Higher-order social cognition in rock-paper-scissors: A simulation study

Harmen de Weerd, Rineke Verbrugge, Bart Verheij
Department of Artificial Intelligence, University of Groningen, P.O. Box 407, 9700 AK, Groningen

Abstract

Opponent modeling in multi-agent game playing and decision making allows agents to recursively model their opponent, creating increasingly complex models of increasingly sophisticated opponents. Human participants show this ability to predict the actions of others through theory of mind, by explicitly attributing unobservable mental content such as beliefs, desires, and intentions to an opponent. However, whereas recursive opponent modeling could continue indefinitely, humans show difficulties in using higher orders of theory of mind. One of the explanations for the limited recursive use of theory of mind in humans can be that recursion is initially advantageous, but that there is little advantage of deeper recursion beyond a certain point. Our previous research supports this explanation through agent-based simulations in the specific setting of a dynamic game, which shows a reduction in advantage after the second step of recursion. In this paper, we aim to determine whether this effect is generalizable to other settings, such as iterated single-shot games. We investigate the advantages of theory of mind in three variations of the well-known rock-paper-scissors game. We find that 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, but that the advantage for deeper recursion to third-order theory of mind is small in comparison. Furthermore, we find that the advantage of second-order theory of mind remains for games with a larger action space when for each opponent action there exists a unique best response.

1 Introduction

In settings where multiple strategic agents perform actions that influence each other's decision making process, it is often necessary to accurately predict the behaviour of others in order to respond appropriately [19]. One option to do so is by modeling an opponent explicitly, e.g. through dynamic epistemic logic [21, 29, 30], epistemic game theory [5, 6], interactive POMDPs [16], networks of influence diagrams [24], or iterated best-response models such as cognitive hierarchy models [9] and level-n theory [2]. 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.

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 as *theory of mind*. Experiments in which humans play games show evidence that humans use theory of mind recursively in their decision making process [13, 17, 18, 23]. For example, when asked to search for a hidden object in one of four boxes, three of which are labeled 'A' and one of them 'B', participants tend to ignore the box labeled 'B', 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 the box labeled 'B' [10]. The ability to use higher-order theory of mind is well established for humans, and may indicate an evolutionary advantage that justifies the recursive use of theory of mind. However, whereas humans make use of higher-order theory of mind, the use of theory of mind of any kind by non-human species is a controversial matter [3, 22, 27, 28]. Also, whereas recursive opponent modeling could continue indefinitely, children between the ages of 6 and 8 struggle to apply higher-order theory of mind [13], and even adults use higher-order theory of mind only up to a certain point [13, 17, 31].

In this paper, we consider agent-based computational models to investigate the advantages of making use of higher-order theory of mind. Agent-based modeling has proven its usefulness as a research tool to investigate how behavioral patterns emerge from the interactions between individuals (cf. [12]). In Evolutionary Game Theory, agent-based models have been used to explain the existence of cooperative behaviour as well as the influence of local interactions in a spatial structure on the global distribution of strategies [15, 20]. We have previously used the approach of agent-based modeling to investigate the effectiveness of recursive opponent modeling in the setting of a specific game called Limited Bidding [11]. Limited Bidding is a two-player extensive form game, in which each player is given an identical set of five tokens of different values. Over the course of five rounds, players simultaneously choose a token to play, with the exception that each token may only be used in one round per game. A round is won by the agent that played the token with the highest value, while the game is won by the agent that won most rounds. The results in this setting show that opponent modeling in the form of first-order theory of mind is successful against naïve agents, and second-order theory of mind is successful against both naïve and first-order theory of mind agents. However, continued recursion, to third-order theory of mind and higher, provides the agent with limited additional advantages [11].

Our results from Limited Bidding support the Machiavellian intelligence hypothesis[1] [33] in the sense that theory of mind can give individuals a competitive advantage over their opponent when playing Limited Bidding, but it is unclear whether these effects are caused by the specific game structure. In this paper, we apply the model that we presented in [11] to several variations of repeated rock-paper-scissors (RPS). These games are outlined in Section 2. Section 3 describes how theory of mind may be used to obtain an advantage over less sophisticated opponents within our setting. Section 4 gives a detailed description of the theory of mind agents that are placed in competition with one another. The results of a tournament between agents with different abilities for theory of mind are presented in Section 5. Finally, Section 6 discusses our findings.

2 Outline of three games

We test the effectiveness of theory of mind across three variations of the rock-paper-scissors game. In the following subsections, we will present these variations in more detail and present hypotheses for our agent-based simulations. These games are in a special class of games, in which there is no outcome that is satisfactory for both players. For each outcome, there is at least one player who could have done better by selecting a different action. That is, these games do not have a pure-strategy Nash equilibrium. We believe opponent modeling to be especially relevant in this class of games. These games are therefore a suitable starting point to investigate the possible advantages of higher-order theory of mind.

2.1 Rock-paper-scissors

The game of *rock-paper-scissors* (RPS) [32] 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 players choose the same action, the game ends in a draw. 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. [4]) in which the player chooses each of the options with equal probability. Although randomizing over the possible actions has an expected utility of zero, a randomizing player prevents his opponent from taking advantages of regularities in his strategy. However, experimental evidence suggests that human participants are poor at generating random sequences [25], and do not always play as predicted by game theory [8].

If theory of mind presents individuals with an advantage over their competitors, our results for RPS should show that agents with a higher order of theory of mind outperform opponents that are more limited in their ability to make use of theory of mind.

2.2 Elemental rock-paper-scissors

Elemental rock-paper-scissors (ERPS)[2] is an extension of RPS that includes five actions, 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

[1]For a discussion of alternative hypotheses for the evolution of theory of mind, see [31].
[2]Based on the five elements in Chinese philosophy.

	Rock	Paper	Scissors
Rock	0	-1	1
Paper	1	0	-1
Scissors	-1	1	0

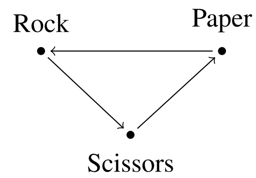


Table 1: **(RPS)** Payoff table and graph representation for the rock-paper-scissors game. The table shows the payoff for the player choosing the row action ‘rock’, ‘paper’ or ‘scissors’, for every possible choice of the player choosing the column action. Arrows in the graph are read as ‘defeats’. For example, the arrow from ‘paper’ to ‘rock’ means that ‘paper’ defeats ‘rock’.

	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

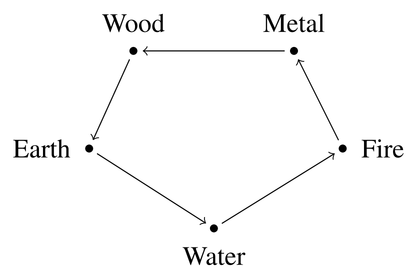


Table 2: **(ERPS)** Payoff table and graph representation for the elemental rock-paper-scissors game, in which agents choose between five different actions. The table shows the payoff for the player choosing the row action ‘wood’, ‘metal’, ‘fire’, ‘water’, or ‘earth’, for every possible choice of the 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’.

	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

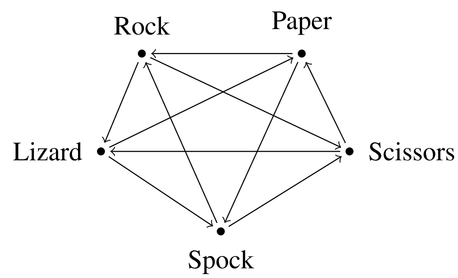


Table 3: **(RPSLS)** Payoff table and graph representation for the rock-paper-scissors-lizard-Spock game. The table shows the payoff for the player choosing the row action ‘rock’, ‘paper’, ‘scissors’, ‘lizard’ and ‘Spock’, for every possible choice of the player choosing the column action. Arrows in the graph are read as ‘defeats’. For example, the arrow between ‘lizard’ and ‘Spock’ means that ‘lizard’ defeats ‘Spock’.

that an opponent may play, there is a unique best response that yields the agent a positive outcome.

The differences in performance of agents playing RPS and those playing ERPS should provide insight in the way theory of mind performance is related to the structure of the game. In ERPS, the predicted behaviour of the opponent is not limited to the three possible actions of RPS. We expect that the larger action space allows higher-order agents to perform better in ERPS than in RPS. However, the additional actions may make it more difficult to accurately predict the opponent, which may result in a decrease in performance of theory of mind agents playing ERPS, as compared to those playing RPS.

2.3 Rock-paper-scissors-lizard-Spock

Rock-paper-scissors-lizard-Spock (RPSLS)[3] is an extension of RPS with a different interaction topology. 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. If an agent predicts its opponent to play 'paper', the agent has no preference for playing either 'scissors' or 'lizard', since both of them will defeat 'paper' equally well. As a result, we expect that the application of increasingly higher orders of theory of mind will result in less of an advantage for the agent than in the cases of RPS and ERPS, in which the best response to a predicted action is unique.

3 Simulation theory of mind

An agent may try to take advantage of regularities in its opponent's strategy by predicting its behaviour. In human participants, one way of generating such predictions is by using simulation-theory of mind [19]. In simulation-theory of mind, a player takes the perspective of his opponent, and simulates the decision making process of the opponent by making the decision for himself. Through the implicit assumption that the opponent's thought process can be accurately modeled by his own thought process, the player then predicts that the opponent will make the same decision he would have made if the roles were reversed. To avoid confusion, we will refer to a focal agent as if it were male, while we refer to opponents as if they were female.

A *zero-order* theory of mind agent is unable to model the mental content of his opponent, including beliefs, desires and intentions. Specifically, a zero-order theory of mind agent cannot represent that his opponent has goals different from his own. When predicting his opponent's behaviour, the agent is therefore limited to his memory of previous events. If, for example, the zero-order theory of mind agent remembers his opponent to have played 'rock' in previous encounters, a natural conclusion would be that his opponent is going to play 'rock' again. Given this belief, the zero-order theory of mind agent would therefore adjust his behaviour to play 'paper' more often. Anecdotal evidence of RPS tournaments held with human participants suggests that inexperienced or frustrated players tend to play the action that would have won in the last round. That is, they play as if they were zero-order theory of mind agents.

In contrast to a zero-order theory of mind, a *first-order* theory of mind 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 first-order theory of mind agent. To predict his opponent's behaviour, the first-order theory of mind agent puts himself in the position of his opponent. For example, suppose that the first-order theory of mind agent remembers that he previously played 'rock' against the same opponent. He realizes that if the roles were reversed, and he would remember his opponent to play 'rock' before, the agent would play 'paper' more often. The first-order theory of mind agent has the ability to attribute this thought process to his opponent, and predict that she will play 'paper' more often. Given this prediction, the first-order theory of mind agent should play 'scissors' more often.

Although the first-order theory of mind agent models a zero-order theory of mind opponent, agents in our setup do not know the extent of the abilities of their opponents with certainty. Through repeated interaction, a first-order theory of mind agent may come to believe that his opponent is not a zero-order theory of mind agent, and that she has no beliefs at all. Such an opponent without beliefs could, for example, play 'rock' irrespective of what the first-order theory of mind agent has previously played. Based on this belief, a first-order theory of mind agent can choose to play as if he were a zero-order theory of mind agent.

Similar to the way a first-order theory of mind agent models his opponent as having zero-order theory of mind, a *second-order* theory of mind agent models a first-order theory of mind opponent. A second-order theory of mind agent thus considers the possibility that his opponent is putting herself in his position, and is modeling him as a zero-order theory of mind agent. If the second-order theory of mind agent remembers his opponent to have played 'rock' in previous encounters, he would therefore believe his opponent to predict that he will be playing 'paper' more often. As a result, the second-order theory of mind agent would predict his opponent to play 'scissors' more often, in which case the agent should play 'rock' more often himself.

[3]Originally by Sam Kass and Karen Bryla.

Order of theory of mind k	
	0
$b^{(0)}(R)$	0.5
$b^{(0)}(P)$	0.3
$b^{(0)}(S)$	0.2

(a) Mental content of the *ToM*$_0$ agent

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

(b) Mental content of the *ToM*$_1$ agent

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

(c) Mental content of the *ToM*$_2$ agent

Table 4: Mental content of (a) the *ToM*$_0$ agent in Example 1, (b) the *ToM*$_1$ agent in Example 2, and (c) the *ToM*$_2$ agent in Example 3, in a game of rock-paper-scissors.

4 An agent-based model of theory of mind

The agents that have been described in Section 3 have been implemented as computational agents. In this section, we discuss the implementation of thse types of computational agents that play the three games RPS, ERPS and RPSLS. We closely follow the model presented for Limited Bidding [11], in which agents differ in their ability to explicitly represent beliefs, and therefore in their ability to make use of theory of mind. In the remainder, the shorthand *ToM*$_i$ will be used to refer to an agent that has abilities up to and including ith-order theory of mind, but not higher.

4.1 Zero-order theory of mind

Theory of mind agents differ in their order of theory of mind, which specifies the maximum nesting of beliefs that they can represent, and their learning speed $0 \leq \lambda \leq 1$. A zero-order theory of mind agent models his opponent's behaviour, but cannot represent her beliefs. A zero-order theory of mind agent models his opponent in the form of a probability distribution $b^{(0)}$ over an action space $\mathcal{S}$, such that $b^{(0)}(s)$ represents what the agent believes to be the probability that his opponent will play action $s \in \mathcal{S}$. We assume:

$$b^{(0)}(s) \geq 0 \qquad \text{for all } s \in \mathcal{S}, \tag{1}$$

$$\sum_{s \in \mathcal{S}} b^{(0)}(s) = 1 \tag{2}$$

That is, (1) agents assign non-negative probability to their opponent playing a certain action $s \in S$, and (2) the probabilities assigned to possible opponent action sum up to 1.

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 decide what action to perform. To reduce predictability of lower-order theory of mind agents, agents choose probabilistically. Specifically, the probability that an agent chooses a certain action is proportional to the probability that performing the action will result in winning the game (i.e. yield a positive payoff), given the distribution $b^{(0)}$ of opponent actions. This is represented by the function $t^*(b^{(i)})$, which is a probability distribution over the action space $\mathcal{S}$. Agents therefore do not always choose the action that they believe to yield them the highest payoff (cf. [1]). Instead, when an agent is playing against an unpredictable opponent, the agent will also play more randomly.

A *ToM*$_0$ agent's ability to predict his opponent's behaviour is limited to his beliefs $b^{(0)}$. A *ToM*$_0$ agent therefore selects what action to play by using the function t^* directly. That is, given his zero-order beliefs $b^{(0)}$, a *ToM*$_0$ agent will play $t^*(b^{(0)})$. Example 1 shows the decision process of a *ToM*$_0$ agent in a specific situation.

4.2 First-order theory of mind

A *ToM*$_1$ agent has zero-order beliefs $b^{(0)}$ of his opponent's behaviour, and additionally attributes similar beliefs to his opponent in the form of an additional probability distribution $b^{(1)}$. Here, $b^{(1)}(t)$ represents what the agent believes his opponent to judge what the probability is that he will play action $t \in \mathcal{S}$. To predict what action his opponent will

Example 1. Consider the game of rock-paper-scissors. In this game, the action space is $\mathcal{S} = \{R, P, S\}$, while the payoffs are given by Table 1.

We consider a *ToM*$_0$ agent, whose mental content is listed in Table 4a. Based on the agent's zero-order beliefs $b^{(0)}$, the agent believes that if he would play R, there is a 20% probability that he will win (i.e. that his opponent will play S). Similarly, he believes that if he should choose to play P, there is a 50% probability that he will win the game, and if he plays S, there is a 30% probability that he will win.

The agent chooses which actions to play proportionally to the probability that it will result in winning the game, given by the agent's zero-order beliefs. In the case of RPS, the probabilities of winning the game when playing R, P, or S add up to one. The probability that an agent chooses to play an action is therefore the same as the probability that this action will result in winning the game. In this case,

$$t^*(b^{(0)})(R) = b^{(0)}(S) = 0.2 \quad (3)$$

$$t^*(b^{(0)})(P) = b^{(0)}(R) = 0.5 \quad (4)$$

$$t^*(b^{(0)})(S) = b^{(0)}(P) = 0.3. \quad (5)$$

This means that the *ToM*$_0$ agent has a 20% probability of playing R, a 50% probability of playing P, and a 30% probability of playing S.

perform, a *ToM*$_1$ agent makes use of his first-order beliefs $b^{(1)}$ to determine what action he would most likely play if the situation were reversed, and the *ToM*$_1$ was faced with the decision his opponent is making. This is achieved by $\hat{s}^{(1)} = \arg\max_{s \in \mathcal{S}} t^*(b^{(1)})$.

A *ToM*$_1$ agent's first-order theory of mind produces a prediction for his opponent's next action, if she is a *ToM*$_0$ opponent. However, agents in our setting form beliefs about the order of theory of mind their opponent is reasoning at. This allows a *ToM*$_1$ agent to believe that his opponent is not reacting to his actions at all, but plays according to a strategy that does not change over repeated games. In this case, the *ToM*$_1$ agent should play as if he were a *ToM*$_0$ agent, and base his decision of what action to play solely on his zero-order beliefs $b^{(0)}$.

The extent to which first-order theory of mind governs the decisions of the *ToM*$_1$ 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 weighs his zero-order beliefs $b^{(0)}$ against the prediction $\hat{s}^{(1)}$ 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 $\hat{a}_j^{(1)}$ increased, while his integrated belief that his opponent will be playing any other action is decreased. Specifically,

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

Using the belief integration function U, the *ToM*$_1$ agent combines the prediction $\hat{s}^{(1)}$ with his zero-order beliefs $b^{(0)}$. The *ToM*$_1$ agent then decides what action to do based on these combined beliefs. That is, a *ToM*$_1$ agent decides what action to use by calculating

$$t^*\Big(U\big(b^{(0)}, \hat{s}^{(1)}, c_1\big)\Big) = t^*\left(U\Big(b^{(0)}, \arg\max_{s \in \mathcal{S}} t^*\big(b^{(1)}\big), c_1\Big)\right). \quad (7)$$

An example of the decision process of a *ToM*$_1$ agent is shown in Example 2.

4.3 Second-order theory of mind

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

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 4b. 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 $b^{(1)}(R) = 0.4$, 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 is most likely to do in her place. That is, the agent first calculates the probability that he would play each of the possible actions, if his first-order beliefs $b^{(1)}$ were actually his zero-order beliefs.

$$\begin{aligned} t^*(b^{(1)})(R) &= b^{(1)}(S) = 0.1 && (8)\\ t^*(b^{(1)})(P) &= b^{(1)}(R) = 0.4 && (9)\\ t^*(b^{(1)})(S) &= b^{(1)}(P) = 0.5 && (10) \end{aligned}$$

The *ToM*$_1$ agent's first-order theory of mind predicts that his opponent is most likely to select

$$\hat{s}^{(1)} = \arg\max_{s \in \mathcal{S}} t^*(b^{(1)}) = S. \tag{11}$$

In this case, the agent's prediction $\hat{s}^{(1)}$ conflicts with his zero-order beliefs $b^{(0)}$. According to his first-order theory of mind, his opponent is most likely going to play S, while the agent's zero-order beliefs only assigns a 20% probability that his opponent is going to play S. 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. The *ToM*$_1$ agent makes use of these integrated beliefs to decide his strategy.

$$\begin{aligned} t^*(U(b^{(0)}, S, 0.9))(R) &= U(b^{(0)}, S, 0.9)(S) = (1 - 0.9) \cdot b^{(0)}(S) + 0.9 = 0.92 && (12)\\ t^*(U(b^{(0)}, S, 0.9))(P) &= U(b^{(0)}, S, 0.9)(R) = (1 - 0.9) \cdot b^{(0)}(R) = 0.05 && (13)\\ t^*(U(b^{(0)}, S, 0.9))(S) &= U(b^{(0)}, S, 0.9)(P) = (1 - 0.9) \cdot b^{(0)}(P) = 0.03 && (14) \end{aligned}$$

This means that the *ToM*$_1$ agent has a 92% probability of playing R, a 5% probability of playing P, and a 3% probability of playing S.

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 Equation (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[4]. Based on second-order theory of mind, the *ToM*$_2$ agent therefore predicts that his opponent will be playing

$$\hat{s}^{(2)} = \arg\max_{s \in \mathcal{S}} t^* \left(U\left[b^{(1)}, \arg\max_{s \in \mathcal{S}} t^*\left(b^{(2)}\right), 0.8 \right] \right). \tag{15}$$

This prediction $\hat{s}^{(2)}$ based on second-order theory of mind is integrated with the *ToM*$_2$ agent's zero-order beliefs $b^{(0)}$ and his prediction $\hat{s}^{(1)}$ 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

[4]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.

Example 3. Consider a *ToM*$_2$ agent that plays rock-paper-scissors, and whose mental content is given in Table 4c. 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 by $b^{(2)}$. This is what the agent believes his opponent to believe his first-order beliefs to be. Firstly, the agent determines what action he would assign the highest probability if his second-order beliefs $b^{(2)}$ were actually his zero-order beliefs.

$$\arg\max_{s\in\mathcal{S}} t^*(b^{(2)}) = P \tag{17}$$

That is, the *ToM*$_2$ agent believes that a *ToM*$_1$ opponent would predict that he himself would play P.

Secondly, the *ToM*$_2$ agent determines how his *ToM*$_1$ opponent's prediction that he will be playing P 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 P. These integrated beliefs are then used to predict the opponent's behaviour.

$$\begin{aligned}
U(b^{(1)}, P, 0.8)(R) &= 0.2 \cdot b^{(1)}(R) = 0.2 \cdot 0.4 + 0.8 = 0.08 && (18)\\
U(b^{(1)}, P, 0.8)(P) &= 0.2 \cdot b^{(1)}(R) + 0.8 = 0.2 \cdot 0.5 = 0.90 && (19)\\
U(b^{(1)}, P, 0.8)(S) &= 0.2 \cdot b^{(1)}(R) = 0.2 \cdot 0.1 = 0.02 && (20)\\
\hat{s}^{(2)} &= \arg\max_{s\in\mathcal{S}} t^*\left(U(b^{(1)}, P, 0.8)\right) = S && (21)
\end{aligned}$$

Using his second-order theory of mind, the *ToM*$_2$ agent therefore predicts that his opponent will play S.

To make a decision, the agent integrates his zero-order beliefs $b^{(0)}$, his first-order prediction $\hat{s}^{(1)} = S$ (see Equation (11)), and his second-order prediction $\hat{s}^{(2)} = S$. 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 S. In this example, the agent has confidence $c_2 = 0.1$ in second-order theory of mind. This results in the following integrated beliefs:

$$\begin{aligned}
U(U(b^{(0)}, S, 0.9), S, 0.1)(R) &= 0.9 \cdot (0.1 \cdot b^{(0)}(R)) = 0.045 && (22)\\
U(U(b^{(0)}, S, 0.9), S, 0.1)(P) &= 0.9 \cdot (0.1 \cdot b^{(0)}(P)) = 0.027 && (23)\\
U(U(b^{(0)}, S, 0.9), S, 0.1)(S) &= 0.9 \cdot (0.1 \cdot b^{(0)}(S) + 0.9) + 0.1 = 0.928 && (24)
\end{aligned}$$

Based on his integrated beliefs of what the opponent is going to do, the *ToM*$_2$ agent decides his strategy. In this case, the *ToM*$_2$ agent plays R with a 92.8% probability, P with a 4.5% probability and S with a 2.7% probability.

weighs the integrated beliefs in Equation (7) against his prediction of opponent behaviour $\hat{s}^{(2)}$ based on second-order theory of mind. The resulting integrated beliefs are then used to determine the probability that the *ToM*$_2$ agent is going to play a certain action. The decision of the *ToM*$_2$ agent is therefore given by

$$t^*\left(U\left[U\left(b^{(0)}, \underbrace{\arg\max_{s\in\mathcal{S}} t^*\left(b^{(1)}\right)}_{\hat{s}^{(1)}}, c_1\right), \underbrace{\arg\max_{s\in\mathcal{S}} t^*\left(U\left[b^{(1)}, \arg\max_{s\in\mathcal{S}} t^*\left(b^{(2)}\right), 0.8\right]\right)}_{\hat{s}^{(2)}}, c_2\right]\right). \tag{16}$$

The decision process of a *ToM*$_2$ agent is illustrated in a specific setting by Example 3.

4.4 Higher orders of theory of mind

For every additional order of theory of mind available to the agent, say order i, the agent maintains an additional model of the opponent in the form of belief structure $b^{(i)}$ and a corresponding confidence c_i. These beliefs are used to model the decision process of an $(i-1)$th-order theory of mind opponent and generate a prediction $\hat{s}^{(i)}$ of her behaviour. This prediction is weighed against the integrated beliefs over the previous $(i-1)$ orders of theory of mind to reach a decision of what action to play. We restrict our investigation to *ToM*$_i$ agents for $i = \{0, 1, 2, 3, 4\}$.

A *ToM*$_3$ agent expands on the decision process of a *ToM*$_2$ agent by modeling his opponent as a *ToM*$_2$ agent. Based on third-order theory of mind, the *ToM*$_3$ agent uses his third-order beliefs $b^{(3)}$ to predict that his opponent is most likely going to perform the action

$$\hat{s}^{(3)} = \arg\max_{s\in\mathcal{S}} t^*\Big(U\Big[b^{(1)}, \arg\max_{s\in\mathcal{S}} t^*\Big(U\Big[b^{(2)}, \arg\max_{s\in\mathcal{S}} t^*\big(b^{(3)}\big), 0.8\Big]\Big), 0.8\Big]\Big). \tag{25}$$

In this prediction, the *ToM*$_3$ agent assumes that his opponent does not make use of first-order theory of mind. Equation (25) models a 'pure' *ToM*$_2$ opponent in the sense that the value for the confidence c_1 in first-order theory of mind attributed to his opponent is set to zero. Note that does not mean that the *ToM*$_3$ agent is unable to model a *ToM*$_1$ opponent. To represent the possibility that he is playing against a *ToM*$_1$ opponent, the *ToM*$_3$ agent makes use of the prediction based on application of second-order theory of mind shown in Equation (15).

The prediction $\hat{s}^{(3)}$ based on third-order theory of mind is integrated with his beliefs based on lower orders of theory of mind, such that the *ToM*$_3$ agent decides to play according to

$$\begin{aligned} t^*\Big(U\Big\{U\Big[U\Big(b^{(0)}, \underbrace{\arg\max_{s\in\mathcal{S}} t^*\big(b^{(1)}\big)}_{\hat{s}^{(1)}}, c_1\Big), \underbrace{\arg\max_{s\in\mathcal{S}} t^*\Big(U\Big[b^{(1)}, \arg\max_{s\in\mathcal{S}} t^*\big(b^{(2)}\big), 0.8\Big]\Big)}_{\hat{s}^{(2)}}, c_2\Big], \\ \underbrace{\arg\max_{s\in\mathcal{S}} t^*\Big(U\Big[b^{(1)}, \arg\max_{s\in\mathcal{S}} t^*\Big(U\Big[b^{(2)}, \arg\max_{s\in\mathcal{S}} t^*\big(b^{(3)}\big), 0.8\Big]\Big), 0.8\Big]\Big)}_{\hat{s}^{(3)}}, c_3\Big\}\Big). \end{aligned} \tag{26}$$

4.5 Belief adjustment and learning speed

Up to now, we discussed how agents of different orders of theory of mind decide what action to play, based on their current beliefs $b^{(i)}$ and confidence levels c_i. By placing himself in the position of his opponent, and viewing the game from her perspective, an agent makes predictions for the action his opponent 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. In the remainder of this sections, we describe how agents update their beliefs $b^{(i)}$ and confidence levels c_i when they observe the outcome of a game.

For each agent, the belief structures $b^{(i)}$ are initialized randomly, while confidence levels c_i are initially zero. When facing an unfamiliar opponent, an agent therefore plays as if he were a *ToM*$_0$ agent. After each round, the action $\tilde{t}$ actually performed by the agent and action $\tilde{s}$ performed by his opponent are revealed. Based on this information, an agent updates his confidence in theory of mind to reflect the accuracy of his predictions. Given an agent-specific learning speed $0 \leq \lambda \leq 1$, a *ToM*$_1$ agent updates his confidence in first-order theory of mind c_1 according to

$$c_1 := \begin{cases} (1-\lambda)\cdot c_1 & \text{if } \tilde{s} \neq \hat{s}^{(1)} \\ \lambda + (1-\lambda)\cdot c_1 & \text{if } \tilde{s} = \hat{s}^{(1)}. \end{cases} \tag{27}$$

Equation (27) shows that a *ToM*$_1$ agent increases his confidence in the use of first-order theory of mind if it yields accurate predictions, and lowers his confidence if predictions are inaccurate. For higher orders of theory of mind, a *ToM*$_i$ agent additionally adjusts his confidences c_i according to

$$c_i := \begin{cases} (1-\lambda)\cdot c_i & \text{if } \tilde{s} \neq \hat{s}^{(i)} \\ c_i & \text{if there is a } 1 \leq j < i \text{ such that } \tilde{s} = \hat{s}^{(i)} = \hat{s}^{(j)} \\ \lambda + (1-\lambda)\cdot c_i & \text{otherwise.} \end{cases} \tag{28}$$

This update is similar to the updating of the confidence in first-order theory of mind, except that for higher-order theory of mind, an agent does not change his confidence in ith-order theory of mind if there is some order of theory of mind j with $j < i$ for which both ith-order and jth-order theory of mind yield correct predictions. That is, theory

Example 4. Consider the *ToM*$_2$ agent from Example 3, whose mental content is given in Table 4c. 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. We assume that the *ToM*$_2$ agent played action $\tilde{t} = R$, and that his opponent played $\tilde{s} = S$, and that the agent learns at a learning speed of $\lambda = 0.6$.

Depending on the accuracy of the prediction of application of ith-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 predicted that the opponent would play S (Equation (11)). Since $\tilde{s} = \hat{s}^{(1)}$, first-order theory of mind resulted in an accurate prediction. The new confidence c_1, as calculated by Equation (27), becomes

$$c_1 := (1 - \lambda) \cdot c_1 + \lambda = (1 - 0.6) \cdot 0.9 + 0.6 = 0.96. \tag{29}$$

Using second-order theory of mind, the agent obtained the prediction $\hat{s}^{(2)} = S$. Second-order theory of mind therefore also correctly predicted that the opponent would play S. 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 the following holds:

$$\begin{aligned} b^{(0)}(R) &:= U(b^{(0)}, S, 0.6)(R) = (1 - 0.6) \cdot 0.5 = 0.2 & (30)\\ b^{(0)}(P) &:= U(b^{(0)}, S, 0.6)(P) = (1 - 0.6) \cdot 0.3 = 0.12 & (31)\\ b^{(0)}(S) &:= U(b^{(0)}, S, 0.6)(S) = (1 - 0.6) \cdot 0.2 + 0.6 = 0.68 & (32) \end{aligned}$$

This means that after the belief update, the agent believes that there is a 68% probability that his opponent is going to repeat the action S in the next round.

Note that the agent's second-order beliefs $b^{(2)}$ are updated in a similar way. However, the agent's first-order beliefs $b^{(1)}$ are updated using the action $\tilde{t}$ that the agent played himself.

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.

Each agent also updates his belief structures $b^{(i)}$. Since the zeroth-order beliefs $b^{(0)}$ represent the agent's beliefs concerning his opponent's behaviour, these beliefs are updated using his opponent's choice $\tilde{s}$. However, second-order beliefs $b^{(2)}$ specify what the agent believes his opponent to believe about what he believes about her actions. 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 $\tilde{s}$ as well. Conversely, 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 $\tilde{t}$. Using the belief adjustment function U, the beliefs are adjusted given an agent-specific learning speed $0 \leq \lambda \leq 1$, such that

$$\begin{aligned} b^{(i)}(s) &:= U\left(b^{(i)}, \tilde{s}, \lambda\right)(s) \qquad \text{if } i \text{ is even, and} & (33)\\ b^{(i)}(t) &:= U\left(b^{(i)}, \tilde{t}, \lambda\right)(t) \qquad \text{if } i \text{ is odd.} & (34) \end{aligned}$$

That is, agents in this setting make use of exponential smoothing [7] to model future opponent behaviour, and believe that their opponent is doing the same. The agent's learning speed λ determines how quickly it learns. A higher value of λ indicates that the agent changes his beliefs more rapidly based on new information. At the maximum of $\lambda = 1$, a *ToM*$_0$ agent believes that the previous observation is representative for future behaviour. At the other extreme of $\lambda = 0$, the agent does not change his beliefs at all when new information is presented. The equilibrium strategy of random play therefore corresponds to the behaviour of a *ToM*$_0$ agent with $\lambda = 0$ and $b^{(0)}(s) = \frac{1}{|\mathcal{S}|}$ for $s \in \mathcal{S}$.

Note that the agents we describe implicitly assume that their opponents update their beliefs using the same learning speed λ as they do themselves, and do not consider the possibility that their opponent reacts differently to new information than they do themselves. As a result, none of the agents actively tries to model the learning speed λ of their opponent.

Example 4 shows the process of belief adjustment for a specific situation. Due to the restrictions on the learning speed λ, equations (33) and (34) preserve the normalization and non-negativity of beliefs. Similarly, the confidences c_i in the application of ith-order theory of mind remain limited to the range $[0, 1]$.

5 Results

The agents described in Section 4 have been implemented in Java and their performance has been tested in competition of the games RPS, ERPS and RPSLS described in Section 2. Agents played against each other in trials that consist of 20 consecutive games. An agent's trial score is the average of the agent's game scores over all games in the trial, ranging from -1 to 1. We have compared the results for trials of 20 games to longer trials of 50 and 100 games and found no qualitative differences.

To determine the advantage of the ability to make use of higher orders of theory of mind, we simulated focal agents in competition with an opponent that was exactly one order of theory of lower than the agent. The figures in this section depict the focal agent's trial score averaged over 1000 runs, as a function of his learning speed λ_i, as well as the learning speed λ_j of his opponent. Simulation results are shown for every 0.02 step in learning speeds over the range $\lambda_i, \lambda_j \in [0, 1]$. Higher and lighter areas represent that the focal agent performed better than his opponent, while lower and darker areas show that his opponent obtained a higher average score. 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.

5.1 Rock-paper-scissors

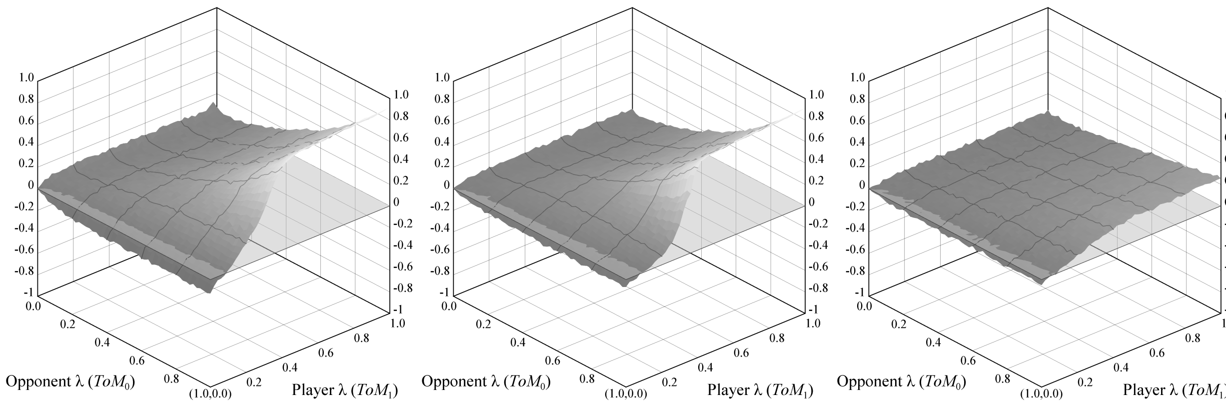
Figure 1a shows the performance of a *ToM*$_1$ agent in competition with a *ToM*$_0$ opponent in the rock-paper-scissors game. The *ToM*$_1$ agent is at a clear advantage when playing against a *ToM*$_0$ opponent. When his learning speed $\lambda_i > 0.1$, the *ToM*$_1$ agent will on average win more rounds than his *ToM*$_0$ opponent, regardless of her learning speed. The *ToM*$_1$ agent's performance is highest when both players learn at high speed, although performance levels out when his learning speed exceeds $\lambda_i = 0.5$.

As shown by Figure 2a, performance of a *ToM*$_2$ agent when playing against a *ToM*$_1$ opponent is similar to the performance of a *ToM*$_1$ agent playing against a *ToM*$_0$ opponent,. However, the *ToM*$_2$ agent is less effective at lower learning speeds. As a result, the *ToM*$_2$ agent requires a learning speed of at least $\lambda_i > 0.15$ in order to obtain a positive score on average.

Figure 3a shows that although a *ToM*$_3$ agent can still obtain an average positive score when playing RPS against a *ToM*$_2$ opponent, performance is poor compared to the performance of a *ToM*$_2$ agent playing against a *ToM*$_1$ opponent. Even in the most favourable circumstances, the performance of a *ToM*$_3$ agent never reaches 0.3. For even higher orders of theory of mind, the additional advantage is even more limited. Figure 4a shows that when a *ToM*$_4$ agent faces a *ToM*$_3$ opponent in RPS, the one with the highest learning speed is expected to win. In this case, the ability to make use of fourth-order theory of mind does not yield the agent advantages beyond those that can be obtained through third-order theory of mind.

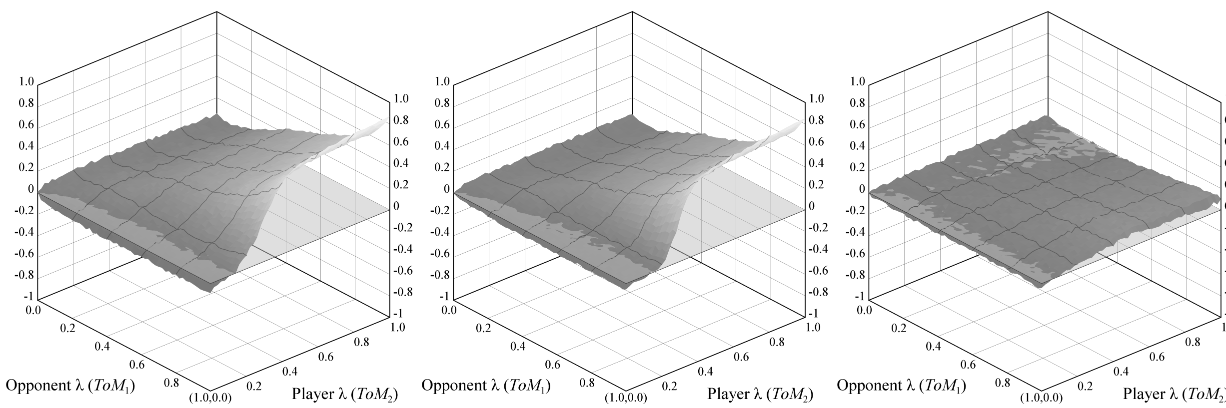
5.2 Elemental rock-paper-scissors

The results of theory of mind agents in the game of ERPS turn out to be qualitatively equivalent to the results for RPS. Figures 1b and 2b show that both the *ToM*$_1$ agent and the *ToM*$_2$ agent perform well against an opponent that is more limited in her ability to apply theory of mind. Similarly, Figure 3b shows that a *ToM*$_3$ agent performs poorly when playing against a *ToM*$_2$ opponent compared to the performance of a *ToM*$_2$ agent playing against a *ToM*$_1$ opponent, while Figure 4b shows that a *ToM*$_4$ agent only outperforms a *ToM*$_3$ opponent when he is learning at a higher rate.



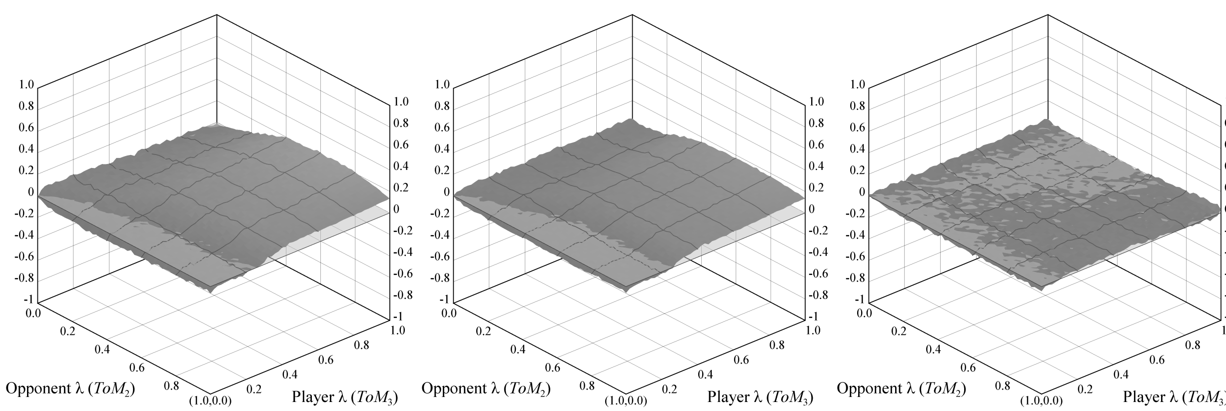
(a) Rock-paper-scissors (b) Elemental rock-paper-scissors (c) Rock-paper-scissors-lizard-Spock

Figure 1: Performance of a ToM_1 agent playing the three games against a ToM_0 opponent.



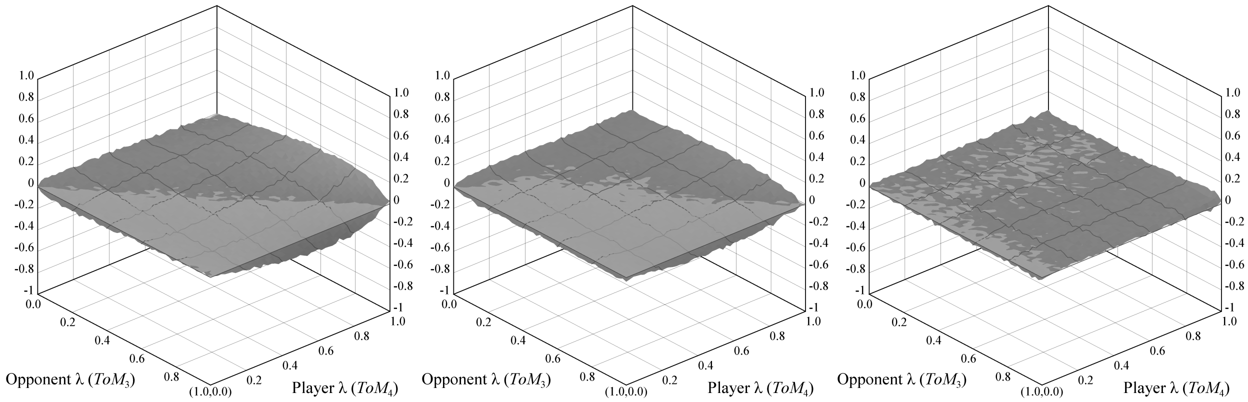
(a) Rock-paper-scissors (b) Elemental rock-paper-scissors (c) Rock-paper-scissors-lizard-Spock

Figure 2: Performance of a ToM_2 agent playing the three games against a ToM_1 opponent.



(a) Rock-paper-scissors (b) Elemental rock-paper-scissors (c) Rock-paper-scissors-lizard-Spock

Figure 3: Performance of a ToM_3 agent playing the three games against a ToM_2 opponent.



(a) Rock-paper-scissors (b) Elemental rock-paper-scissors (c) Rock-paper-scissors-lizard-Spock

Figure 4: Performance of a *ToM*$_4$ agent playing the three games against a *ToM*$_3$ opponent.

Although agent performance in the game of ERPS is almost identical to agent performance in the game of RPS, the results show that the smaller action space of RPS limits the performance of a *ToM*$_3$ agent in RPS. Figure 3a shows that, when playing the RPS game against an opponent with zero learning speed, a *ToM*$_3$ agent has an average negative score when his own learning speed is 1. In this case, the *ToM*$_3$ agent cannot distinguish correctly between an opponent that is playing a stationary mixed strategy and a *ToM*$_2$ opponent. Figure 3b shows that this is no longer the case in ERPS.

5.3 Rock-paper-scissors-lizard-Spock

Figure 1c shows the performance of a *ToM*$_1$ agent playing the game of RPSLS against a *ToM*$_0$ opponent. As the Figure shows, performance of the *ToM*$_1$ agent decreases when the best response to an action is no longer unique. Best-case performance of a *ToM*$_1$ agent playing against a *ToM*$_0$ opponent decreases from 0.9 in RPS to 0.3 in RPSLS.

The decrease in performance in RPSLS also applies to theory of mind agents of higher orders. Whereas a *ToM*$_2$ agent is guaranteed an average positive score when playing RPS or ERPS against a *ToM*$_1$ opponent, Figure 2c shows that this not true for RPSLS. When playing against a *ToM*$_1$ opponent with low learning speed, a *ToM*$_2$ agent with learning speed $\lambda_i > 0.5$ is unable to consistently win the RPSLS game. Moreover, Figure 3c shows that a *ToM*$_3$ agent that learns too quickly will on average lose the RPSLS game when facing a *ToM*$_2$ opponent that learns at a low rate.

Although Figure 4c shows that the *ToM*$_4$ agent is not guaranteed an average positive score when playing RPSLS against a *ToM*$_3$ opponent, the *ToM*$_4$ agent will not lose the RPSLS game on average to a *ToM*$_3$ opponent if his learning speed λ is high enough. In fact, the *ToM*$_4$ agent has the advantage when his *ToM*$_3$ opponent learns quickly enough. However, this advantage may be related to the specific problems faced by a *ToM*$_3$ agent that learns too quickly.

6 Discussion and conclusion

According to the Machiavellian intelligence hypothesis [33], opponent modeling through theory of mind presents individuals with a competitive advantage. In the present study, we have investigated whether such a competitive advantage for theory of mind is present through agent-based simulations. In our approach, agents are limited in their ability to explicitly represent nested beliefs. For example, a first-order theory of mind agent knows that his opponent may have beliefs about his behaviour, but is unable to represent the possibility that his opponent has beliefs about his own beliefs. This approach is similar to iterated best-response models such as the cognitive hierarchy model [9] and level-n theory [2]. In iterated best-response models, players are categorized according to the number of iterative steps of reasoning they are able to perform before making a decision. However, although our approach is similar to iterated best-response models, there are a number of differences.

Iterated best-response models generally deal with player responses in unrepeated simultaneous-choice games. The least sophisticated step 0 or level-0 player is assumed to play according to some known probability distribution. Players of higher levels of iterated reasoning, say level k, form beliefs about the relative distributions of players of levels 0 to $k-1$. This results in an expected distribution of actions, to which players of level k play the best response. In contrast, the approach in this paper is focused on predicting an opponent's future behaviour over repeated games. An agent has no information on the order of theory of mind his opponent is playing at, but attempts to model her behaviour through repeated interaction, and adjust his own behaviour accordingly. However, in our approach the opponent is trying to model the agent as well, and changes the behaviour that the agent is trying to model.

Using agent-based simulations, we determined whether the competitive advantage suggested by the Machiavellian intelligence hypothesis are present in the repeated RPS game, perhaps the most transparent non-trivial setting in which the role of theory of mind can be investigated. Our results from the RPS game show that first-order and second-order theory of mind agents clearly outperform opponents that are more limited in their ability to model others in the game of RPS, while performance of a third-order theory of mind agent playing against a second-order theory of mind opponent is poor in comparison. Qualitatively, our results in the repeated single-shot RPS game are equivalent to those obtained in the dynamic game setting of Limited Bidding [11]. Additionally, in the settings we investigated, the ability to make use of fourth-order theory of mind allowed agents to obtain little to no advantage over third-order theory of mind opponents.

Our results in the ERPS game suggest that the advantage of higher-order theory of mind, and the diminishing returns on higher orders of theory of mind found in RPS are not related to the number of actions available to the agents. When agents choose from five instead of three possible actions, performance of first-order and second-order agents is comparable to their performance then playing RPS, while third-order theory of mind agents only improved when playing against an opponent with a stationary mixed strategy.

We also found that the effectiveness of theory of mind is dependent on the existence of a unique best response. In the game of RPSLS, each opponent action can be defeated by two other actions. As a result, performance of theory of mind agents in the game of RPSLS is greatly reduced compared to performance in the games of RPS and ERPS. One possible explanation for the low performance in the game of RPSLS is that when an agent is indifferent between two actions, he chooses either one with equal probability. It is possible that a slight asymmetry, such that one option is preferable over the other, would benefit higher-order theory of mind. One direction of future research is to determine whether such asymmetries create a focal point [26] for agents with a lower order of theory of mind, resulting in more predictable behaviour.

Our findings provide further support for the Machiavellian intelligence hypothesis. In the competitive games we presented, first-order and second-order theory of mind provide a clear advantage over an opponent of a lower order, while deeper levels of recursion help less, and show that the results reported in [11] are not exclusive to the specific game setting studied there, but are generalizable to competitive games of different designs. In future research, we aim to determine the advantage of the ability to make use of theory of mind in cooperative settings, for example in teamwork, and in mixed-motive settings such as negotiations.

Acknowledgments

This work was supported by the Netherlands Organisation for Scientific Research (NWO) Vici grant NWO 277-80-001.

References

[1] C. Bach and A. Perea. Utility proportional beliefs. In Ghosh et al. [14].
[2] M. Bacharach and D. Stahl. Variable-frame level-n theory. *Game Econ. Behav.*, 32(2):220–246, 2000.
[3] M. Balter. 'Killjoys' challenge claims of clever animals. *Science*, 335(6072):1036–1037, 2012.
[4] K. Binmore. *Playing for Real.* Oxford University Press, Oxford, UK, 2007.
[5] A. Brandenburger and E. Dekel. Hierarchies of beliefs and common knowledge. *J. Econ. Theory*, 59:189–189, 1993.

[6] A. Brandenburger, A. Friedenberg, and H. Keisler. Admissibility in games. *Econometrica*, 76(2):307–352, 2008.
[7] R. Brown. Smoothing, forecasting and prediction of discrete time series. 1963.
[8] C. Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2003.
[9] C. Camerer, T. Ho, and J. Chong. A cognitive hierarchy model of games. *Q. J. Econ.*, 119(3):861–898, 2004.
[10] V. Crawford and N. Iriberri. Fatal attraction: Salience, naïveté, and sophistication in experimental "Hide-and-Seek" games. *Am. Econ. Rev.*, 97(5):1731–1750, 2007.
[11] H. de Weerd and B. Verheij. The advantage of higher-order theory of mind in the game of limited bidding. In J. van Eijck and R. Verbrugge, editors, *Proc. Workshop Reasoning About Other Minds*, CEUR Workshop Proceedings 751, pages 149–164, 2011.
[12] J. Epstein. Agent-based computational models and generative social science. *Complexity*, 4(5):41–60, 1999.
[13] L. Flobbe, R. Verbrugge, P. Hendriks, and I. Krämer. Children's application of theory of mind in reasoning and language. *J. Log. Lang. Inf.*, 17(4):417–442, 2008.
[14] S. Ghosh, J. van Benthem, and R. Verbrugge, editors. *Modeling Strategic Reasoning*. Texts in Logic and Games (FOLLI subseries of LNCS). Springer Verlag, Berlin, to appear.
[15] H. Gintis. *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Behavior*. Princeton University Press, Princeton, NJ, 2000.
[16] P. Gmytrasiewicz and P. Doshi. A framework for sequential planning in multiagent settings. *J. Artif. Intell. Res.*, 24(1):49–79, 2005.
[17] T. Hedden and J. Zhang. What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, 85(1):1–36, 2002.
[18] B. Meijering, H. van Rijn, N. Taatgen, and R. Verbrugge. I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. In *Proc. 33rd Conf. Cogn. Sci. Soc.*, pages 2486–2491. Cognitive Science Society, 2011.
[19] S. Nichols and S. Stich. *Mindreading*. Oxford University Press, USA, 2003.
[20] M. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, Cambridge, MA, 2006.
[21] E. Pacuit. Strategic reasoning in games. In Ghosh et al. [14].
[22] D. Penn and D. Povinelli. On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philos. T. R. Soc. B*, 362(1480):731, 2007.
[23] J. Perner and 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(3):437–71, 1985.
[24] A. Pfeffer. Networks of influence diagrams: A formalism for representing agents' beliefs and decision-making processes. *J. Artif. Intell. Res.*, 33:109–147, 2008.
[25] A. Rapoport and D. Budescu. Randomization in individual choice behavior. *Psychol. Rev.*, 104(3):603–617, 1997.
[26] R. Sugden. A theory of focal points. *Econ. J.*, 105(430):533–550, 1995.
[27] M. Tomasello. *Why we Cooperate*. MIT Press, Cambridge, MA, 2009.
[28] E. van der Vaart, R. Verbrugge, and C. Hemelrijk. Corvid re-caching without theory of mind: A model. *PLoS ONE*, 7(3):e32904, 2012.
[29] H. Van Ditmarsch. Knowledge games. *B. Econ. Res.*, 53(4):249–273, 2001.
[30] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, 2007.
[31] R. Verbrugge. Logic and social cognition: The facts matter, and so do computational models. *J. Philos. Logic*, 38:649–680, 2009.
[32] J. Von Neumann. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, 100(1):295–320, 1928.
[33] A. Whiten and R. Byrne. *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge University Press, Cambridge, 1997.