subjects. This allows us to simulate computational agents in game settings, 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. By varying game settings, this allows us to determine scenarios in which the ability to make use of theory of mind is beneficial to an agent, as well as whether increasingly higher orders of theory of mind provide individuals with increasing advantages over competitors.

To test the Machiavellian intelligence hypothesis, we consider a number of competitive zero-sum games in which we let our computational agents compete to determine whether the ability to make use of higher-order theory of mind is advantageous in a competitive setting. We consider four different games. First, we consider three variations on repeated single-shot

---

<sup>2</sup> For a discussion of alternative hypotheses, see [37].

**Table 1**

(RPS) Payoff table and graph representation for the rock–paper–scissors game. The table shows the payoff for the player choosing the row action ‘rock’, ‘paper’ or ‘scissors’, for every possible choice of the player choosing the column action. Arrows in the graph are read as ‘defeats’. For example, the arrow from ‘paper’ to ‘rock’ means that ‘paper’ defeats ‘rock’.

	Rock	Paper	Scissors
Rock	0	−1	1
Paper	1	0	−1
Scissors	−1	1	0

rock–paper–scissors (RPS) games. The transparent setup of RPS allows us to relate differences in the effectiveness of higher-order theory of mind more easily to the structure of the game. The fourth game is Limited Bidding, which involves planning over multiple rounds of play. We also consider a more complex, extensive form game to judge how well the evolutionary advantage of making use of higher-order theory of mind generalizes across games. The four games are described in detail in Section 2.

Agents may benefit from theory of mind in these games by considering the position of their opponent, and determining what mental content they would have if the roles were reversed. This process is first described intuitively in Section 3. A formal description of the model we use is presented in Section 4. To determine whether the use of theory of mind presents agents with an advantage over opponents without such abilities, we placed agents of different orders of theory of mind in competition with one another. The results of these can be found in Section 5. Finally, Section 6 provides discussion and gives directions for future research.

Throughout this paper, we will be considering agents engaged in a competitive two-player game. To avoid confusion, we will refer to the focal agent or player as if he were male, and his opponent as if she were female.

## 2. Game settings

We investigate theory of mind in four game settings. The games we describe are strictly competitive games in the sense that they are zero-sum games; there is no possibility for a win–win situation in these games. In each of the games we present, each player can guarantee an expected outcome of zero, irrespective of how his opponent plays. That is, the value of each of these games is zero. By playing a mixed strategy, in which the player randomly selects one of the actions he can perform, a player can prevent his opponent from structurally winning the game. However, through repeated games, a player may learn regularities in his opponent’s strategy over time, which he might be able to use to his advantage.

### 2.1. Rock–paper–scissors variations

In the following subsections, we describe the well-known game rock–paper–scissors (RPS), as well as two variations. The rock–paper–scissors game, also known as RoShamBo, is a game settings in which the ability to model an opponent has informally demonstrated its relevance and applicability [50,51]. Although no strategy can consistently defeat an agent that plays RPS randomly, an agent that repeatedly encounters the same opponent in the setting of an RPS game may use regularities in the opponent’s strategy to its advantage. In programming competitions [50], the random strategy only results in an average score. The existence of agents that play according to a non-randomizing strategy allows stronger players to increase their score at the expense of weaker players. The champion of the programming competition in 2000 made use of strategies that detect regularities in the opponent’s behaviour, but also considered the possibility that the opponent was using similar strategies to model the champion’s behaviour [51].

#### 2.1.1. Rock–paper–scissors

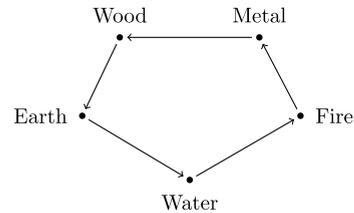
The game of *rock–paper–scissors* (RPS) [52] is a two-player symmetric zero-sum game in which both players simultaneously choose one of the three possible actions ‘rock’, ‘paper’, or ‘scissors’. If both choose the same action, the game ends in a tie. Otherwise, the player that chooses ‘rock’ wins from the one that chooses ‘scissors’, ‘scissors’ wins from ‘paper’, and ‘paper’ wins from ‘rock’. The game can be represented as shown in Table 1, which shows the payoff table and a graph representation for the RPS game. The matrix shows the payoff for the player choosing the row action for every possible choice of the player choosing the column action. In the graph, an arrow from action A to action B denotes the relation ‘A defeats B’.

RPS is known to have a unique mixed-strategy Nash equilibrium (see e.g. [53]) in which the player chooses each of the options with equal probability. That is, when player strategies are known, there is always a player that can improve his expected outcome unless both players play by randomly choosing one of the possible actions. When agents repeatedly play RPS against the same opponent, an agent that randomizes his actions prevents his opponent from taking advantages of regularities in his strategy. However, randomizing also prevents the agent from exploiting regularities in his opponent’s behaviour that may show up over repeated games. By correctly modeling regularities in an opponent’s behaviour, an agent

**Table 2**

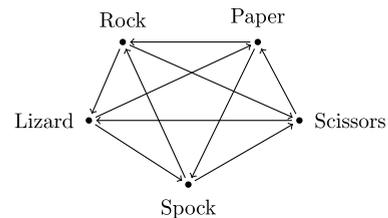
(ERPS) Payoff table and graph representation for the elemental rock–paper–scissors game, in which agents choose between five different actions. The table shows the payoff for the player choosing the row action ‘wood’, ‘metal’, ‘fire’, ‘water’, or ‘earth’, for every possible choice of the player choosing the column action. Arrows in the graph are read as ‘defeats’. For example, the arrow between ‘wood’ and ‘earth’ means that ‘wood’ defeats ‘earth’.

	Wood	Metal	Fire	Water	Earth
Wood	0	−1	0	0	1
Metal	1	0	−1	0	0
Fire	0	1	0	−1	0
Water	0	0	1	0	−1
Earth	−1	0	0	1	0

**Table 3**

(RPSLS) Payoff table and graph representation for the rock–paper–scissors–lizard–Spock game. The table shows the payoff for the player choosing the row action ‘rock’, ‘paper’, ‘scissors’, ‘lizard’ and ‘Spock’, for every possible choice of the player choosing the column action. Arrows in the graph are read as ‘defeats’. For example, the arrow between ‘lizard’ and ‘Spock’ means that ‘lizard’ defeats ‘Spock’.

	Rock	Paper	Scissors	Lizard	Spock
Rock	0	−1	1	1	−1
Paper	1	0	−1	−1	1
Scissors	−1	1	0	1	−1
Lizard	−1	1	−1	0	1
Spock	1	−1	1	−1	0



can increase his score at the expense of his opponent. Experimental evidence suggests that human participants are poor at generating random sequences [54,55], and play RPS in a non-random way [56,57].

We expect that the ability to make use of theory of mind will present an agent with an advantage over opponents without such abilities. The champion of the first international RPS programming competition in 2000 made use of a strategy that detected regularities in the opponent’s behaviour, but also considered the possibility that the opponent was using a similar strategy to model the behaviour of the champion’s [51]. That is, this program engaged in theory of mind, by attributing the intention to win the game to his opponent. However, due to the limited action space, there may be a limit to the effectiveness of theory of mind that is specific to this particular game.

### 2.1.2. Elemental rock–paper–scissors

Although the simple structure of RPS is appealing, the limitation to three actions may influence the effectiveness of higher-order theory of mind. To address this issue, we also consider *elemental rock–paper–scissors* (ERPS). ERPS extends RPS such that it includes the five actions ‘wood’, ‘metal’, ‘fire’, ‘water’, and ‘earth’, as shown in Table 2. The ERPS game preserves the property of RPS that each action is defeated by exactly one response. That is, for each action that an opponent may play, there exists a unique best response that guarantees a positive outcome for the agent.<sup>3</sup>

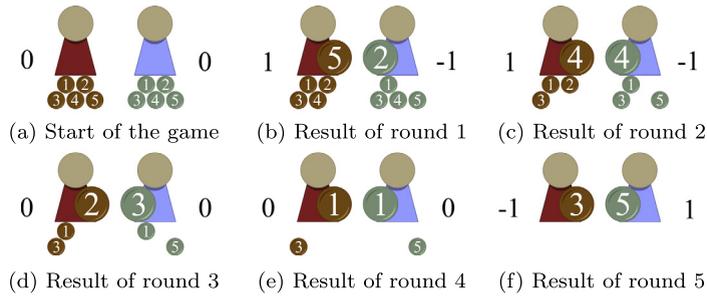
As in the case of RPS, the unique mixed-strategy Nash equilibrium for ERPS is to randomize over all possible actions. However, due to the increased action space, ERPS may have an increased support for theory of mind. That is, we expect theory of mind agents to perform at least as well on ERPS as they would in RPS. Moreover, any differences in the performance of theory of mind agents playing ERPS, compared to those playing RPS, can be attributed to the differences in the structure of the games. In particular, increased performance of higher-order theory of mind agents in ERPS indicates that a limited action space influences the effectiveness of theory of mind.

### 2.1.3. Rock–paper–scissors–lizard–Spock

*Rock–paper–scissors–lizard–Spock* (RPSLS) [58] is an extension of RPS, which adds the actions ‘lizard’ and ‘Spock’ to the actions ‘rock’, ‘paper’, and ‘scissors’ from RPS. Like ERPS, RPSLS has five actions, but in RPSLS each action wins from exactly two other actions, while being defeated by the remaining two other actions. Table 3 shows the payoff matrix and a graph representation of the RPSLS game.

Unlike the previous two games, the best response to an action in RPSLS is not unique. This means that when an agent attributes mental content to his opponent, this does not result in a clear prediction of opponent behaviour. An agent that predicts his opponent to play ‘paper’ has no preference for playing either ‘scissors’ or ‘lizard’, since either will defeat ‘paper’ equally well. As a result, an agent that believes his opponent to believe that the agent will play ‘paper’, will predict that

<sup>3</sup> The payoffs in elemental rock–paper–scissors are based on the overcoming cycle of elements in the Chinese philosophy Wu Xing.



**Fig. 1.** Example of the way limited bidding is played. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

she is going to play either ‘scissors’ or ‘lizard’. Similarly, the opponent’s behaviour in RPSLS is less informative than in RPS and ERPS. After all, if an agent plays ‘paper’, it may have believed that his opponent would play ‘rock’, but also that she would play ‘Spock’. As a result, playing RPSLS repeatedly against the same opponent provides less information about the opponent’s behaviour than in RPS and ERPS. Due to this increased difficulty in modeling the opponent, we expect that theory of mind agents perform more poorly in RPSLS than in RPS and ERPS.

### 2.2. Limited Bidding

Unlike rock–paper–scissors and the two variations on it we presented in the previous subsections, *Limited Bidding* (LB)<sup>4</sup> is a game that plays across several rounds. When the game starts, each player is handed an identical set of 5 tokens each, valued 1 to 5. Over the course of 5 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 tokens that are still available to them they should place as the bid. For example, a player that selects the token with value 5 in the first round will make sure that the first round will not result in a win for his opponent. However, this also means that for the remaining 4 rounds, the token with value 5 will not be available to this player. Players therefore have to weigh the additional probability that they will win the current round against the loss of competitive strength in later rounds that results from using a higher valued token. Fig. 1 shows an example of the way LB is played. In this case, the game is won by the light blue player on the right.

Note that in LB, it is not possible to win all the rounds. Instead, any player can win a maximum of four rounds, in which case the last round is won by his opponent. As a result, a player can achieve a maximum score of 3 in LB. As for the variations on RPS described earlier, a player can prevent his opponent from winning the game. He can do so by randomly choosing to play one of the tokens still available to him at each round of the game. Averaged over repeated games, this mixed strategy of randomizing over all available choices will result in a score of zero for both the player and his opponent.

### 2.3. Rational players

In game theory, it is common to make the assumption that every player is rational, and that this fact is known by all players. Moreover, players are assumed to know that everyone knows that every player is rational, continuing in this fashion ad infinitum. In terms of theory of mind, this *common knowledge of rationality* [60,61] means that players possess the ability to make use of theory of mind of any depth or order. In this section, we will explain how rational players 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.

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 4 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 4.

<sup>4</sup> Limited Bidding is an adaptation of a game presented in [59].

**Table 4**

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.

Player 1	Player 2					
	123	132	213	231	312	321
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 5**

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.

Player 1	Player 2		
	1	2	3
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)

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 [62]. 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 his 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 his opponent is playing randomly, a rational agent has no strategy available that will yield a better expected payoff than playing randomly as well.

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 5. Note that this is a variation of the rock–paper–scissors game. 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.

#### 2.4. Hypotheses about the effectiveness of theory of mind

In this section, we described four different games: rock–paper–scissors, elemental rock–paper–scissors, rock–paper–scissors–lizard–Spock and Limited Bidding. In the game of rock–paper–scissors, agents choose from the three possible actions, each of which is defeated by exactly one of the other actions. The game of elemental rock–paper–scissors resembles RPS in that each action is defeated by exactly one other action, but agents playing ERPS have five different actions to choose from. Rock–paper–scissors–lizard–Spock allows for five different actions as well, but unlike ERPS, each action is defeated by exactly two other actions. Finally, Limited Bidding is a game that spans several rounds, in which agents decide the order in which they play tokens from an initial set of five.

The game of RPS serves as a transparent base scenario to determine whether theory of mind benefits agents in competitive settings. We expect that the ability to make use of theory of mind is advantageous in competitive settings. Specifically, we expect that the ability to make use of higher orders of theory of mind allows an agent to outperform an opponent that is of a lower order of theory of mind in the game of rock–paper–scissors. In the remainder, we will refer to this expectation as hypothesis  $H_{RPS}$ .

The small number of actions that agents choose from in RPS may limit the effectiveness of higher-order theory of mind. The ERPS game, in which agents have a larger action space, addresses this issue. We expect that the larger action space in the ERPS game allows higher-order theory of mind agents to outperform opponents of a lower order of theory of mind at least as well as in the RPS game, which we will refer to as hypothesis  $H_{ERPS}$ .

Agents make use of theory of mind to model the opponent in an attempt to predict her behaviour. As a result, theory of mind is likely to be more effective when the opponent is more predictable. We therefore selected the RPSLS game, in which there is no unique best-response to each action, which should make opponent behaviour harder to predict. Hypothesis

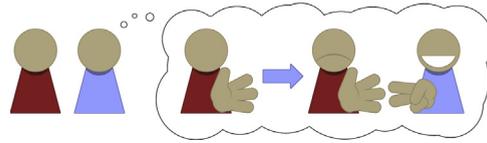


Fig. 2. Example of a possible thought process of a zero-order theory of mind agent.

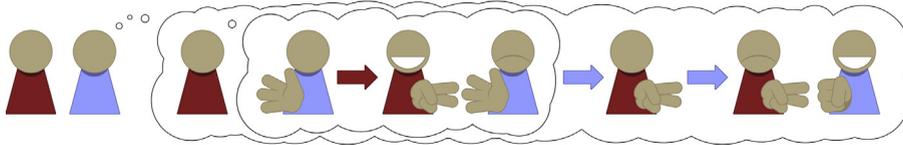


Fig. 3. Example of a possible thought process of a first-order theory of mind agent.

$H_{RPSLS}$  states that we expect theory of mind agents to have difficulty predicting their opponent in RPSLS, and to perform more poorly compared to performance in RPS and ERPS.

Limited Bidding is a multi-stage game, and represents a more complex situation than the single-shot games RPS, ERPS, and RPSLS. This game has been selected to determine whether theory of mind is advantageous, and whether the results from simple single-shot games translate towards a more complex setting. We expect that the results in Limited Bidding may be quantitatively different to those of rock–paper–scissors, but that the results will be qualitatively similar; we will call this hypothesis  $H_{LB}$ .

### 3. Playing the games using simulation-theory of mind

In the games described in Section 2, players can prevent their opponent from winning the game by playing randomly. However, this strategy not only prevents an agent from losing the game from his opponent, but also prevents the agent from winning the game for himself. As a result, the randomizing strategy only results in an average score in the RoShamBo programming competitions [50] discussed in Section 2.1. An agent that believes its opponent to play in a non-random way may try to predict the opponent's behaviour to take advantage of regularities in its opponent's strategy, and win the game.

For humans, one way of generating predictions of opponent behaviour is by using simulation-theory of mind [63–65]. In simulation-theory of mind, a player takes the perspective of its opponent, and determines what its own decision would be if the player had been in the position faced by its opponent. Using the implicit assumption that the opponent's thought process can be accurately modeled by its own thought process, the player then predicts that the opponent will make the same decision the player would have made if the roles were reversed.

In this section, we describe the intuition behind the process of perspective-taking for agents that differ in their abilities to explicitly model mental states, and illustrate how this affects their choices in playing RPS. In Section 4, this intuition is described in a computational model. In the remainder, we will speak of a  $ToM_k$  agent to indicate an agent that has the ability to use theory of mind up to and including the  $k$ -th order, but not beyond.

#### 3.1. Zero-order theory of mind

A zero-order theory of mind ( $ToM_0$ ) agent is unable to model the mental content such as beliefs, desires and intentions of his opponent. In particular, a  $ToM_0$  agent is unable to represent that his opponent has goals that are different from his own goals. When predicting his opponent's behaviour, the agent is limited to his memory of previous events. The  $ToM_0$  agent is intended to model an inexperienced or frustrated player, who only consider his opponent's behaviour rather than thinking about the way she reacts to his actions.

A  $ToM_0$  agent believes that what happened in the past is a good predictor for what is going to happen in the future. This reflects human players' tendency to interpret repetition as indicative for a pattern [66]. Fig. 2 illustrates a possible thought process of a  $ToM_0$  agent. If a  $ToM_0$  agent remembers that his opponent mostly played 'paper' in previous RPS games, he concludes that his opponent is most likely to play 'paper' in the next game. Given this belief, the  $ToM_0$  agent would therefore adjust his behaviour to play 'scissors'.

#### 3.2. First-order theory of mind

In contrast to a  $ToM_0$  agent, a first-order theory of mind ( $ToM_1$ ) agent considers the possibility that his opponent is trying to win the game for herself, and that she reacts to the choices made by the  $ToM_1$  agent. To predict his opponent's behaviour, the  $ToM_1$  agent puts himself in the position of his opponent, and considers the information available to him from her perspective. Fig. 3 shows an example of such a thought process. Suppose that the  $ToM_1$  agent remembers that he mostly played 'paper' in previous RPS games against the same opponent. He realizes that if the roles were reversed, and

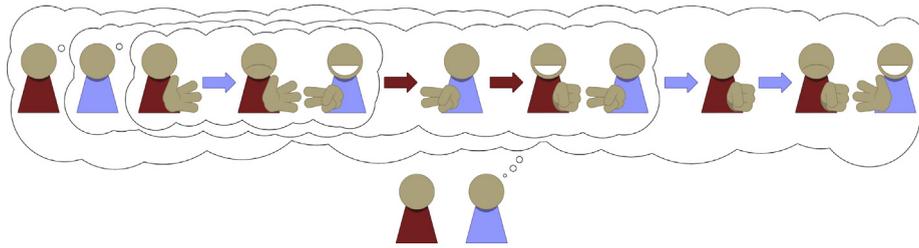


Fig. 4. Example of a possible thought process of a second-order theory of mind agent.

he would remember that his opponent mostly played ‘paper’, he would conclude that his opponent would most likely be playing ‘paper’ again, and that he should play ‘scissors’. The  $ToM_1$  agent has the ability to attribute this thought process to his opponent, and predict that she is likely to play ‘scissors’. Given this prediction, the  $ToM_1$  agent should play ‘rock’.

Although the  $ToM_1$  agent models his opponent as being able use zero-order theory of mind, agents in our setup do not know the extent of the abilities of their opponents with certainty. Through repeated interaction, a  $ToM_1$  agent may come to believe that his opponent is not a  $ToM_0$  agent, and that she has no beliefs at all. Such an opponent without beliefs could, for example, play ‘rock’ irrespective of what the  $ToM_1$  agent has previously played. Based on this belief, a  $ToM_1$  agent can choose to play as if he were a  $ToM_0$  agent, and follow a thought process such as the one presented in Fig. 2.

### 3.3. Second-order theory of mind

Just as a  $ToM_1$  agent models his opponent as having a zero-order theory of mind, a second-order theory of mind ( $ToM_2$ ) agent models a  $ToM_1$  opponent. That is, a  $ToM_2$  agent considers the possibility that his opponent is putting herself in his position, and is modeling him as a  $ToM_0$  agent. Fig. 4 depicts a possible though process of a  $ToM_2$  agent. If the  $ToM_2$  agent remembers his opponent to have mostly played ‘paper’ in previous encounters, he would believe his opponent to predict that he will be playing ‘scissors’ more often. As a result, the  $ToM_2$  agent would predict his opponent to play ‘rock’ more often, in which case the agent should play ‘paper’ more often himself.

Each additional order of theory of mind allows for deeper recursion of mental state attribution. Note that each order of theory of mind represents an additional model for opponent behaviour, with a corresponding prediction. A  $ToM_2$  agent therefore considers three predictions of his opponent’s behaviour, based on the application of zero-order, first-order, and second-order theory of mind.

## 4. Model

We implemented computational agents that make use of simulation-theory of mind, and play similarly to intuitive description of Section 3. In this section, we discuss the implementation of these agents, which play the competitive games described in Section 2. The agents presented here differ in their ability to explicitly represent beliefs, and therefore in their ability to make use of theory of mind.

### 4.1. Representation of the games

In the model we discuss, a game is a tuple  $\mathcal{G} = \langle \mathcal{N}, \mathcal{S}, \mathcal{A}, T, \pi \rangle$ , where:

- $\mathcal{N} = \{i, j\}$  is the set of agents, where  $i$  denotes the focal agent, and  $j$  denotes his opponent;
- $\mathcal{S}$  is the set of possible states of the game;
- $\mathcal{A} = \mathcal{A}_i \times \mathcal{A}_j$  is the set of possible action pairs, where  $\mathcal{A}_i$  is the set of actions that can be performed by the agent  $i$ , and  $\mathcal{A}_j$  is the set of actions that can be performed by his opponent  $j$ ;
- $T$  is a partial transition function  $T: \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$ , which describes the results of the pair of actions of the focal agent and his opponent on the game state; and
- $\pi = (\pi_i, \pi_j)$  is the pair of payoff functions  $\pi_i, \pi_j: \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ .

An instance of a game as played by two players then consists of an initial game state and a sequence of action pairs. For example, in Limited Bidding, each game state  $s \in \mathcal{S}$  encodes the tokens that are still available to the agent, as well as those still available to his opponent. This allows agents that are playing LB to distinguish between individual rounds, and make their beliefs concerning the opponent’s gameplay conditional on the tokens that can still be played.

Fig. 5 shows a representation of an instance of an LB game, which corresponds to the example game shown in Fig. 1. In Fig. 5, states are represented by boxes, while arrows show the state transitions. Each state transition shows the action pair that caused the transition, as well as the payoff  $\pi_i$  for the focal agent and the payoff  $\pi_j$  for his opponent. The game starts in the initial state in which both the agent and his opponent each have an identical set of five tokens. The actions that the

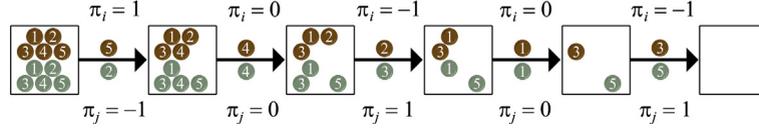


Fig. 5. Representation of the instance of the Limited Bidding game shown in Fig. 1. The figure shows the action pairs that transition the game from the initial state to the final state.

agent can perform correspond to selecting one of the tokens that is still available to him. Once the agent and his opponent have selected an action, the game transitions to a new state, and both players receive a payoff. This process is repeated until a final state is reached in which no combination of actions leads to a change in game state. For Limited Bidding, this final state is the situation after five rounds, when there are no more tokens to play.

The game states  $\mathcal{S}$  are intended to model the different stages of a multi-stage game such as Limited Bidding. Single-shot games, such as the variations on rock–paper–scissors described in Section 2, can be represented as a game that contains two states  $\mathcal{S} = \{s_0, s_1\}$ , such that the game transitions from the start state  $s_0$  to the end state  $s_1$  through any possible action pair  $(a_i, a_j) \in \mathcal{A}$ . That is,  $T(s_0, (a_i, a_j)) = s_1$  for all  $(a_i, a_j) \in \mathcal{A}$ .

Additional to transitioning to a new game state, agents receive a payoff based on their actions. The payoff functions  $\pi_i, \pi_j: \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$  determine the payoff  $\pi_i(s, (a_i, a_j))$  for the focal agent  $i$ , and payoff  $\pi_j(s, (a_i, a_j))$  for his opponent for each combination of game state  $s \in \mathcal{S}$  and action pair  $(a_i, a_j) \in \mathcal{A}$ . Note that since we consider only zero-sum games here,  $\pi_i(s, (a_i, a_j)) = -\pi_j(s, (a_i, a_j))$ .

#### 4.2. Zero-order theory of mind agents

In the present setup, we assume that agents understand the game. Furthermore, we assume no agent considers the possibility that his opponent does not understand the game, or that his opponent believes that he does not understand the game, continuing in this fashion ad infinitum. This means that, for example, none of the agents considers it possible that his opponent will perform an action that is not in her action space  $\mathcal{A}_j$ . Similarly, no agent believes that his opponent considers it a possibility that he himself will play an action that is not in his action space  $\mathcal{A}_i$ .

Note that this assumption is similar to the assumption that the rules and dynamics of the game are *common knowledge* [5,67,68], which requires that each agent understands the game, and knows that his opponent understands the game, and so forth. However, the simulated agents described here are limited in their ability to make use of theory of mind, and may not be able to represent their opponent's mental content. That is, in this case it is not possible to assume that the rules and dynamics of the game are common knowledge. Instead, we assume that no agent has beliefs that conflict with common knowledge of the rules and dynamics of the game.

Agents form beliefs  $b^{(0)}$  in the form of a probability distribution over the opponent's actions  $\mathcal{A}_j$  for every game state, such that  $b^{(0)}(a_j; s)$  represents what the agent believes to be the probability that his opponent will play action  $a_j \in \mathcal{A}_j$  given that the game is in situation  $s \in \mathcal{S}$ . We assume:

$$b^{(0)}(a_j; s) \geq 0 \quad \text{for all } a_j \in \mathcal{A}_j, s \in \mathcal{S}. \quad (1)$$

$$\sum_{a_j \in \mathcal{A}_j} b^{(0)}(a_j; s) = 1 \quad \text{for all } s \in \mathcal{S}. \quad (2)$$

That is, (1) agents assign non-negative probability to their opponent playing a certain action in a certain game state, and (2) the probabilities assigned to each possible opponent action sum up to 1 for each possible game state.

For a  $ToM_0$  agent, the belief structure  $b^{(0)}$  represents the extent of his beliefs concerning his opponent's behaviour. Given the game's payoff function and his beliefs about the way his opponent plays the game, a  $ToM_0$  agent is able to assign a subjective value  $\Phi_i(a_i; b^{(0)}, s)$  to playing a certain action  $a_i \in \mathcal{A}_i$  in game state  $s \in \mathcal{S}$ , given his beliefs  $b^{(0)}$  concerning his opponent's behaviour. To determine this value, the agent considers how likely he considers it to be that his opponent is going to play some action  $a_j \in \mathcal{A}_j$ . If the opponent would play  $a_j$ , playing  $a_i$  would yield the agent an immediate payoff  $\pi_i(s, (a_i, a_j))$ , but it would also cause the game to move forward, and end up in a new state  $s' = T(s, (a_i, a_j))$ . The agent takes this into account by planning ahead, and determining the maximum value he can achieve when the game reaches state  $s'$ . The combination of immediate payoff  $\pi_i(s, (a_i, a_j))$  and the maximum value that can be achieved in the state  $T(s, (a_i, a_j))$  are weighted by what the agent believes to be the probability  $b^{(0)}(a_j; s)$  that his opponent is actually going to play action  $a_j$ . The value  $\Phi_i(a_i; b^{(0)}, s)$  that the focal agent  $i$  assigns to playing action  $a_i$  in game state  $s$ , based on his belief  $b^{(0)}$  concerning his opponent's behaviour, is given by

$$\Phi_i(a_i; b^{(0)}, s) = \sum_{a_j \in \mathcal{A}_j} b^{(0)}(a_j; s) \cdot \left( \pi_i(s, (a_i, a_j)) + \max_{a' \in \mathcal{A}_i} \Phi_i(a'; b^{(0)}, T(s, (a_i, a_j))) \right). \quad (3)$$

**Table 6**  
Possible mental contents of a  $ToM_0$  agent and a  $ToM_1$  agent in a game of rock–paper–scissors.

(a) Mental content of the $ToM_0$ agent in Example 1.		(b) Mental content of the $ToM_1$ agent in Example 2.	
Order of theory of mind $k$		Order of theory of mind $k$	
	0	0	1
$b^{(0)}(R, s_0)$	0.5	$c_k$	0.9
$b^{(0)}(P, s_0)$	0.3	$b^{(k)}(R, s_0)$	0.5 0.4
$b^{(0)}(S, s_0)$	0.2	$b^{(k)}(P, s_0)$	0.3 0.5
		$b^{(k)}(S, s_0)$	0.2 0.1

We assume that agents choose rationally given their beliefs. That is, agents choose to play the action  $a_i$  that maximizes the value function. This is represented by the decision function  $t_i^*$ , given by

$$t_i^*(b^{(0)}; s) = \arg \max_{a_i \in \mathcal{A}_i} \Phi_i(a_i; b^{(0)}, s). \quad (4)$$

**Example 1.** Consider an agent that plays rock–paper–scissors against his opponent. The RPS game consists of two states  $S = \{s_0, s_1\}$ , where the first state  $s_0$  represents the start of the game, and the second state  $s_1$  is the end of the game. The action spaces of the agent and his opponent are the same,  $\mathcal{A}_i = \mathcal{A}_j = \{R, P, S\}$ . The transition function  $T$  is defined such that  $T(s_0, a_i, a_j) = s_1$  for all  $a_i \in \mathcal{A}_i$ , and  $a_j \in \mathcal{A}_j$ . The payoffs in state  $s_0$  are given by Table 1, while payoffs are zero in state  $s_1$ .

We consider a  $ToM_0$  agent, whose mental content is listed in Table 6a. The agent's zero-order beliefs  $b^{(0)}$  indicate that the agent believes that there is a 50% probability that his opponent is going to play  $R$ , a 30% probability that his opponent is going to play  $P$ , and a 20% probability that his opponent is going to play  $S$ . Based on these zero-order beliefs, the agent can determine the value for each of the actions  $R$ ,  $P$ , and  $S$ , based on the expected payoff. For example, the agent believes that if he plays  $R$ , there is a 30% probability that he will lose because his opponent played  $P$ , and a 20% probability that he will win because his opponent played  $S$ . This results in the following values:

$$\begin{aligned} \Phi_i(R; b^{(0)}, s_0) &= b^{(0)}(R; s_0) \cdot \pi_i(s_0, R, R) + b^{(0)}(P; s_0) \cdot \pi_i(s_0, R, P) + b^{(0)}(S; s_0) \cdot \pi_i(s_0, R, S) \\ &= 0.5 \cdot 0 + 0.3 \cdot (-1) + 0.2 \cdot 1 = -0.1 \\ \Phi_i(P; b^{(0)}, s_0) &= 0.5 \cdot 1 + 0.3 \cdot 0 + 0.2 \cdot (-1) = 0.3 \\ \Phi_i(S; b^{(0)}, s_0) &= 0.5 \cdot (-1) + 0.3 \cdot 1 + 0.2 \cdot 0 = -0.2 \end{aligned}$$

The agent then chooses to play the action that has maximum value. In this case:

$$t_i^*(b^{(0)}; s_0) = \arg \max_{a_i \in \mathcal{A}_i} \Phi_i(a_i; b^{(0)}, s_0) = P$$

That is, the  $ToM_0$  agent described in Table 6a chooses to play  $P$ .

#### 4.3. First-order theory of mind agents

A  $ToM_1$  agent attributes beliefs to his opponent in the form of an additional probability distribution  $b^{(1)}$ . Here,  $b^{(1)}(a_i; s)$  represents what the agent believes his opponent to judge what the probability is that he will play action  $a_i \in \mathcal{A}_i$  in game state  $s \in S$ . However, a  $ToM_1$  agent also has zero-order beliefs  $b^{(0)}$  about what his opponent will do. The decision process of the  $ToM_1$  agent consists of roughly three steps:

1. making a prediction  $\hat{a}_j^{(1)}$  of opponent behaviour, based on the agent's first-order beliefs  $b^{(1)}$ ;
2. integrating the first-order prediction  $\hat{a}_j^{(1)}$  of opponent behaviour and the zero-order belief  $b^{(0)}$ ; and
3. selecting the action that maximizes the agent's expected payoff, given his integrated beliefs about opponent behaviour.

Let us describe each step more precisely.

(1) First, the  $ToM_1$  agent makes a prediction of opponent behaviour based on his first-order beliefs  $b^{(1)}$ . Using simulation-theory of mind, the agent uses his own decision function  $t^*$  to make a prediction of the action his opponent will play. To do so, the agent determines the action  $\hat{a}_j^{(1)} \in \mathcal{A}_j$  that maximizes the value function from the perspective of the opponent, given that the agent believes his opponent to have zero-order beliefs  $b^{(1)}$ . That is,

$$\hat{a}_j^{(1)} = t_j^*(b^{(1)}; s) = \arg \max_{a_j \in \mathcal{A}_j} \Phi_j(a_j; b^{(1)}, s). \quad (5)$$

Note that Eq. (5) is similar to Eq. (4). That is, the  $ToM_1$  agent determines his prediction of opponent behaviour similar to the way a  $ToM_0$  agent determines his own behaviour. In calculating the prediction  $\hat{a}_j^{(1)}$ , the agent makes use of his own value function and his beliefs  $b^{(1)}$ . Note that by specifying  $\hat{a}_j^{(1)}$ , the agent makes a single prediction of the opponent's behaviour rather than assigning probabilities to each possible opponent action. This allows the agent to check the validity of his prediction more easily, by comparing the prediction  $\hat{a}_j^{(1)}$  with the opponent's actual behaviour. However, this also means that slight differences between the agent's value function and that of his opponent may render the prediction incorrect.

(2) A  $ToM_1$  agent's first-order theory of mind provides the agent with a prediction  $\hat{a}_j^{(1)}$  of opponent behaviour. This prediction may conflict with his zero-order beliefs  $b^{(0)}$ . The extent to which first-order theory of mind governs the decisions of the agent's actions is determined by his confidence  $0 \leq c_1 \leq 1$  that first-order theory of mind accurately predicts his opponent's behaviour. The value of his confidence  $c_1$  allows the agent to distinguish between different types of opponents, and he weights his zero-order beliefs against the prediction of first-order theory of mind accordingly. This weighting process is captured by a belief integration function  $U$ . This function integrates the agent's first-order prediction  $\hat{a}_j$  with his zero-order beliefs  $b^{(0)}$  of opponent behaviour. Compared to his zero-order beliefs  $b^{(0)}$ , the agent's integrated belief that his opponent will be playing action  $\hat{a}_j^{(1)}$  is increased, while his integrated belief that his opponent will be playing any other action is decreased. Specifically,

$$U(b^{(0)}, \hat{a}_j^{(1)}, c_1)(a_j; s) = \begin{cases} (1 - c_1) \cdot b^{(0)}(a_j; s) & \text{if } a_j \neq \hat{a}_j^{(1)}, \\ (1 - c_1) \cdot b^{(0)}(a_j; s) + c_1 & \text{if } a_j = \hat{a}_j^{(1)}. \end{cases} \quad (6)$$

(3) After integrating his zero-order beliefs  $b^{(0)}$  and his prediction of opponent behaviour  $\hat{a}_j^{(1)}$  based on first-order theory of mind, the agent chooses what action to play. This decision is made analogously to the way a  $ToM_0$  agent decides (Eq. (4)). However, the  $ToM_1$  agent decides based on his integrated beliefs  $U(b^{(0)}, \hat{a}_j^{(1)}, c_1)$  of opponent behaviour, instead of his zero-order beliefs  $b^{(0)}$  directly. That is, a  $ToM_1$  agent chooses to play the action given by

$$t_i^*(U(b^{(0)}, \hat{a}_j^{(1)}, c_1); s) = t_i^*(U(b^{(0)}, t_j^*(b^{(1)}; s), c_1); s). \quad (7)$$

In the special case where the agent has no confidence in first-order theory of mind,  $c_1 = 0$ , the  $ToM_1$  agent's decision is only influenced by his zero-order beliefs. In this case, the agent chooses as if he were a  $ToM_0$  agent.

**Example 2.** Consider a  $ToM_1$  agent that plays rock–paper–scissors, similar to the agent in [Example 1](#), whose mental content is given in [Table 6b](#). The table shows that the  $ToM_1$  agent has zero-order beliefs  $b^{(0)}$ , which indicate the agent's beliefs concerning his opponent's actions, as well as first-order beliefs  $b^{(1)}$ . For example, since  $b^{(1)}(R, s_0) = 0.4$ , the agent believes that his opponent believes that there is a 40% probability that he is going to play  $R$ . Taking the perspective of his opponent, the agent determines what he would do in her place. That is, the agent first calculates the value that he would assign to each of the actions available to his opponent, if his first-order beliefs  $b^{(1)}$  were actually his zero-order beliefs, and his opponent's payoffs were actually his payoffs.

$$\Phi_j(R; b^{(1)}, s_0) = 0.4 \cdot 0 + 0.5 \cdot (-1) + 0.1 \cdot 1 = -0.4$$

$$\Phi_j(P; b^{(1)}, s_0) = 0.4 \cdot 1 + 0.5 \cdot 0 + 0.1 \cdot (-1) = 0.3$$

$$\Phi_j(S; b^{(1)}, s_0) = 0.4 \cdot (-1) + 0.5 \cdot 1 + 0.1 \cdot 0 = 0.1$$

The agent's first-order theory of mind predicts that his opponent will select the action that will yield her the highest payoff.

$$\hat{a}_j^{(1)} = t_j^*(b^{(1)}; s_0) = \arg \max_{a_j \in \mathcal{A}_j} \Phi_j(a_j; b^{(1)}, s_0) = P$$

Using his first-order theory of mind, the agent predicts that his opponent is going to play  $P$ .

Note that the agent's prediction  $\hat{a}_j^{(1)}$  conflicts with his zero-order beliefs  $b^{(0)}$ . According to his first-order theory of mind, his opponent is going to play  $P$ , while the agent's zero-order beliefs assign a 50% probability that his opponent is going to play  $R$ . To be able to make a decision, the agent integrates his first-order prediction with his zero-order beliefs  $b^{(0)}$ . In this case, the agent's confidence  $c_1$  in first-order theory of mind is 0.9. This means that the agent's integrated beliefs are determined for 90% by his prediction based on first-order theory of mind, and for 10% by his zero-order beliefs.

$$U(b^{(0)}, P, 0.9)(R; s_0) = (1 - 0.9) \cdot b^{(0)}(R; s_0) = 0.1 \cdot 0.5 = 0.05$$

$$U(b^{(0)}, P, 0.9)(P; s_0) = (1 - 0.9) \cdot b^{(0)}(P; s_0) + 0.9 = 0.93$$

$$U(b^{(0)}, P, 0.9)(S; s_0) = (1 - 0.9) \cdot b^{(0)}(S; s_0) = 0.1 \cdot 0.2 = 0.02$$

After integrating his zero-order beliefs and first-order prediction, the agent believes there is a 5% probability that his opponent is going to play *R*, a 93% probability that his opponent is going to play *P* and a 2% probability that his opponent is going to play *S*.

Based on his integrated beliefs, the agent determines the value for playing each of the actions.

$$\Phi_i(R; U(b^{(0)}, P, 0.9), s_0) = 0.05 \cdot 0 + 0.93 \cdot (-1) + 0.02 \cdot 1 = -0.91$$

$$\Phi_i(P; U(b^{(0)}, P, 0.9), s_0) = 0.05 \cdot 1 + 0.93 \cdot 0 + 0.02 \cdot (-1) = 0.03$$

$$\Phi_i(S; U(b^{(0)}, P, 0.9), s_0) = 0.05 \cdot (-1) + 0.93 \cdot 1 + 0.02 \cdot 0 = 0.88$$

The agent then chooses to play the action that has maximum value. In this case:

$$t_i^*(U(b^{(0)}, P, 0.9); s_0) = \arg \max_{a_i \in \mathcal{A}_i} \Phi_i(a_i; U(b^{(0)}, P, 0.9), s_0) = S$$

That is, the  $ToM_1$  agent described in Table 6b chooses to play *S*.

#### 4.4. Second-order theory of mind agents

Similar to the way a  $ToM_1$  agent models his opponent as a  $ToM_0$  agent, a  $ToM_2$  agent considers the possibility that his opponent may be a  $ToM_1$  agent. As such, the  $ToM_2$  agent has an explicit model of what beliefs he believes his opponent to be attributing to him. In our model, these beliefs are represented by an additional belief structure  $b^{(2)}$ . Using simulation-theory of mind, the agent attributes the decision-making process described by Eq. (7) to his opponent. That is, the agent considers the game from the perspective of his opponent, and determines what he would do in her position, if he were a  $ToM_1$  agent.

To determine his opponent's actions, the  $ToM_2$  agent needs to know her confidence  $c_1$  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.<sup>5</sup> Based on second-order theory of mind, the  $ToM_2$  agent therefore predicts that his opponent will be playing (square brackets are used for readability)

$$\hat{a}_j^{(2)} = t_j^*(U[b^{(1)}, t_i^*(b^{(2)}; s), 0.8]; s). \quad (8)$$

This prediction  $\hat{a}_j^{(2)}$  based on second-order theory of mind is integrated with the  $ToM_2$  agent's zero-order beliefs  $b^{(0)}$  and his prediction  $\hat{a}_j^{(1)}$  based on first-order theory of mind, before he makes his choice of what action to play. As for the  $ToM_1$  agent, a  $ToM_2$  agent does not know at which order of theory of mind his opponent is playing. Instead, the extent to which second-order theory of mind governs the decisions of the  $ToM_2$  agent's actions is determined by his confidence  $0 \leq c_2 \leq 1$  that second-order theory of mind accurately predicts his opponent's behaviour. The  $ToM_2$  agent weights the integrated beliefs in Eq. (7) against his prediction of opponent behaviour  $\hat{a}_j^{(2)}$  based on second-order theory of mind. As a result, the  $ToM_2$  agent's integrated beliefs about his opponent behaviour are given by

$$U[U(b^{(0)}, \underbrace{t_j^*(b^{(1)}; s)}_{\hat{a}_j^{(1)}}), \underbrace{t_j^*(U[b^{(1)}, t_i^*(b^{(2)}; s), 0.8]; s)}_{\hat{a}_j^{(2)}}], c_2]. \quad (9)$$

The  $ToM_2$  agent therefore performs two belief integration steps. First, the agent integrates his zero-order beliefs  $b^{(0)}$  concerning his opponent's behaviour with his prediction  $\hat{a}_j^{(1)}$  based on application of first-order theory of mind. In the second step, his prediction  $\hat{a}_j^{(2)}$  based on second-order theory of mind is integrated into these beliefs as well. The  $ToM_2$  agent then makes his final choice of what action to select based on these beliefs:

$$t_i^*(U[U(b^{(0)}, \hat{a}_j^{(1)}, c_1), \hat{a}_j^{(2)}, c_2]; s). \quad (10)$$

**Example 3.** Consider a  $ToM_2$  agent that plays rock–paper–scissors, similar to Example 2, whose mental content is given in Table 7. When a  $ToM_2$  agent considers his opponent's first-order beliefs about his own actions, the agent performs the decision process of a  $ToM_1$  agent from the viewpoint of his opponent. That is, he calculates what he believes that she predicts that he will do based on her first-order beliefs. The agent's model of his opponent's first-order beliefs are captured

<sup>5</sup> Results from additional simulations using different values of  $c_1 \in [0, 1]$  turned out to be visually indistinguishable from the ones presented here for any value of  $c_1$  over 0.5.

**Table 7**  
Possible mental content of a  $ToM_2$  agent in a game of rock–paper–scissors, as in Example 3.

	Order of theory of mind $k$		
	0	1	2
$c_k$		0.9	0.1
$b^{(k)}(R, s_0)$	0.5	0.4	0.3
$b^{(k)}(P, s_0)$	0.3	0.5	0.3
$b^{(k)}(S, s_0)$	0.2	0.1	0.4

by  $b^{(2)}$ . This is what the agent believes his opponent to believe his first-order beliefs to be. Firstly, the agent determines what he would do if his second-order beliefs  $b^{(2)}$  were actually his zero-order beliefs.

$$\Phi_i(R; b^{(2)}, s_0) = 0.3 \cdot 0 + 0.3 \cdot (-1) + 0.4 \cdot 1 = 0.1$$

$$\Phi_i(P; b^{(2)}, s_0) = 0.3 \cdot 1 + 0.3 \cdot 0 + 0.4 \cdot (-1) = -0.1$$

$$\Phi_i(S; b^{(2)}, s_0) = 0.3 \cdot (-1) + 0.3 \cdot 1 + 0.4 \cdot 0 = 0$$

$$t_i^*(b^{(2)}; s_0) = \arg \max_{a_i \in \mathcal{A}_i} \Phi_i(a_i; b^{(2)}, s_0) = R$$

That is, the  $ToM_2$  agent believes his opponent to predict that he will be playing  $R$ .

Secondly, the agent determines how his opponent's prediction that he will be playing  $R$  influences her zero-order beliefs. The agent does not explicitly model the opponent's confidence in first-order theory of mind. Rather, he assumes a value of 0.8 for this confidence. The agent then integrates his first-order beliefs  $b^{(1)}$ , which he believes to correspond to his opponent's zero-order beliefs, with the prediction that he will play  $R$ .

$$U(b^{(1)}, R, 0.8)(R; s_0) = 0.2 \cdot b^{(1)}(R; s_0) + 0.8 = 0.88$$

$$U(b^{(1)}, R, 0.8)(P; s_0) = 0.2 \cdot b^{(1)}(R; s_0) = 0.2 \cdot 0.5 = 0.10$$

$$U(b^{(1)}, R, 0.8)(S; s_0) = 0.2 \cdot b^{(1)}(R; s_0) = 0.2 \cdot 0.1 = 0.02$$

These integrated beliefs specify what the agent believes what his opponent's beliefs are concerning his actions. For example, based on application of his second-order theory of mind, the  $ToM_2$  agent believes that his opponent believes that there is an 88% probability that he himself will play  $R$ . From the viewpoint of his opponent, the agent then determines what the value would be for playing each of the possible actions, given the integrated beliefs of opponent action.

$$\Phi_i(R; U(b^{(1)}, R, 0.8), s_0) = 0.88 \cdot 0 + 0.10 \cdot (-1) + 0.02 \cdot 1 = -0.08$$

$$\Phi_i(P; U(b^{(1)}, R, 0.8), s_0) = 0.88 \cdot 1 + 0.10 \cdot 0 + 0.02 \cdot (-1) = 0.86$$

$$\Phi_i(S; U(b^{(1)}, R, 0.8), s_0) = 0.88 \cdot (-1) + 0.10 \cdot 1 + 0.02 \cdot 0 = -0.78$$

The action that maximizes this value represents the agent's prediction of the action his opponent is going to play according to his second-order theory of mind.

$$\hat{a}_j^{(2)} = t_j^*(U(b^{(1)}, R, 0.8); s_0) = P.$$

Based on second-order theory of mind, the agent therefore believes his opponent will play  $P$ .

To make a decision, the agent integrates his zero-order beliefs  $b^{(0)}$ , his first-order prediction  $\hat{a}_j^{(1)} = P$  (see Example 2), and his second-order prediction  $\hat{a}_j^{(2)} = P$ . Example 2 shows how the agent's zero-order beliefs and his first-order prediction of opponent behaviour are integrated. Using this confidence  $c_2$ , the agent also integrates his belief that his opponent is going to play  $P$ . In this example, the agent has confidence  $c_2 = 0.1$  in second-order theory of mind. This results in the following integrated beliefs:

$$U(U(b^{(0)}, P, 0.9), P, 0.1)(R; s_0) = 0.9 \cdot 0.05 = 0.045$$

$$U(U(b^{(0)}, P, 0.9), P, 0.1)(P; s_0) = 0.9 \cdot 0.93 + 0.1 = 0.937$$

$$U(U(b^{(0)}, P, 0.9), P, 0.1)(S; s_0) = 0.9 \cdot 0.02 = 0.018$$

Based on these integrated beliefs, the agent determines the value for playing each of the actions.

$$\Phi_i(R; U(U(b^{(0)}, P, 0.9), P, 0.1), s_0) = 0.018 - 0.937 = -0.919$$

$$\Phi_i(P; U(U(b^{(0)}, P, 0.9), P, 0.1), s_0) = 0.045 - 0.018 = 0.027$$

$$\Phi_i(S; U(U(b^{(0)}, P, 0.9), P, 0.1), s_0) = 0.937 - 0.045 = 0.892$$

The agent then chooses to play the action that maximizes the value. In this case:

$$t_i^*(U(U(b^{(0)}, P, 0.9), P, 0.1); s_0) = \arg \max_{a_i \in \mathcal{A}_i} \Phi_i(a_i; U(U(b^{(0)}, P, 0.9), P, 0.1), s_0) = S$$

Based on his integrated beliefs of what the opponent is going to do, the  $ToM_2$  agent's choice is to play  $S$ .

#### 4.5. Higher orders of theory of mind agents

For every order of theory of mind available to the agent beyond the second-order, say order  $k$ , the agent maintains an additional belief structure  $b^{(k)}$ . These beliefs are used to expand his decision process by modeling the decision process of a  $(k - 1)$ st-order theory of mind agent from his opponent's point of view. The resulting prediction is weighted against the decision process of  $(k - 1)$ st-order of theory of mind from his own point of view. For example, a  $ToM_3$  agent expands the decision process of a  $ToM_2$  agent, represented by Eq. (10). He does so by modeling the decision process of a  $ToM_2$  agent from his opponent's point of view. That is, the  $ToM_3$  agent calculates his prediction of opponent behaviour  $\hat{a}_j^{(3)}$  based on third-order theory of mind:

$$\hat{a}_j^{(3)} = t_j^*(U[U(b^{(1)}, t_i^*(b^{(2)}; s), 0.8), t_i^*(U[b^{(2)}, t_j^*(b^{(3)}; s), 0.8]; s), 0.8]; s). \quad (11)$$

Once the  $ToM_3$  agent has determined his prediction based on third-order theory of mind, he weights this prediction against the decision process of a  $ToM_2$  agent, represented by Eq. (10). The extent to which the  $ToM_3$  agent's prediction  $\hat{a}_j^{(3)}$  of a  $ToM_2$  opponent's behaviour is reflected in his own behaviour is determined by his confidence  $0 \leq c_3 \leq 1$  that third-order theory of mind yields accurate predictions of his opponent's behaviour. That is, the choice of a  $ToM_3$  agent is given by

$$t_j^*(U[U(U(b^{(0)}, \underbrace{t_i^*(b^{(1)})}_{\hat{s}^{(1)}}), \underbrace{t_i^*(U(b^{(1)}, t_i^*(b^{(2)}), 0.8))}_{\hat{s}^{(2)}}], c_2), \\ \underbrace{t_j^*(U[U(b^{(1)}, t_i^*(b^{(2)}; s), 0.8), t_i^*(U[b^{(2)}, t_j^*(b^{(3)}; s), 0.8]; s), 0.8]; s)}_{\hat{s}^{(3)}}], c_3)].$$

#### 4.6. Belief adjustment and learning speed

In the previous subsections, we discussed how agents of different orders of theory of mind decide what action to play, based on their current beliefs  $b^{(k)}$  and confidence levels  $c_k$ . By placing himself in the position of his opponent, and viewing the game from her perspective, an agent makes predictions for the action his opponents is going to perform. Each order of theory of mind available to the agent generates such a prediction. The agent can use the accuracy of these predictions to gain information about the opponent's abilities over repeated games, and adjust his beliefs and confidence levels accordingly.

For example, a  $ToM_2$  agent may learn that his opponent is not playing as predicted by his second-order theory of mind, but that his first-order theory of mind consistently makes accurate predictions of her actions. In such a case, the  $ToM_2$  agent may start to play as if he were a  $ToM_1$  opponent, and ignore predictions from his second-order theory of mind altogether. However, it is important to note that while the  $ToM_2$  agent may adjust his behaviour to take advantage of predictable behaviour of his opponent, his opponent is trying to do the same. In this section, we describe how agents update their beliefs  $b^{(k)}$  and confidence levels  $c_k$  when they observe the outcome of a game.

When an agent plays against an unfamiliar opponent for the first time, his beliefs  $b^{(k)}$  are initialized randomly, while his confidence levels  $c_k$  are initialized at zero. After each round, the actual choice  $\tilde{a}_i$  of the agent and  $\tilde{a}_j$  of his opponent are revealed. At this moment, an agent updates his confidence in theory of mind based on the accuracy of his predictions. A  $ToM_1$  agent increases his confidence  $c_1$  in first-order theory of mind when his first-order prediction  $\hat{a}_j^{(1)}$  calculated through Eq. (5) was correct. In other cases, his confidence in first-order theory of mind decreases. This process is represented by the update

$$c_1 := \begin{cases} (1 - \lambda) \cdot c_1 & \text{if } \tilde{a}_j \neq \hat{a}_j^{(1)}, \\ \lambda + (1 - \lambda) \cdot c_1 & \text{if } \tilde{a}_j = \hat{a}_j^{(1)}, \end{cases} \quad (12)$$

where  $0 \leq \lambda \leq 1$  is an agent-specific *learning speed*. An agent's learning speed indicates the relative weight of new information in determining beliefs. An agent with a high learning speed determines whether his opponent is a  $ToM_0$  agent based on his most recent observations. The  $ToM_1$  agent's confidence  $c_1$  in first-order theory of mind reflects the accuracy of first-order theory of mind in the most recent games in this case. An agent with a low learning speed depends on experience built up over a longer period of time.

For higher orders of theory of mind, an agent additionally adjusts each of his confidences  $c_k$  in  $k$ th-order theory of mind for each order  $k$  of theory of mind available to him. Similar to the update of his confidence  $c_1$  in first-order theory of mind, an agent reduces his confidence  $c_k$  in  $k$ th-order theory of mind when the corresponding prediction  $\hat{a}_j^{(k)}$  of opponent

behaviour based on application of  $k$ th-order theory of mind was incorrect, that is  $\hat{a}_j^{(k)} \neq \tilde{a}_j$ . However, an agent only increases his confidence in  $k$ th-order theory of mind when it yields correct predictions, and the predictions made by each order of theory of mind lower than  $k$  were incorrect. If there is some lower order  $n < k$  of theory of mind for which  $\hat{a}_j^{(n)} = \tilde{a}_j$ , the agent does not increase his confidence in  $k$ th-order theory of mind. That is, theory of mind agents only grow more confident in the use of higher-order theory of mind when this results in accurate predictions that could not have been made with a lower order of theory of mind. This feature makes agents less likely to overestimate the theory of mind abilities of their opponent.

$$c_k := \begin{cases} (1 - \lambda) \cdot c_k & \text{if } \tilde{a}_j \neq \hat{a}_j^{(k)}, \\ c_k & \text{if there is a } 1 \leq n < k \text{ such that } \tilde{a}_j = \hat{a}_j^{(n)} = \hat{a}_j^{(k)}, \\ \lambda + (1 - \lambda) \cdot c_k & \text{otherwise.} \end{cases} \quad (13)$$

When the actual choice of the agent  $\tilde{a}_i$  and his opponent  $\tilde{a}_j$  are revealed, the agent also updates his beliefs  $b^{(k)}$ . Since the zero-order beliefs  $b^{(0)}$  represent the agent's beliefs concerning his opponent's behaviour, these beliefs are updated using his opponent's choice  $\tilde{a}_j$ . This is done by increasing the belief the opponent will perform action  $\tilde{a}_j$  in the same game state  $s \in \mathcal{S}$ , while decreasing the belief that she will perform any other action. Second-order beliefs  $b^{(2)}$  specify what the agent believes his opponent to believe about what he believes that she is going to do. That is, an agent's second-order beliefs  $b^{(2)}$  describe beliefs concerning the actions of his opponent, and are therefore updated using her choice  $\tilde{a}_j$  as well. After this update, the agent believes that his opponent believes that he believes more strongly that she will perform the action  $\tilde{a}_j$  in the same game state  $s \in \mathcal{S}$ . This is true for each of the even-numbered orders of theory of mind available to the agent. The belief structure  $b^{(k)}$  of all even-numbered orders of theory of mind are updated using the opponents choice  $\tilde{a}_j$ .

On the other hand, the odd-numbered orders of theory of mind describe the actions of the agent himself. These beliefs are therefore updated using the agent's choice  $\tilde{a}_i$ . For example, after the belief adjustment, the agent believes that his opponent believes more strongly that he will perform action  $\tilde{a}_i$  when the same game state  $s \in \mathcal{S}$  is encountered again. Using the belief updating function  $U$ , the beliefs are adjusted using the agent's learning speed  $\lambda$ , such that

$$b^{(k)}(a_j; s) := U(b^{(i)}, \tilde{a}_j, \lambda)(a_j; s) \quad \text{for } k \text{ even and all } a_j \in \mathcal{A}_j, \quad \text{and} \quad (14)$$

$$b^{(k)}(a_i; s) := U(b^{(i)}, \tilde{a}_i, \lambda)(a_i; s) \quad \text{for } k \text{ odd and all } a_i \in \mathcal{A}_i. \quad (15)$$

That is, the agent adjusts his beliefs based on the forecasting technique of exponential smoothing [69]. Note that these adjustments only apply to the game state  $s$  in which the actions were taken.

The agent's learning speed  $\lambda$  determines how quickly the agent learns. That is, a higher value of  $\lambda$  shows that the agent changes his beliefs more radically based on new information. At the maximum of  $\lambda = 1$ , an agent effectively believes that the last action his opponent performed determines future behaviour. At the other extreme of  $\lambda = 0$ , the agent does not learn, and does not change his beliefs when new information becomes available. This also means that an agent with learning speed  $\lambda = 0$  does not change his behaviour.

The agents we describe do not actively try to model the learning speed  $\lambda$  of their opponent. Instead, an agent assumes that his opponent updates her beliefs using the same learning speed as he does himself. That is, our computational agents do not consider the possibility that their opponent reacts differently to new information than they do themselves. This means that in general, the beliefs that an agent attributes to his opponent are structurally different from her actual beliefs.

An agent makes use of theory of mind by considering the position of his opponent from his own viewpoint. When an agent makes use of second-order theory of mind, he also considers what his opponent knows about his own viewpoint. Since the games we consider have symmetric information, this causes the agent's second-order beliefs  $b^{(2)}$  to resemble his zero-order beliefs  $b^{(0)}$  more closely with each update. That is, the agent eventually believes that a first-order theory of mind opponent knows what his zero-order beliefs are.

Due to the restrictions on the learning speed  $\lambda$ , Eqs. (14) and (15) preserve the normalization and non-negativity of beliefs. Similarly, the confidences  $c_i$  in the application of  $i$ th-order theory of mind remain limited to the range  $[0, 1]$ .

**Example 4.** Consider the  $ToM_2$  agent from Example 3, whose mental content is given in Table 7. Once both the agent and his opponent have decided on an action to play, the actions are revealed to both players, and each receives the payoff based on those actions. Once the outcome of the game is revealed, each agent updates his beliefs based on what is observed. Our calculations showed that the  $ToM_2$  agent we discussed in our example has played action  $\tilde{a}_i = S$ . We assume that his opponent played  $\tilde{a}_j = P$ , and that the agent's learning speed  $\lambda = 0.6$ .

Table 8 lists the agent's predictions, confidences in theory of mind and beliefs before and after the belief update. Depending on the accuracy of the prediction of application of  $i$ th-order theory of mind, the confidence  $c_i$  in that order of theory of mind increases or decreases. In our example, first-order theory of mind accurately predicted that the opponent would play  $P$ , since  $\tilde{a}_j = \hat{a}_j^{(1)}$ . As a result, the new confidence  $c_1$ , as calculated by Eq. (12), becomes

$$c_1 := (1 - \lambda) \cdot c_1 + \lambda = (1 - 0.6) \cdot 0.9 + 0.6 = 0.96.$$

**Table 8**Beliefs and confidences in theory of mind before and after the belief update in the example of a  $ToM_3$  agent playing RPS.

	Order of theory of mind $i$					
	Before update			After update		
	0	1	2	0	1	2
$\hat{a}_j^{(i)}$		$P$	$P$			
$c_i$		0.9	0.1		0.96	0.10
$b^{(i)}(R, s_0)$	0.5	0.4	0.3	0.20	0.20	0.12
$b^{(i)}(P, s_0)$	0.3	0.5	0.3	0.72	0.16	0.72
$b^{(i)}(S, s_0)$	0.2	0.1	0.4	0.08	0.64	0.16

Table 8 shows that  $\hat{a}_j^{(2)} = P = \tilde{a}_j$ , and thus that second-order theory of mind also correctly predicted that the opponent would play  $P$ . However, since first-order theory of mind is of a lower order than second-order theory of mind, and since first-order theory of mind also correctly predicted the action of the opponent, the confidence  $c_2$  in second-order theory of mind remains unchanged.

The actions that were actually played by the agent and his opponent also change the agent's beliefs. Each even-numbered order of theory of mind refers to beliefs concerning the opponent's actions. These beliefs are therefore updated to reflect the action that the opponent has taken most recently. This is done by increasing the belief that the opponent will perform the same action, in our case  $P$ , while decreasing the other beliefs. That is, the agent's zero-order beliefs  $b(0)$  are updated, such that after the update

$$b^{(0)}(R; s_0) := U(b^{(0)}, P, 0.6)(R; s_0) = (1 - 0.6) \cdot 0.5 = 0.2$$

$$b^{(0)}(P; s_0) := U(b^{(0)}, P, 0.6)(P; s_0) = (1 - 0.6) \cdot 0.3 + 0.6 = 0.72$$

$$b^{(0)}(S; s_0) := U(b^{(0)}, P, 0.6)(S; s_0) = (1 - 0.6) \cdot 0.2 = 0.08$$

This means that after the belief update, the agent believes that there is a 72% probability that his opponent is going to repeat the action  $P$  in the next round.

The agent's second-order beliefs  $b^{(2)}$  also concern the actions of the opponent. Specifically, the agent's second-order beliefs  $b^{(2)}$  determine what the agent believes his opponent to believe what he believes about her actions. The agent uses the action  $\tilde{a}_j = P$  actually performed by his opponent to update his second-order beliefs as well.

The odd-numbered orders of theory of mind represent beliefs concerning the agent's own actions. These beliefs are therefore updated to reflect that the agent chose action  $\tilde{a}_i = S$ . For the agent's first-order beliefs  $b^{(1)}$ , this results in

$$b^{(1)}(R; s_0) := U(b^{(1)}, S, 0.6)(R; s_0) = (1 - 0.6) \cdot 0.5 = 0.20$$

$$b^{(1)}(P; s_0) := U(b^{(1)}, S, 0.6)(P; s_0) = (1 - 0.6) \cdot 0.4 = 0.16$$

$$b^{(1)}(S; s_0) := U(b^{(1)}, S, 0.6)(S; s_0) = (1 - 0.6) \cdot 0.1 + 0.6 = 0.64$$

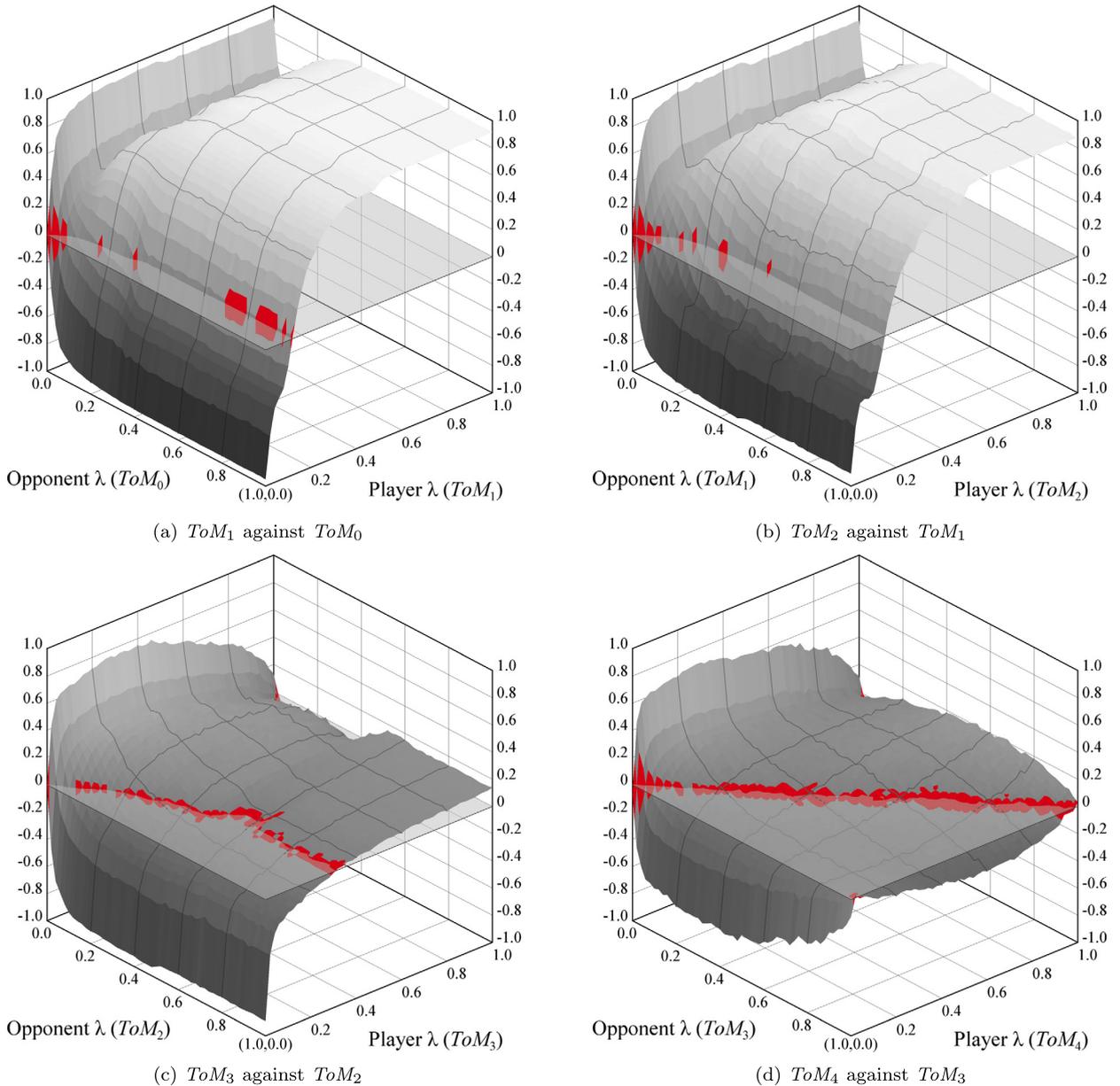
This means that after the belief update, the agent believes that his opponent believes that there is a 64% probability that he will repeat the action  $S$  in the next round.

## 5. Results

The agent model described in Section 4 has been implemented in Java and its performance has been tested in each of the settings described in Section 2. For the rock–paper–scissors game, as well as the variations on this game, each trial consisted of an agent that plays 20 consecutive games against the same opponent.<sup>6</sup> An agent's trial score is the average of the agent's game scores over all games in the trial. The graphs in this section depict the average trial score, averaged over 500 trials. Since Limited Bidding is a more complex game, a longer sequence is needed to learn to model the opponent. Each trial in this game consisted of an agent that plays 50 consecutive games against the same opponent. Our results were qualitatively similar if longer trials of 100 games were used instead.

In this section, performance is measured as the average trial score of the focal agent, as a function of his learning speed  $\lambda_i$ , as well as the learning speed  $\lambda_j$  of his opponent. The figures in this section show simulation results for every 0.02 step in learning speeds over the range  $\lambda_i, \lambda_j \in [0, 1]$ . We report the results of simulations in which a focal agent is exactly one order of theory of mind higher than his opponent. In simulations in which the difference in theory of mind ability of the focal agent and his opponent was larger than one order, performance of the focal agent turned out to be similar.

<sup>6</sup> We have compared the results for trials of 20 games to longer trials of 50 and 100 games and found no qualitative differences.



**Fig. 6.** Average performance of theory of mind agents playing rock–paper–scissors against opponents of a lower order of theory of mind. Performance was averaged over 500 trials of 20 consecutive games each. Insignificant results ( $p > 0.01$ ) are highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 5.1. Rock–paper–scissors

Fig. 6 shows how the ability to represent mental content of others affects the performance of agents in the RPS game as a function of the learning speed  $\lambda_i$  of the focal agent and of the learning speed  $\lambda_j$  of his opponent. Higher and lighter areas indicate that the focal agent won more games than he lost, while lower and darker areas show that his opponent had the upper hand. To emphasize the shape of the surface, the grid that appears on the bottom plane has been projected onto the surface, and the plane of zero performance appears as a semi-transparent surface in the figure. Red areas indicate where performance was not significantly different from zero, at a significance level  $\alpha = 0.01$ .

Fig. 6a shows that a  $ToM_1$  agent that has a learning speed  $\lambda_i = 0$  cannot compete with his opponent. When the agent does not learn from his opponent’s behaviour, he loses nearly all rounds. Similarly, his opponent loses nearly all rounds when she does not learn at all ( $\lambda_j = 0$ ). This shows that both zero-order theory of mind agents and first-order theory of mind agents can successfully model an opponent that always performs the same action.

The figure also shows that a  $ToM_1$  agent mostly outperforms a  $ToM_0$  opponent. Whenever the  $ToM_1$  agent's learning speed is at least  $\lambda_i > 0.1$ , he will on average win more rounds than he loses, and obtain a positive score. The  $ToM_1$  agent's score is particularly high when both he and his opponent learn at a high rate, in which case the agent wins almost all rounds. When his  $ToM_0$  opponent learns at a low rate, the average score of the  $ToM_1$  agent is reduced.

The relatively low performance of the  $ToM_1$  agent against slow learning opponents is due to the fact that learning speed determines an agent's memory. A  $ToM_0$  agent with high learning speed adapts to new situations quickly, but also quickly forgets information from previous rounds. When faced with an unpredictable opponent, a  $ToM_0$  agent with high learning speed will therefore choose erratically, but with confidence. In this case, the  $ToM_0$  agent believes that his opponent will repeat the same action she has performed the last time they met. For a  $ToM_1$  opponent, this represents a predictable situation that she can use to her advantage.

Conversely, a  $ToM_0$  agent with low learning speed retains his former beliefs for a longer time. When encountering an unpredictable opponent, such a  $ToM_0$  agent will therefore start playing with little confidence. That is, the probability distribution modeled by  $b^{(0)}$  may gradually come to resemble a uniform distribution. This causes a  $ToM_0$  agent with low learning speed to play the action that he weakly believes to be a slightly better choice than the rest. This also makes it more difficult for a  $ToM_1$  opponent to predict which token the  $ToM_0$  agent with low learning speed will play, since his choice is not robust against small deviations in his beliefs.

Although a  $ToM_1$  agent performs better against an opponent that learns quickly than against an opponent that learns slowly, performance of the  $ToM_1$  agent is largely independent of the quality of his model. When a  $ToM_1$  agent makes use of his theory of mind, he assumes that his opponent reacts to new information the same way he does. That is, the agent assumes that he and his opponent share the same learning speed. However, the figure does not show an increase in performance along the line of equal learning speeds. That is, the cost of assuming equal learning speeds is low in RPS.

Fig. 6b shows the performance of a  $ToM_2$  agent playing RPS against a  $ToM_1$  opponent. Note that Fig. 6b is similar to Fig. 6a. As for the  $ToM_1$  agent, a  $ToM_2$  agent performs best when playing RPS against a  $ToM_1$  opponent when both he and his opponent learn at a high speed, while the  $ToM_2$  agent has more difficulty modeling a  $ToM_1$  agent that learns slowly. This shows that application of higher-order theory of mind can benefit an agent when playing RPS. Performance of the  $ToM_2$  agent playing RPS against a  $ToM_1$  opponent is nonetheless slightly lower than that of a  $ToM_1$  agent playing RPS against a  $ToM_0$  opponent.

Figs. 6a and 6b suggest that application of higher orders of theory of mind benefits an agent. However, performance of a  $ToM_3$  agent playing RPS against a  $ToM_2$  agent, as shown in Fig. 6c, is poor in comparison. Although the  $ToM_3$  agent still outperforms a  $ToM_2$  opponent, he does so at a lower margin. The average score of the  $ToM_3$  agent only exceeds 0.5 when his opponent has learning speed zero. When facing a  $ToM_2$  opponent that has a low learning speed, the average score of a  $ToM_3$  agent that learns quickly even becomes negative.

Although the  $ToM_3$  agent can still outperform a  $ToM_2$  opponent at a small margin, Fig. 6d shows that a  $ToM_4$  agent no longer outperforms a  $ToM_3$  opponent in RPS. In this scenario, the outcome of the game is mostly dependent on which of the agents has the highest learning speed, and no longer on theory of mind abilities. In Fig. 6d, this can be seen by the fact that the  $ToM_4$  agent obtains a positive outcome on average only if his learning speed  $\lambda_i$  is higher than the learning speed  $\lambda_j$  of his opponent.

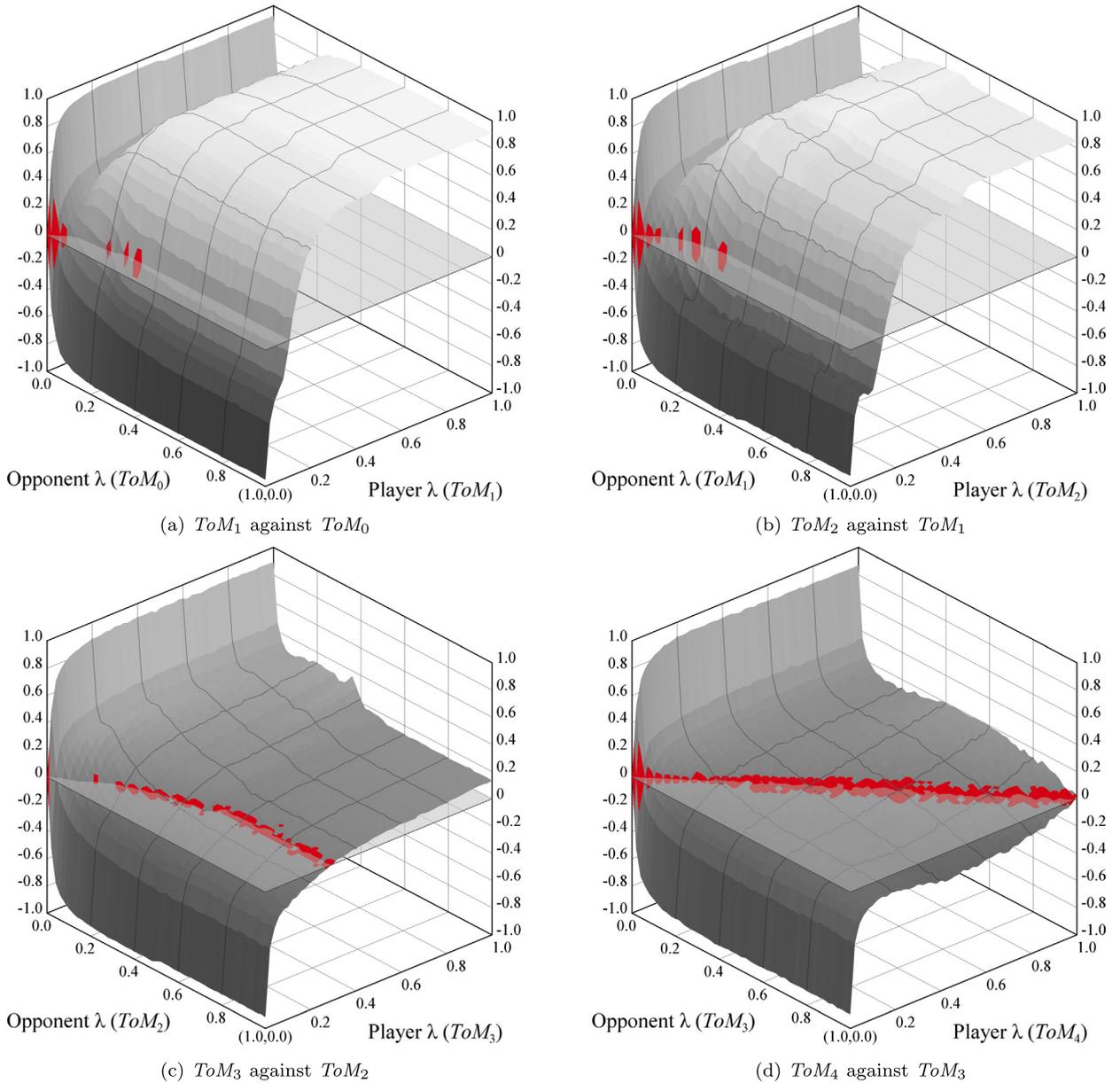
In summary, the ability to make use of theory of mind can benefit an agent in the game of RPS. As we hypothesized (cf. hypothesis  $H_{RPS}$ , Section 2.4), both the  $ToM_1$  agent and the  $ToM_2$  agent outperform opponents of a lower order of theory of mind. The performance of the  $ToM_3$  agent and the  $ToM_4$  agent suggests that there may be a limit to the effectiveness of application of higher orders of theory of mind. However, since rock–paper–scissors involves three possible opponent actions, the game leaves room for only three unique predictions of the opponent's next action. The low performance of the  $ToM_3$  and  $ToM_4$  agents may therefore be caused by specific characteristics of the RPS game, rather than a limit to the effectiveness of application of higher orders of theory of mind. The next section describes a game with more than three actions in order to differentiate between these alternative explanations.

## 5.2. Elemental rock–paper–scissors

In Section 2.1.2, we introduced elemental rock–paper–scissors, a variation on the classical RPS game in which agents choose from an action set of five actions. ERPS preserves the feature of RPS that each action is defeated by exactly one other action. Differences in performance of theory of mind agents that play RPS and those that play ERPS allow us to determine whether features of the game structure affect the effectiveness of higher orders of theory of mind in competitive games.

The results for ERPS are shown in Fig. 7. Our expectation that performance of theory of mind agents in playing ERPS would be at least as good as performance in RPS is only partially correct. Similar to our results of theory of mind agents playing RPS, Fig. 7a shows that a  $ToM_1$  agent outperforms a  $ToM_0$  opponent, while Fig. 7b shows that a  $ToM_2$  agent outperforms a  $ToM_1$  opponent as well. However, performance in the game of ERPS is slightly reduced compared to the situation in which they were playing RPS. Especially when either the agent or his opponent learns at a low speed, it is more difficult for a theory of mind agent to model his opponent in a game of ERPS than it is in RPS.

The main qualitative difference between RPS and ERPS is shown by the performance of the  $ToM_3$  agent and performance of the  $ToM_4$  agent, depicted in Fig. 7c. Our results in RPS showed that it is difficult for a  $ToM_3$  agent to model his oppo-



**Fig. 7.** Average performance of theory of mind agents playing ERPS against opponents of a lower order of theory of mind. Performance was averaged over 500 trials of 20 consecutive games each. Insignificant results ( $p > 0.01$ ) are highlighted in red.

nent correctly. In Fig. 6c, this presents itself as relatively low performance against an opponent with zero learning speed  $\lambda = 0$ . In contrast, Fig. 7c shows that the  $ToM_3$  agent does not have this difficulty when playing ERPS against a similar opponent.

Since the richer action space of ERPS increases performance of the  $ToM_3$  agent when playing against an opponent that does not learn, the structure of the game influences the effectiveness of theory of mind. However, performance of a  $ToM_3$  agent playing ERPS against a  $ToM_2$  opponent is still poor in comparison to performance of the  $ToM_1$  and  $ToM_2$  agents playing ERPS against opponents of a lower order of theory of mind. Although the  $ToM_1$  and  $ToM_2$  agents clearly outperform opponents of a lower order of theory of mind, the  $ToM_3$  agent outperforms the  $ToM_2$  agent at a small margin only.

Fig. 7d shows the performance of a  $ToM_4$  agent playing ERPS against a  $ToM_3$  opponent. Like the  $ToM_3$  agent, the peak performance of the  $ToM_4$  agent when playing against an opponent with learning speed  $\lambda_j = 0$  shown in the figure indicates that the  $ToM_4$  agent has no difficulty distinguishing agents that have learning speed zero from agents of a lower order of theory of mind. However, the ability to make use of fourth-order theory of mind does not present an agent with advantages in ERPS beyond those of third-order theory of mind. Fig. 7d shows that a  $ToM_4$  agent that plays ERPS against a  $ToM_3$

opponent only obtains a positive score on average if his learning speed  $\lambda_i$  is higher than the learning speed  $\lambda_j$  of his opponent. That is, when a  $ToM_4$  agent plays ERPS against a  $ToM_3$  opponent, whoever has the highest learning speed is expected to win.

In summary, we investigate the game of ERPS to determine whether the limited choice of actions for agents playing RPS had an effect on the advantage of making use of theory of mind. The results confirm our expectations (cf. hypothesis  $H_{ERPS}$ , Section 2.4) that when agents choose from a limited action space, higher orders of theory of mind may experience difficulty modeling their opponent. However, the limited action space does not explain the relatively poor performance of a  $ToM_3$  agent when playing against a  $ToM_2$  opponent, which was found both in RPS and ERPS.

### 5.3. Rock–paper–scissors–lizard–Spock

The game of rock–paper–scissors–lizard–Spock, described in Section 2.1.3, is a variation on ERPS in which each action is defeated by exactly two other actions. As a result, the best response to each action is not unique. Our expectation was that it would be harder to predict an opponent's behaviour in this case, and that performance of theory of mind agents would be reduced. Fig. 8 shows that this is indeed generally the case. In the game of RPSLS, the advantage of making use of theory of mind is reduced compared to RPS and ERPS.

Fig. 8a shows the performance of a  $ToM_1$  agent when playing RPSLS against a  $ToM_0$  opponent. Unlike in RPS and ERPS, a  $ToM_1$  agent performs better when his learning speed matches the learning speed of his opponent. In Fig. 8, this is reflected by high scores along the line of equal learning speeds  $\lambda_i = \lambda_j$ . In this case, the  $ToM_1$  agent's model of his opponent's beliefs matches her actual beliefs. However, even though modeling his opponent's beliefs correctly yields the agent a higher score, he is still expected to win in most cases when his learning speed does not match that of his opponent.

Performance of the  $ToM_1$  agent is particularly low when his opponent has the maximum learning speed  $\lambda_j = 1$ . In this case, she only considers the agent's actions in the previous game, and ignores all information from previous games. For example, if the  $ToM_1$  agent plays 'paper' in a game of RPSLS, the  $ToM_0$  opponent will believe that he will repeat the same action in future games. This means that the  $ToM_0$  opponent has two actions, 'lizard' or 'scissors', which maximize her expected payoff, and chooses either one of these two actions with 50% probability.

On the other hand, when the  $ToM_0$  opponent learns at a lower speed,  $\lambda_j < 1$ , she does not completely replace her beliefs when new information becomes available. In this case, the  $ToM_0$  opponent believes that there is a small probability that the  $ToM_1$  agent will play some action other than 'paper'. In general, this prevents two actions from having exactly the same expected payoffs. Since agents choose the action that yields them the highest expected payoff, this causes the  $ToM_0$  opponent to choose one of the possible actions with certainty.

As Fig. 8a shows, these two distinct types of behaviour make it more difficult for the  $ToM_1$  agent to accurately model his opponent. In the present model, a  $ToM_1$  agent that has a learning speed  $\lambda_i < 1$  believes that his opponent has the same learning speed. As a result, he believes that there is a single action that maximizes the opponent's expected payoff. However, when his opponent has the maximal learning speed  $\lambda_j = 1$ , she actually randomizes her choice over two possible actions. The  $ToM_1$  agent is therefore expected to predict his opponent's behaviour incorrectly in half the cases.

Fig. 8b shows the performance of a  $ToM_2$  agent when playing RPSLS against a  $ToM_1$  opponent. Similar to the  $ToM_1$  agent, performance of the  $ToM_2$  agent is low when playing RPSLS against an opponent that learns at maximum speed,  $\lambda_o = 1$ . The  $ToM_2$  agent also has particular difficulties modeling a  $ToM_1$  opponent in RPSLS when his own learning speed  $\lambda_i$  is low. In this case, the  $ToM_2$  agent is outperformed by an opponent of lower order of theory of mind. However, the  $ToM_2$  agent will on average win when his learning speed  $\lambda_i$  is over 0.7.

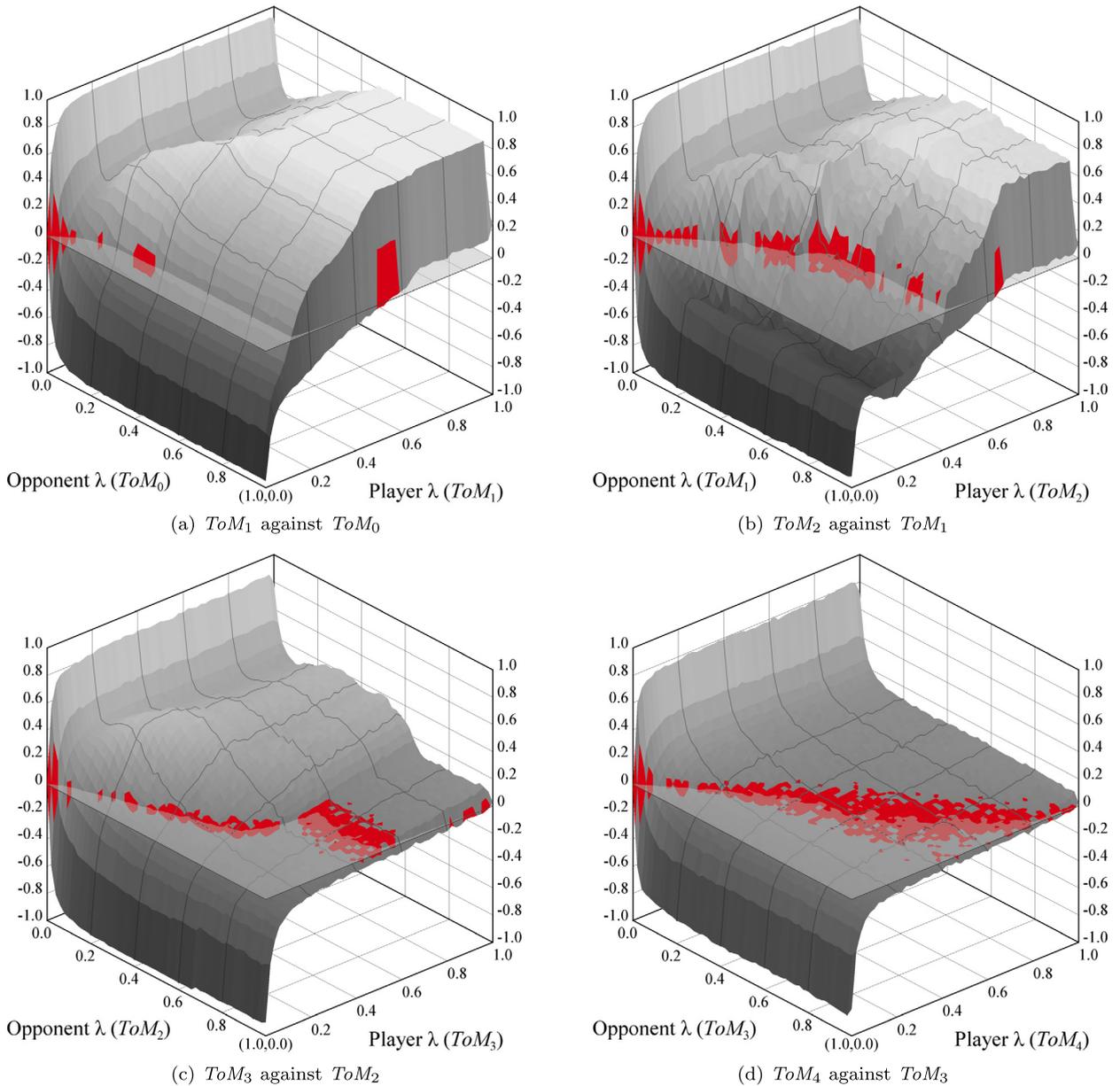
The low performance of the  $ToM_2$  agent in RPSLS when he learns at a low speed translates to a benefit for the  $ToM_3$  agent. Fig. 8c shows the performance of the  $ToM_3$  agent when playing RPSLS against a  $ToM_2$  opponent. When his opponent's learning speed  $\lambda_o$  is low, the  $ToM_3$  agent performs better in RPSLS than he would have in the games of RPS and ERPS. However, the  $ToM_3$  agent performs poorly when his  $ToM_2$  opponent learns quickly enough. In particular, when facing a  $ToM_2$  opponent that learns at the maximal learning speed  $\lambda_j = 1$ , the  $ToM_3$  agent only obtains a positive score on average when he learns at the maximal learning speed  $\lambda_i = 1$  as well.

Similar to the games of RPS and ERPS, performance of a  $ToM_4$  agent playing RPSLS against a  $ToM_3$  opponent is mostly determined by which player has the highest learning speed, as shown in Fig. 8d. However, unlike in RPS and ERPS, the  $ToM_4$  agent is at a very small advantage over his  $ToM_3$  opponent. That is, when the learning speed  $\lambda_i$  of the  $ToM_4$  agent and the learning speed  $\lambda_j$  of his  $ToM_3$  opponent are close together, the  $ToM_4$  agent is expected to win more than predicted by chance performance.

In summary, our results from the game of RPSLS show that the effectiveness of theory of mind is strongly related to the predictability of lower-order agents. Theory of mind agents perform more poorly when their opponent is indifferent between two possible actions and her behaviour is less predictable. This confirms our expectations about the relationship between the performance of theory of mind agents and the predictability of their opponents (cf. hypothesis  $H_{RPSLS}$ , Section 2.4).

### 5.4. Limited Bidding

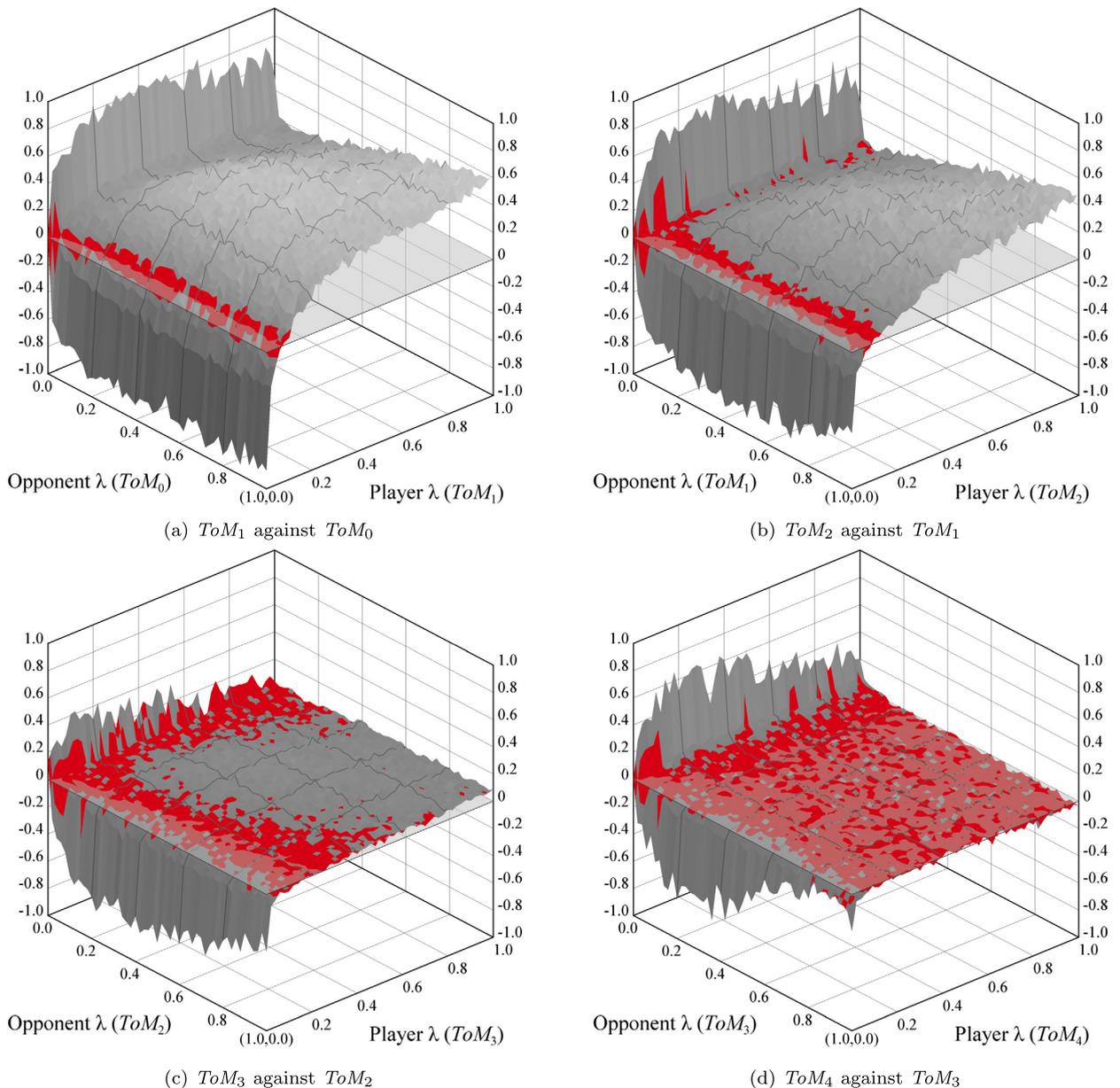
Unlike the variations on rock–paper–scissors, Limited Bidding is an extensive form game that spans several rounds. Although there is a unique best-response to each opponent action, there are multiple responses that yield a positive outcome.



**Fig. 8.** Average performance of theory of mind agents playing RPSLS against opponents of a lower order of theory of mind. Performance was averaged over 500 trials of 20 consecutive games each. Insignificant results ( $p > 0.01$ ) are highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To determine the advantage of having the ability to explicitly represent mental states of others in the game of LB, agents that differ in their order of theory of mind have been placed in competition. Fig. 9 shows the performance of theory of mind agents as a function of the learning speed  $\lambda_i$  of the focal agent and the learning speed  $\lambda_j$  of his opponent. Performance has been normalized to range from 1, which means that the focal agent achieved the maximum possible payoff, to  $-1$ , in which case his opponent achieved the maximum possible payoff. As before, lighter areas highlight that the agent performed better than his opponent, while darker areas show that his opponent obtained a higher average score.

Fig. 9a shows that  $ToM_1$  agents predominantly obtain a positive score when playing against  $ToM_0$  opponents. A  $ToM_1$  agent performs well when facing an opponent that does not learn, as shown by the high scores when the opponent's learning speed is zero ( $\lambda_j = 0$ ). The bright area along the line of equal learning speeds indicates that the advantage of the  $ToM_1$  agent is also 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. Fig. 9a 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  $\lambda_i > 0.08$ .



**Fig. 9.** Average performance of theory of mind agents playing Limited Bidding against opponents of a lower order of theory of mind. Performance was averaged over 50 trials of 50 consecutive games each. Insignificant results ( $p > 0.01$ ) are highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As for the cases of RPS and ERPS described above, applying theory of mind appears to be least effective when a  $ToM_1$  agent is playing against a  $ToM_0$  opponent that has a low learning speed. In LB, a  $ToM_0$  opponent with high learning speed changes her beliefs radically, but with high confidence. That is, the effect of the random initialization of beliefs has less impact on opponent behaviour when her learning speed is high than when her learning speed is low. For a  $ToM_1$  agent, a  $ToM_0$  opponent with a high learning speed represents a more predictable situation, which he can use to his advantage.

Fig. 9b shows that a  $ToM_2$  agent is at an advantage over a  $ToM_1$  opponent. However, although Fig. 9b shows many of the same features as Fig. 9a, such as the brighter area along the main diagonal of equal learning speeds,  $ToM_2$  agents playing against  $ToM_1$  opponents obtain a score that is on average 0.13 lower than the score of  $ToM_1$  agents playing against  $ToM_0$  agents. As a result, a  $ToM_2$  agent needs a higher learning speed of at least  $\lambda_i > 0.12$  in order to obtain, on average, a positive score when playing against a  $ToM_1$  agent. Note that like a  $ToM_1$  agent, a  $ToM_2$  agent has more difficulty obtaining an advantage when playing against an opponent with low learning speed than when her learning speed is high.

Similar to the results found for the variations on RPS, the application of first-order and second-order theory of mind present an agent with a clear advantage over opponents of a lower order of theory of mind. However, the advantage of a

$ToM_3$  agent over a  $ToM_2$  opponent is only marginal. Fig. 9c shows that a  $ToM_3$  agent barely outperforms a  $ToM_2$  agent, with an average score that only exceeds 0.1 when the  $ToM_2$  opponent has zero learning speed. Moreover, although it appears as if a  $ToM_3$  agent can still on average obtain a positive score when his learning speed is at least  $\lambda_i > 0.32$ , Fig. 9c shows that when the  $ToM_2$  opponent has learning speed  $0 < \lambda_j < 0.1$ , performance of the  $ToM_3$  agent may still fall below the plane of zero performance. That is, a  $ToM_3$  agent is no longer guaranteed to win when playing against a  $ToM_2$  opponent for any value of his learning speed  $\lambda_j$ .

Fig. 9d shows that a  $ToM_4$  agent fails to obtain an advantage of any kind over a  $ToM_3$  agent when playing LB. When neither the agent nor his opponent learns at a low speed, the game will, on average, end in a tie. The learning speed of the agent and the learning speed of his opponent do not have a strong effect on the expected outcome of the game.

In summary, agent performance in LB clearly shows diminishing returns on higher orders of theory of mind. The use of first-order and second-order theory of mind allows agents to obtain a reliable advantage over opponents that are more limited in their ability to explicitly represent mental states of others. However, a specialized system for third-order theory of mind barely allows  $ToM_3$  agents to outperform  $ToM_2$  agents, while a fourth-order theory of mind does not yield an agent any advantage that could not have been obtained with a third-order theory of mind. Qualitatively, the results are similar to those described for the RPS game in Section 5.1.

### 5.5. Summary of results

To determine the effectiveness of theory of mind, we simulated computational theory of mind agents, as described in Section 4, playing competitive games against one another. In hypothesis  $H_{RPS}$  (Section 2.4), we predicted that higher orders of theory of mind would benefit agents in competitive settings. Our results support this conclusion in the sense that the ability to make use of first-order and second-order theory of mind allows agents to obtain a clear advantage over opponents of a lower order of theory of mind. However, for orders of theory of mind beyond the second, the additional advantage is marginal.

This pattern of results was consistent across the variations on rock–paper–scissors we investigated. As we predicted in hypothesis  $H_{ERPS}$ , the larger action space of elemental rock–paper–scissors was advantageous for higher-order theory of mind agents in some instances. However, the larger action space did not remove the diminishing returns on higher orders of theory of mind. Qualitatively similar results were found for the multi-stage limited bidding game, which confirms hypothesis  $H_{LB}$ .

The relatively limited advantage of  $ToM_3$  agents playing against  $ToM_2$  opponents appears to be caused by the model that the  $ToM_2$  opponent holds of the  $ToM_3$  agent. Agents start out by playing as if they were  $ToM_0$  agents. When a  $ToM_3$  agent is in competition with a  $ToM_2$  opponent, both of them will notice that their predictions based on first-order theory of mind are correct. This causes both agents to grow more confident in application of first-order theory of mind. As a result, they both gradually start to play more as if they were  $ToM_1$  agents. When this happens, predictions based on first-order theory of mind will become less accurate, but predictions based on second-order theory of mind become increasingly accurate, increasing confidence in the application of second-order theory of mind. Both the agent and his opponent will therefore start playing as if they were  $ToM_2$  agents. At this point, the  $ToM_2$  opponent can no longer model the behaviour of the agent. That is, she will notice that none of her predictions are correct. Because of this, she will lose confidence in the application of both first-order and second-order theory of mind, and gradually start to play as if she were a  $ToM_0$  agent again. When the  $ToM_3$  agent tries to take advantage of this by playing as if he were a  $ToM_1$  agent, the  $ToM_2$  opponent is once again able to recognize this behaviour, and she will grow more confident in her predictions based on second-order theory of mind again. This causes the  $ToM_2$  opponent to constantly keep changing her strategy, which hinders the  $ToM_3$  agent in his efforts of trying to model her behaviour.

The relation between the performance of a theory of mind agent and the predictability of his opponent's behaviour is also reflected in the results of the rock–paper–scissors–lizard–Spock game. As predicted in hypothesis  $H_{RPSLS}$ , higher-order theory of mind agents perform more poorly in this game than in RPS and ERPS.

## 6. Discussion and conclusion

The Machiavellian intelligence hypothesis [38] on the evolution of theory of mind predicts that there are competitive settings in which the use of higher-order theory of mind presents individuals with an evolutionary advantage. But the benefits of making use of higher-order theory of mind may not always outweigh the costs. For example, in settings in which a pure-strategy Nash equilibrium exists, individuals that make use of theory of mind are unlikely to outperform individuals that play the Nash strategy without explicitly reasoning about their opponent's mental states. In other cases, simple heuristics may be superior to methods that rely on sophisticated cognitive abilities like theory of mind [70,71]. However, humans possess the ability to make use of higher-order theory of mind, which suggests that there may be settings in which this cognitively demanding skill is useful. For example, in using secret codes or negotiating climate change control, heuristics alone may not be enough.

In this paper, we have used agent-based models to show how the ability to make use of theory of mind can present individuals with an advantage over opponents that lack such an ability in certain competitive settings. The advantage was found to be qualitatively similar across the four competitive games we discussed, which included repeated single-shot

games rock–paper–scissors, elemental rock–paper–scissors, rock–paper–scissors–lizard–Spock, and the repeated extensive form game limited bidding.

To our surprise, the results show diminishing returns on higher orders of theory of mind. Although both first-order and second-order theory of mind agents clearly outperform opponents that are more limited in their abilities to represent mental content of others, third-order theory of mind agents only marginally outperform second-order theory of mind opponents. Fourth-order theory of mind was only found to be beneficial under specific circumstances. These diminishing returns on higher orders of theory of mind were found not to be related to the number of actions available to the agents. Increasing the action space from which agents choose did not increase performance of a third-order theory of mind agent in competition with a second-order theory of mind opponent.

Although theory of mind allows agents to outperform opponents that are more limited in their ability to explicitly represent mental states, theory of mind may not always be an efficient use of memory capacity. Additional experiments show that in simple games such as rock–paper–scissors, an agent seems to benefit more from remembering past behaviour of his opponent rather than representing her mental states. However, for more complex games such as Limited Bidding, theory of mind appears to have benefits that go beyond remembering past opponent behaviour. Agents that are capable of both associative learning strategies and theory of mind strategies may therefore choose not to use their theory of mind when the task is simple. Tasks may need to be sufficiently complex to elicit a theory of mind response.

In our model, we have assumed that agents choose what action to perform rationally. That is, agents choose to perform the action that they believe to yield them the highest possible payoff. This results in a predictability that benefits theory of mind agents, as shown by our results in the game of rock–paper–scissors–lizard–Spock. When an opponent is indifferent between two actions in the sense that both actions maximize the expected payoff, the effectiveness of theory of mind suffers. However, when there is a slight asymmetry between the two actions, such that one action appears to be a slightly better alternative than the other, this creates a focal point [72] for agents. In this case, the opponent will choose the action that she believes to yield the better payoff. However, this behaviour can be predicted by higher-order theory of mind agents.

An agent of a lower order of theory of mind may therefore be able to avoid falling victim to an opponent capable of theory of mind of a higher order when he does not choose what action to play completely rationally. For example, agents could choose the action to perform with a probability proportional to the expected payoff. Similarly, utility proportional beliefs [73] may benefit the effectiveness of theory of mind agents, through the belief that opponents choose an action proportionally to its utility. In this case, the theory of mind agent is less reliant on his opponent playing completely rationally. Future research may reveal how a balance can be achieved between exploiting weaknesses in the opponent's actions, while remaining unpredictable enough to avoid exploitation.

In our model, a zero-order theory of mind agent does not believe that his opponent behaves randomly [10,11], but attempts to model the opponent's behaviour by assuming her past actions predict what she will do in the future. A higher-order theory of mind agent therefore simultaneously updates his model of the mental content of the opponent and his belief about the opponent's theory of mind abilities. It would be interesting to compare the effectiveness of theory of mind in direct competition with more classical strategies and heuristics.

In future work, we aim to investigate whether theory of mind is effective in more complex interaction settings including various partners as well. Theory of mind may play an important role in cooperative settings, for example in teamwork, as well as mixed-motive settings such as negotiations (cf. [37]). This may provide further insights for automated agents that share their environment with human agents, such as in automated negotiation [3,4].

## Acknowledgements

This work was supported by the Netherlands Organisation for Scientific Research (NWO) Vici grant NWO 277-80-001, awarded to Rineke Verbrugge for the project 'Cognitive systems in interaction: Logical and computational models of higher-order social cognition'. We would like to thank the three anonymous reviewers for their helpful comments.

## References

- [1] H. de Weerd, B. Verheij, The advantage of higher-order theory of mind in the game of limited bidding, in: J. van Eijck, R. Verbrugge (Eds.), Proc. Workshop Reason. Other Minds: Log. Cogn. Perspect., CEUR Workshop Proceedings, 2011, pp. 149–164.
- [2] H. de Weerd, R. Verbrugge, B. Verheij, Higher-order social cognition in the game of rock–paper–scissors: A simulation study, in: G. Bonanno, H. van Ditmarsch, W. van der Hoek (Eds.), Proc. 10th Conf. Log. Found. Game Decis. Theory, 2012, pp. 218–232.
- [3] S. Kraus, Negotiation and cooperation in multi-agent environments, *Artif. Intell.* 94 (1997) 79–97.
- [4] R. Lin, S. Kraus, J. Wilkenfeld, J. Barry, Negotiating with bounded rational agents in environments with incomplete information using an automated agent, *Artif. Intell.* 172 (2008) 823–851.
- [5] R. Fagin, J. Halpern, Y. Moses, M. Vardi, Reasoning About Knowledge, MIT Press, Cambridge, MA, 1995; second edition 2003.
- [6] H. van Ditmarsch, W. van der Hoek, B. Kooi, Dynamic Epistemic Logic, Springer, 2007.
- [7] P. Gmytrasiewicz, E. Durfee, A rigorous, operational formalization of recursive modeling, in: Proc. First Int. Conf. on Multi-Agent Syst., 1995, pp. 125–132.
- [8] P. Gmytrasiewicz, P. Doshi, A framework for sequential planning in multiagent settings, *J. Artif. Intell. Res.* 24 (2005) 49–79.
- [9] A. Pfeffer, Networks of influence diagrams: A formalism for representing agents' beliefs and decision-making processes, *J. Artif. Intell. Res.* 33 (2008) 109–147.
- [10] W. Yoshida, R. Dolan, K. Friston, Game theory of mind, *PLoS Comput. Biol.* 4 (2008) e1000254.

- [11] C. Camerer, T. Ho, J. Chong, A cognitive hierarchy model of games, *Q. J. Econ.* 119 (2004) 861–898.
- [12] D. Stahl, P. Wilson, On players' models of other players: Theory and experimental evidence, *Games Econ. Behav.* 10 (1995) 218–254.
- [13] M. Bacharach, D.O. Stahl, Variable-frame level-*n* theory, *Games Econ. Behav.* 32 (2000) 220–246.
- [14] H. Simon, A mechanism for social selection and successful altruism, *Science* 250 (1990) 1665–1668.
- [15] D. Kahneman, Maps of bounded rationality: Psychology for behavioral economics, *Am. Econ. Rev.* (2003) 1449–1475.
- [16] D. Premack, G. Woodruff, Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1 (1978) 515–526.
- [17] J. Perner, H. Wimmer, "John thinks that Mary thinks that...". Attribution of second-order beliefs by 5 to 10 year old children, *J. Exp. Child Psychol.* 39 (1985) 437–471.
- [18] T. Hedden, J. Zhang, What do you think I think you think?: Strategic reasoning in matrix games, *Cognition* 85 (2002) 1–36.
- [19] L. Flobbe, R. Verbrugge, P. Hendriks, I. Krämer, Children's application of theory of mind in reasoning and language, *J. Log. Lang. Inf.* 17 (2008) 417–442.
- [20] B. Meijering, H. van Rijn, N. Taatgen, R. Verbrugge, I do know what you think I think: Second-order theory of mind in strategic games is not that difficult, in: *Proc. 33rd Annu. Conf. Cogn. Sci. Soc.*, 2011, pp. 2486–2491.
- [21] V. Crawford, N. Iriberry, Fatal attraction: Saliency, naïveté, and sophistication in experimental "Hide-and-Seek" games, *Am. Econ. Rev.* (2007) 1731–1750.
- [22] H. Wimmer, J. Perner, Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception, *Cognition* 13 (1983) 103–128.
- [23] I. Apperly, *Mindreaders: The Cognitive Basis of "Theory of Mind"*, Psychology Press, Hove, UK, 2011.
- [24] B. Meijering, L. Van Maanen, H. Van Rijn, R. Verbrugge, The facilitative effect of context on second-order social reasoning, in: *Proc. 32nd Annu. Conf. Cogn. Sci. Soc.*, 2010, pp. 1423–1429.
- [25] M. Tomasello, *Why We Cooperate*, MIT Press, Cambridge, MA, 2009.
- [26] M. Schmelz, J. Call, M. Tomasello, Chimpanzees know that others make inferences, *Proc. Natl. Acad. Sci. USA* 108 (2011) 3077–3079.
- [27] J. Burkart, A. Heschl, Understanding visual access in common marmosets, *Callithrix jacchus*: Perspective taking or behaviour reading? *Anim. Behav.* 73 (2007) 457–469.
- [28] J. Kaminski, J. Call, M. Tomasello, Goats' behaviour in a competitive food paradigm: Evidence for perspective taking? *Behaviour* 143 (2006) 1341–1356.
- [29] J. Kaminski, J. Brauer, J. Call, M. Tomasello, Domestic dogs are sensitive to a human's perspective, *Behaviour* 146 (2009) 979–998.
- [30] N. Clayton, J. Dally, N. Emery, Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist, *Philos. Trans. R. Soc. B, Biol. Sci.* 362 (2007) 507.
- [31] T. Bugnyar, Knower-guesser differentiation in ravens: Others' viewpoints matter, *Proc. R. Soc. B, Biol. Sci.* 278 (2011) 634–640.
- [32] D. Penn, D. Povinelli, On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind', *Philos. Trans. R. Soc. B, Biol. Sci.* 362 (2007) 731.
- [33] P. Carruthers, Meta-cognition in animals: A skeptical look, *Mind Lang.* 23 (2008) 58–89.
- [34] E. van der Vaart, R. Verbrugge, C. Hemelrijk, Corvid re-caching without 'theory of mind': A model, *PLoS ONE* 7 (2012) e32904.
- [35] M. Balter, 'Killjoys' challenge claims of clever animals, *Science* 335 (2012) 1036–1037.
- [36] B. Hare, J. Call, M. Tomasello, Do chimpanzees know what conspecifics know? *Anim. Behav.* 61 (2001) 139–151.
- [37] R. Verbrugge, Logic and social cognition: The facts matter, and so do computational models, *J. Philos. Log.* 38 (2009) 649–680.
- [38] A. Whiten, R. Byrne, *Machiavellian Intelligence II: Extensions and Evaluations*, Cambridge University Press, Cambridge, 1997.
- [39] J. Epstein, *Generative Social Science: Studies in Agent-based Computational Modeling*, Princeton University Press, Princeton, NJ, 2006.
- [40] J. Epstein, Agent-based computational models and generative social science, *Complexity* 4 (1999) 41–60.
- [41] W. Jäger, R. Popping, H. Van de Sande, Clustering and fighting in two-party crowds: Simulating the approach-avoidance conflict, *J. Artif. Soc. Soc. Simul.* 4 (2001) 1–18.
- [42] M. Harbers, R. Verbrugge, C. Sierra, J. Debenham, The examination of an information-based approach to trust, in: *Coord., Organ., Inst., and Norms in Agent Syst. III*, 2008, pp. 71–82.
- [43] E. van der Vaart, B. de Boer, A. Hankel, B. Verheij, Agents adopting agriculture: Modeling the agricultural transition, in: *Proc. 9th Int. Conf. from Anim. to Animats: Simul. Adapt. Behav.*, 2006, pp. 750–761.
- [44] H. Gintis, Strong reciprocity and human sociality, *J. Theor. Biol.* 206 (2000) 169–179.
- [45] R. Boyd, H. Gintis, S. Bowles, P. Richerson, The evolution of altruistic punishment, *Proc. Natl. Acad. Sci.* 100 (2003) 3531–3535.
- [46] H. de Weerd, R. Verbrugge, Evolution of altruistic punishment in heterogeneous populations, *J. Theor. Biol.* 290 (2011) 88–103.
- [47] A. Cangelosi, D. Parisi, *Simulating the Evolution of Language*, Springer, 2002.
- [48] B. de Boer, *The Origins of Vowel Systems*, Oxford University Press, USA, 2001.
- [49] I. Slingerland, M. Mulder, E. van der Vaart, R. Verbrugge, A multi-agent systems approach to gossip and the evolution of language, in: *Proc. 31st Annu. Meet. Cogn. Sci. Soc.*, 2009, pp. 1609–1614.
- [50] D. Billings, The first international RoShamBo programming competition, *ICGA J.* 23 (2000) 42–50.
- [51] D. Egnor, locaine powder, *ICGA J.* 23 (2000) 33–35.
- [52] J. Von Neumann, Zur Theorie der Gesellschaftsspiele, *Math. Ann.* 100 (1928) 295–320.
- [53] K. Binmore, *Playing for Real*, Oxford University Press, Oxford, UK, 2007.
- [54] W. Wagenaar, Generation of random sequences by human subjects: A critical survey of literature, *Psychol. Bull.* 77 (1972) 65.
- [55] A. Rapoport, D. Budescu, Randomization in individual choice behavior, *Psychol. Rev.* 104 (1997) 603.
- [56] R. West, C. Lebiere, D. Bothell, Cognitive architectures, game playing, and human evolution, in: *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*, Cambridge University Press, 2006, pp. 103–123.
- [57] R. Cook, G. Bird, G. Lünsler, S. Huck, C. Heyes, Automatic imitation in a strategic context: Players of rock-paper-scissors imitate opponents' gestures, *Proc. R. Soc. B, Biol. Sci.* (2011).
- [58] S. Kass, K. Bryla, Rock paper scissors Spock lizard, <http://www.samkass.com/theories/RPSSL.html>, 2009, accessed 29/12/2012.
- [59] E. De Bono, *Edward de Bono's Super Mind Pack: Expand Your Thinking Powers with Strategic Games & Mental Exercises*, Dorling Kindersley Publishers Ltd., London, UK, 1998.
- [60] M. Osborne, A. Rubinstein, *A Course in Game Theory*, MIT Press, Cambridge, MA, 1994.
- [61] C. Bicchieri, Common knowledge and backward induction: A solution to the paradox, in: *Proc. 2nd Conf. Theor. Asp. Reason. Knowl.*, 1988, pp. 381–393.
- [62] J. Von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, 1944, commemorative edition, 2007.
- [63] M. Davies, The mental simulation debate, *Philos. Issues* 5 (1994) 189–218.
- [64] S. Nichols, S. Stich, *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford University Press, USA, 2003.
- [65] S. Hurley, The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading, *Behav. Brain Sci.* 31 (2008) 1–22.
- [66] R. Falk, C. Konold, Making sense of randomness: Implicit encoding as a basis for judgment, *Psychol. Rev.* 104 (1997) 301.
- [67] J. Barwise, On the model theory of common knowledge, in: *The Situation in Logic*, CSLI Press, Stanford, CA, 1989, pp. 201–220.

- [68] H. van Ditmarsch, J. van Eijck, R. Verbrugge, Common knowledge and common belief, in: J. van Eijck, R. Verbrugge (Eds.), *Discourses on Social Software*, Amsterdam University Press, Amsterdam, 2009, pp. 99–122.
- [69] R. Brown, *Smoothing, Forecasting and Prediction of Discrete Time Series*, Prentice–Hall, Englewood Cliffs, NJ, 1963.
- [70] G. Gigerenzer, R. Hertwig, T. Pachur, *Heuristics: The Foundations of Adaptive Behavior*, Oxford University Press, 2011.
- [71] D. Kahneman, *Thinking, Fast and Slow*, Farrar, Straus and Giroux, New York, 2011.
- [72] R. Sugden, A theory of focal points, *Econ. J.* 105 (1995) 533–550.
- [73] C. Bach, A. Perea, Utility proportional beliefs, <http://epicenter.name/Research.html>, 2011, accessed 29/12/2012.