# Higher-order theory of mind in negotiations under incomplete information

Harmen de Weerd[1], Rineke Verbrugge[1], Bart Verheij[1,2]

[1] Institute of Artificial Intelligence, University of Groningen
[2] CodeX, Stanford University

**Abstract.** Theory of mind refers to the ability to reason explicitly about unobservable mental content such as beliefs, desires, and intentions of others. People are known to make use of theory of mind, and even reason about what other people believe about their beliefs. Although it is unknown why such a higher-order theory of mind evolved in humans, exposure to mixed-motive situations may have facilitated its emergence. In such mixed-motive situations, interacting parties have partially overlapping goals, so that both competition and cooperation play a role. In this paper, we consider negotiation using alternative offers in a particular mixed-motive situation known as Colored Trails, and determine to what extent higher-order theory of mind is beneficial to computational agents. Our results show limited effectiveness of first-order theory of mind, while second-order theory of mind turns out to benefit agents greatly by allowing them to reason about the way they communicate their interests.

## 1 Introduction

In everyday life, people regularly reason about what other people know and believe. People use this *theory of mind* [1] to understand why other people behave in a certain way, to predict their future behaviour, and to distinguish between intentional and accidental behaviour. People also take this ability one step further, and consider that others have a theory of mind as well. This second-order theory of mind allows people to consider and even expect that others will understand why they behave the way that they do. In this paper, we make use of agent-based computational models to explain why our ability to reason about mental content of others may have evolved.

Second-order theory of mind allows people to reason explicitly about belief attributions made by others. For example, in the sentence "Alice knows that Bob knows that Carol is throwing him a surprise party", a *second-order knowledge attribution* is made to Alice, in which she attributes knowledge to Bob. The human ability to make use of higher-order (i.e. at least second-order) theory of mind is well-established, both through tasks that require explicit reasoning about second-order belief attributions [2, 3], as well as in strategic games [4, 5]. However, the use of theory of mind of any kind by non-human species is a controversial matter [6–8]. These differences in the ability to make use of theory of mind raise the issue of the reason for the evolution of a system that allows humans to 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.

A possible explanation for the emergence of higher-order theory of mind is that it is needed in situations that involve mixed-motive interactions such as negotiations [9] or crisis management [10]. In these situations, interactions are partially cooperative in the sense that the interaction can lead to a mutually beneficial outcome, but also partially competitive in the sense that there is no outcome that is optimal for everyone involved. Mixed-motive situations can be understood as the task of sharing a pie [11]. When negotiating parties cooperate to find mutually beneficial solutions, they are searching for ways to enlarge the pie that they are trying to share. At the same time, negotiating parties compete to receive as large a portion of the pie as possible for themselves.

In this paper, we make use of agent-based computational models to investigate the advantages of making use of higher-order theory of mind in mixed-motive settings. Agent-based modeling has proven its usefulness as a research tool to investigate how behavioral patterns emerge from the interactions between individuals (cf. [12]), by allowing precise control and monitoring of the mental content of agents, including application of theory of mind. This approach differs from related work on prescriptive models of negotiation (see for example [13–15]) in that we simulate interactions between agents that differ in their theory of mind abilities to determine the extent to which higher-order theory of mind provides agents with an advantage over those that are more restricted in their use of theory of mind.

We have selected to investigate the effectiveness of higher-order theory of mind in the influential Colored Trails setting, introduced by Grosz, Kraus and colleagues [16–18], which provides a useful test-bed to study interactions in mixed-motive situations. In single-shot negotiations in Colored Trails, first-order theory of mind has been shown to benefit agents greatly, while the advantage of second-order theory of mind only appears when negotiations involve more than two players [19]. In the current paper, we change the Colored Trails setting to include incomplete information about the goals of the partner and to allow for multiple rounds of negotiation, where two agents alternate in making offers until an agreement is reached.

The remainder of the paper is structured as follows. Section 2 describes the details of the Colored Trails setting we investigate, after which Section 3 presents how theory of mind agents negotiate in this setting. To determine the effectiveness of theory of mind, we simulated negotiations between agents of different orders of theory of mind. The results of these can be found in Section 4. Finally, Section 5 provides discussion and gives directions for future research.

## 2 Colored Trails

To determine the effectiveness of higher orders of theory of mind in negotiations, we compare performance of computational agents in the setting of Colored Trails

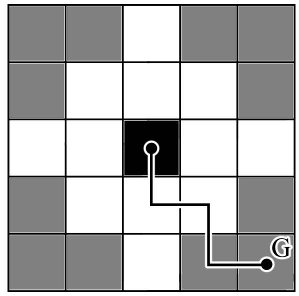



Fig. 1: The Colored Trails game is played on a 5 by 5 board. Players are initially placed on the black tile, and aim to approach their goal tile as closely as possible. To follow the black path from the initial tile to goal location $G$, a player would have to hand in two white chips and two gray chips.

(CT). Colored Trails is a board game designed as a research test-bed for investigating decision-making in groups of people and computer agents [17, 18]. Colored Trails is played by two or more players on a board of colored tiles. Each player starts the game at a given initial tile with a set of colored chips. The colors of the chips match those of the tiles of the board. A player can move to a tile adjacent to his current location by handing in a chip of the same color as the destination tile. Each player is also assigned a goal location, which the player has to approach as closely as possible. To achieve this goal, players are allowed to trade chips among each other. The Colored Trails setting represents a multi-issue bargaining situation, where each issue is represented by a color, while different paths towards the goal location represent different acceptable solutions.

We follow the scoring rules in [17] and award a player that reaches his goal tile 500 points. If a player is unable to reach the goal tile, he pays a penalty of 100 points for each tile in the shortest path from his current location to his goal location. Chips that have not been used to move on the board are worth 50 points. For each player, reaching the goal location is therefore the most valuable. Since a player generally needs a different set of chips to achieve his goal than his trading partner, there may be an opportunity for a trade that would allow both players to reach a higher score. But since unused chips increase a player's score as well, players compete to own as many chips as possible.

In our setting, the Colored Trails game is played by two players on a 5 by 5 board such as the one depicted in Figure 1. Both players start at the center of the board, indicated by the black square. This initial location is publicly announced to the players, so that each player knows the initial location of the other player. Each player also knows his own goal location, which has been randomly chosen from the 12 possible tiles that are at least three steps away from the goal location, indicated by the gray tiles in Figure 1. Although players know their own goal location, they do not know the goal location of their trading partner. This ensures that at the start of a game, players are uncertain about the bargaining position of their trading partner. That is, they do not know the score of their trading partner if negotiations should fail, or his preferences over chips. Through the use of theory of mind, agents can extract information from the offers made by the trading partner to try to learn his goal location.

Negotiation between players takes the form of a sequence of offers. Players take turns suggesting a redistribution of chips, which their trading partner can choose to accept or counter with an offer of his own. The game ends as soon as an offer is accepted. Alternatively, when a player believes that it is impossible to reach an agreement, he can end the negotiation and the initial distribution of chips becomes final. Players can make any offer they wish. For example, a player may repeat an offer that has been previously rejected by his trading partner, or make an offer that he himself has previously rejected. However, both players pay a 1 point penalty for each round of play. That is, when negotiations end after five offers, the final score of each player is reduced by five points.

In Colored Trails, players can achieve a higher score by trading chips in such a way that both players can move closer to their respective goal locations, thereby enlarging the pie they share. At the same time, players compete to obtain as large a piece of the pie as possible through trades that will increase their own score more than it will increase the score of their trading partner.

## 3 Theory of mind agents in Colored Trails

To investigate the effectiveness of theory of mind in mixed-motive settings, we constructed theory of mind agents that are able to play the game as outlined in Section 2. These agents are inspired by the theory of mind agents used by [20] to investigate the effectiveness of theory of mind in competitive settings. To focus our investigations to the effectiveness of theory of mind, we assume that agents make no mistakes in finding routes between locations and do not consider the possibility that mistakes could be made in finding these routes. In the following subsections, we describe how agents of different orders of theory of mind negotiate with their trading partner in the Colored Trails setting. [3]

### 3.1 Zero-order theory of mind agent

The zero-order theory of mind ($ToM_0$) agent is unable to attribute mental content to others. Instead, the $ToM_0$ agent forms zero-order beliefs about the likelihood of his trading partner accepting a certain offer. The $ToM_0$ agent uses these beliefs to calculate the expected value of making an offer, which takes into account the change in the score of the $ToM_0$ agent if his trading partner should accept his offer, as well as the cost of another round of negotiation. A negative expected value for some offer $O$ means that the $ToM_0$ agent believes that it would be better to withdraw from negotiations rather than to make offer $O$.

[3] The formal model of theory of mind agents is available as an online appendix at http://www.ai.rug.nl/SocialCognition/Experiments/.

The $ToM_0$ agent bases his zero-order beliefs on his observations of the behaviour of his trading partner. For example, if the trading partner rejects an offer made by the $ToM_0$ agent, the $ToM_0$ agent believes that his trading partner will also reject any offer that assigns fewer chips to his trading partner and more to himself. Similarly, when the trading partner makes an offer that assigns many red chips to the trading partner, the $ToM_0$ agent concludes that it is unlikely that his trading partner is willing to accept an offer that assigns few red chips to the trading partner and adjusts his beliefs accordingly.

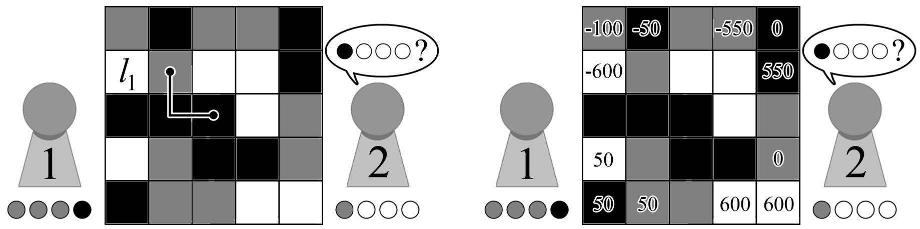
The degree to which the $ToM_0$ agent adjusts his beliefs based on the observed behaviour of his trading partner is represented by a learning speed parameter $\lambda$ ($0 \leq \lambda \leq 1$). A $ToM_0$ agent with zero learning speed does not adjust his beliefs over the course of negotiation. Such a $ToM_0$ agent rarely withdraws from negotiation, and keeps making the same offer until his trading partner either accepts or makes an acceptable counteroffer. Such a counteroffer made by the trading partner should increase of the score of the $ToM_0$ agent that is at least as much as the expected value the $ToM_0$ agent assigns to his own offer.

If the $ToM_0$ agent has the maximum learning speed ($\lambda = 1$), he radically changes his beliefs based on the behaviour of his trading partner. For example, when his trading partner rejects an offer made by the $ToM_0$ agent, the $ToM_0$ agent considers it impossible that his trading partner would accept this offer at any future point in the negotiations. Similarly, when the trading partner makes an offer that assigns three red chips to the trading partner, a $ToM_0$ agent with learning speed $\lambda = 1$ believes that his trading partner will not accept any offer that assigns two or fewer red chips to the trading partner. With a learning speed $\lambda = 1$, a $ToM_1$ agent is therefore quick to end negotiations, either by accepting an offer made by his trading partner or by withdrawing from negotiation.

*Example 1.* Suppose two agents play the Colored Trails game as shown in Figure 2a, in which agent 1 is a $ToM_0$ agent who wants to move from the central square to the white square marked $l_1$. Each agent has an initial set of chips, which holds 3 gray chips and 1 black chip for agent 1, while agent 2 has 3 white chips and 1 gray chip. Figure 2a also shows that with the initial distribution, agent 1 can move two tiles towards his goal location, which results in a score of 0.

Suppose that agent 2 makes an offer to trade the black chip held by agent 1 for the gray chip held by agent 2. Since this offer assigns all black and white chips to agent 2, agent 1 decreases his belief that agent 2 will accept an offer in which agent 1 receives any of the black or white chips. After this belief update, agent 1 decides how to respond. Since accepting the offer would decrease the score of agent 1, agent 1 would rather withdraw from negotiation than accept the offer. However, agent 1 may still consider to make a counteroffer.

Agent 1 can reach his goal either with 1 gray chip, 1 black chip, and 1 white chip, or alternatively with 2 white chips and 1 gray chip. Both options would increase the score of agent 1 by 500. Agent 1 could also ask for more chips than he needs to reach his goal location. Each additional unused chip would increase the score of agent 1 by 50, but also decreases his belief that agent 2 will accept




(a) (b)

Fig. 2: Example of a negotiation setting in Colored Trails. Agent 1 wants to move from the central square to his goal location $l_1$. With his initial set of chips, agent 1 can move two tiles towards his goal location, as shown by the path in (a). When agent 1 is a $ToM_1$ agent, he tries to determine the goal location of agent 2 by calculating how accepting the offer an offer from agent 2 would change the score of agent 2, for all possible goal locations.

the offer. Depending on his exact beliefs, agent 1 decides whether there is any counteroffer that is worth risking the cost of rejection.

### 3.2 First-order theory of mind agent

The first-order theory of mind ($ToM_1$) agent considers the possibility that his trading partner has beliefs and goals, which determine whether or not his trading partner will accept an offer. The $ToM_1$ agent is able to consider his own offer from the perspective of his trading partner, and decide whether he would accept that offer if he were in the position of his trading partner. However, the $ToM_1$ agent does not know the actual goal location and zero-order beliefs of his trading partner, but forms beliefs about his trading partner's goal location and beliefs.

Unlike settings that have complete information like rock-paper-scissors [20], offers in Colored Trails can reveal information about a player's preferences. Using his first-order theory of mind, the $ToM_1$ agent can extract information about his trading partner's goal location from the offers he receives. By putting himself in the position of his trading partner, the $ToM_1$ agent knows that his trading partner would not make an offer that, if the $ToM_1$ agent were to accept it, would reduce the score of the trading partner. The $ToM_1$ agent therefore believes that locations for which this is the case cannot be the goal location of his trading partner. For other locations, the $ToM_1$ agent determines what offer he would have made if he were in the position of his trading partner. Those locations for which the $ToM_1$ agent would have made an offer similar to the one he received from his trading partner are assigned the highest likelihood.

To decide on the best offer to make, the $ToM_1$ agent explicitly considers how his trading partner will respond. This allows the $ToM_1$ agent to subtly manipulate his trading partner. While deciding what offer to make to his trading partner, the $ToM_1$ agent determines what counteroffer he believes his trading partner will be making, and whether he would be willing to accept that counteroffer. The $ToM_1$ agent may therefore decide to make an offer $O$ that he believes will be rejected by his trading partner, but which he also expects to change the beliefs of the trading partner in such a way that the trading partner makes counteroffer $O'$, which the $ToM_1$ agent is willing to accept. However, to accurately predict the behaviour of the trading partner, the $ToM_1$ agent needs to know the goal location of the trading partner. Thus, first-order theory of mind has limited use for making the opening bid of a negotiation.

Although the $ToM_1$ agent forms explicit beliefs about goal location and beliefs of his trading partner, the $ToM_1$ agent makes no attempt to model the learning speed $\lambda$ of his trading partner. Instead, the $ToM_1$ agent makes use of his own learning speed when he updates the beliefs he attributes to his trading partner, implicitly assuming that every agent has the same learning speed. This means that a $ToM_1$ agent with zero learning speed $\lambda = 0$ believes that the beliefs of his trading partner do not change in response to any offer he makes, while a $ToM_1$ agent with maximal learning speed $\lambda = 1$ believes that his trading partner will quickly withdraw from negotiations if he makes his trading partner offers that are too demanding. This implies that the $ToM_1$ agent will have an incorrect representation of the beliefs of his trading partner unless the $ToM_1$ agent and his trading partner have the same learning speed.

*Example 2.* We consider the game shown in Figure 2a, in which agent 1 is a $ToM_1$ agent. When agent 2 offers to trade the black chip held by agent 1 for the gray chip held by agent 2, agent 1 uses this to extract information about the goal location of agent 2. For each of the possible goal locations, Figure 2b shows the change in the score of agent 2 if agent 1 were to accept the offer. For example, if the goal location of agent 2 is the white square at the bottom right of the board, accepting the offer would increase the score of agent 2 by 600 points.

The $ToM_1$ agent 1 believes that the goal location of agent 2 cannot be any of the locations that have a negative or zero change in score, since in these cases, agent 2 would have been better off withdrawing from negotiation. Furthermore, by considering the game from the perspective of his trading partner, agent 1 believes that if the goal location of agent 2 had been any of the locations that show an increase by 50 points, agent 2 would have made a different offer. Agent 1 concludes that the offer made by agent 2 is most consistent with his goal location being one of the three remaining locations. If agent 1 were to accept the offer, agent 2 would be able to reach any one of these three locations.

After updating his goal location beliefs, $ToM_1$ agent 1 decides whether or not to make a counteroffer. Using his first-order theory of mind, agent 1 knows that agent 2 will only accept an offer that would increase his score. For two of the three most likely goal locations, agent 2 would be able to reach his goal location with 1 gray, 1 black, and 1 white chip, while agent 1 would be able to reach his goal location by using 2 white chips and 1 gray chip. The final possible goal location also allows both agents to reach their goal location, with agent 2 using 2 white chips and 2 gray chips, while agent 1 uses 1 black, 1 gray, and 1 white chip. However, there is no offer that is guaranteed to allow both agents to reach respective their goal locations given the information about the goal location of agent 2. The final decision depends on the beliefs of agent 1.

### 3.3 Higher orders of theory of mind agent

Agents that are able to use orders of theory of mind beyond the first can use this ability to attempt to manipulate the beliefs of lower orders of theory of mind to obtain an advantage. For example, a second-order theory of mind ($ToM_2$) agent models his trading partner as a $ToM_1$ agent, which means that the $ToM_2$ agent believes that his trading partner may be interpreting the offers he makes to find out what his goal location is. This allows the $ToM_2$ agent to construct an offer which will inform his trading partner about his goal location. This could speed up the process of finding a mutually beneficial offer.

Second-order theory of mind also allows the $ToM_2$ agent to deceive and manipulate his trading partner more effectively. By careful construction of the offers he makes, the $ToM_2$ agent can provide his trading partner with incomplete or ambiguous information about his goal location, or induce a false belief in his trading partner concerning his goal location. This may cause a $ToM_1$ trading partner to make an offer that is more generous towards the $ToM_2$ agent than the trading partner believes it to be.

For each additional order of theory of mind, agents also take an additional round of play into consideration. Where a $ToM_0$ agent only judges whether his trading partner is likely to accept an offer $O$, a $ToM_1$ agent believes that the way a $ToM_0$ trading partner reacts to an offer $O$ depends on how likely the trading partner thinks it is that the $ToM_1$ agent will accept a possible alternative offer $O'$. A $ToM_2$ agent considers an additional round of play by realizing that the reaction of a $ToM_1$ trading partner to an offer $O$ depends on what this trading partner believes to be the reaction of a $ToM_0$ agent to a possible alternative offer $O'$, which in turn depends on how likely a $ToM_0$ agent considers it to be that his trading partner will accept a possible alternative offer $O''$.

*Example 3.* We consider the game shown in Figure 2a. Example 2 showed that if agent 1 is a $ToM_1$ agent, he believes that there are three possible goal locations for agent 2, and has to make a decision of what offer to make under this uncertainty of goal location. If agent 1 is a $ToM_2$ agent, he can decide to make an offer that is unlikely to be accepted, but may provide agent 2 with enough information to construct an offer that is acceptable to both agents. For example, $ToM_2$ agent 1 believes that if his offer assigns the black chip to agent 2, this provides agent 2 with information about his goal location. Agent 1 can make an offer that assigned 2 gray and 2 white chips to himself to signal that he does not assign a high value to the black chip. However, since the offer of agent 2 assigns the black chip to agent 2, this chip may be of high value for agent 2. By constructing an offer that assigns the black chip to himself, agent 1 can attempt to make agent 2 believe that the black chip is valuable to agent 1 as well. This may encourage agent 2 to give up more chips in exchange for the black chip.

## 4 Simulation results

We performed simulations where the theory of mind agents described in Section 3 negotiated in the Colored Trails setting described in Section 2 according to an alternating offers protocol. Games were played by two agents on a 5 by 5 board of tiles, randomly colored with one of five possible colors. At the start of the game, each player received an initial set of four randomly colored chips, drawn from the same colors as those on the board. Since each player needs at most four chips to reach his or her goal location, it is sometimes possible that after a trade, both players can reach their respective goal location. However, this was not always the case. To ensure that both players have an incentive to negotiate to increase their score, game settings in which some player could reach his goal location with the initial set of chips without trading were excluded from analysis.

To determine the effectiveness of theory of mind, pairs of agents played 1,000 consecutive Colored Trails games. Although agents alternated in making offers, the literature suggests that the opening bid is influential [11, 21], because it serves as an anchor for the negotiation process. Because of this, we differentiate between *initiators*, who make the first offer in a game, and *responders*.

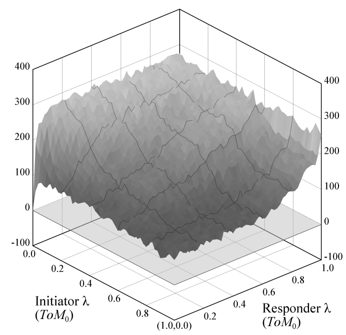
In our simulations, we determined the negotiation score of a $ToM_i$ initiator playing Colored Trails with a $ToM_j$ responder, for each combination of $i, j = 0, 1, 2$. The negotiation score is calculated as the average difference between the initiator's final score after negotiation ended and his initial score at the start of negotiation. A negative score therefore indicates that the initiator paid a higher cost for negotiation than he gained from the resulting trade. Agents started every game reasoning at their highest theory of mind ability. That is, a $ToM_2$ agent always started the game by taking into account the beliefs his trading partner might have about his own beliefs. Although negotiations could theoretically take infinitely long, games that continued for more than 100 rounds of offers made were considered to be unsuccessful. In this case, the initial situation became final, and both agents incurred the cost of 100 rounds of play. In our model, agents were unable to reason about this limit.

In the following subsections, we present competitive and cooperative aspects of negotiation in Colored Trails separately. In Section 4.1, we present the individual performance of agents, which shows how well agents compete. Section 4.2 focuses on the cooperative element of negotiation, and describes the effect of theory of mind on the combined score of the agents in the Colored Trails setting.

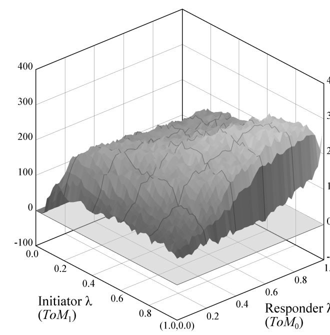
### 4.1 Individual performance results

In this section, we describe the individual performance of theory of mind agents when negotiating in Colored Trails. How large a piece of pie the agents end up with shows how theory of mind influences the competitive abilities of agents.

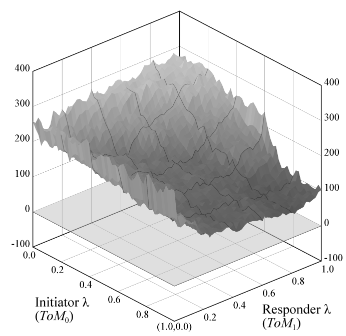
Figure 3a shows the average negotiation score of a $ToM_0$ initiator playing Colored Trails with a $ToM_0$ responder. The three-dimensional figure shows the negotiation score as a function of the learning speeds of both initiator and responder. Since most points in the figure are above the semi-transparent plane of



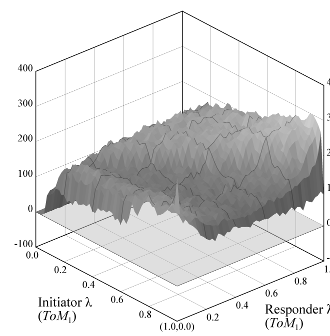

(a) $ToM_0$ initiator, $ToM_0$ responder



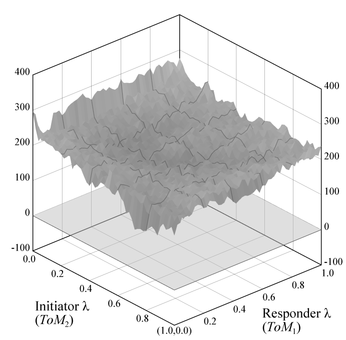

(b) $ToM_1$ initiator, $ToM_0$ responder



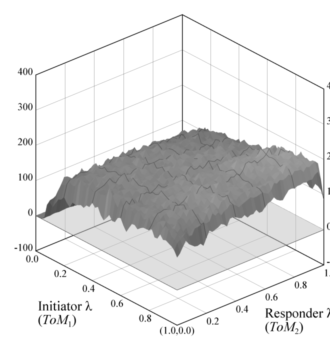

(c) $ToM_0$ initiator, $ToM_1$ responder




(d) $ToM_1$ initiator, $ToM_1$ responder




(e) $ToM_2$ initiator, $ToM_1$ responder




(f) $ToM_1$ initiator, $ToM_2$ responder

Fig. 3: Average negotiation score of theory of mind agents playing Colored Trails.

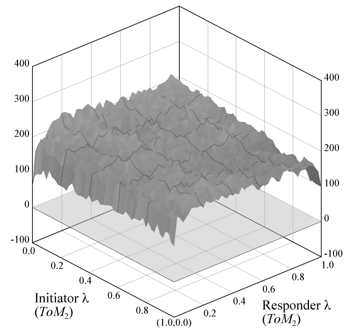



Fig. 4: Average negotiation score of a $ToM_2$ agent negotiating with a $ToM_2$ trading partner in Colored Trails.

zero negotiation score, the figure shows that $ToM_0$ agents are often able to increase their score through negotiation despite their inability to reason explicitly about the goals and desires of their trading partner.

Figure 3a shows that the negotiation score of the $ToM_0$ initiator generally increases as his own learning speed decreases. An agent with a low learning speed changes his behaviour little in response to the offers his trading partner makes, and behaves as if he is unwilling to make concessions. An agent with zero learning speed ($\lambda = 0$) does not adjust his behaviour at all, but keeps repeating the same offer. When both agents follow this strategy, and neither is willing to accept the offer of ther trading partner, the agents will be unable to reach an agreement and only carry the burden of a failed negotiation. Summing up, there is an evolutionary pressure on $ToM_0$ agents to decrease their learning speed, which may result in the worst possible outcome in which negotiation fails.

Figure 3b shows that the score of a $ToM_1$ initiator negotiating with a $ToM_0$ responder is not always higher than that of a $ToM_0$ initiator in the same position. Although the $ToM_1$ obtains a higher score when both agents have a high learning speed, the $ToM_1$ initiator performs poorly when his learning speed is low. Moreover, a $ToM_1$ initiator with learning speed $\lambda = 0$ withdraws from negotiation before making his initial offer. The reason for this is that the $ToM_1$ agent attributes his own learning speed to his trading partner. A $ToM_1$ agent with zero learning speed predicts that his trading partner will keep repeating the same offer until the $ToM_1$ agent makes an acceptable offer. As a result, the $ToM_1$ agent believes that the only way to successfully complete negotiations is for him to give his trading partner what he wants. At the start of a game, however, the $ToM_1$ initiator does not know the goal location of the responder. As a result, there is no offer that the $ToM_1$ initiator believes to be successful, and the $ToM_1$ initiator chooses not to engage in negotiation.

Although the $ToM_1$ agent has an incentive to keep his learning speed high, Figure 3c shows that even in the presence of a $ToM_1$ responder, the $ToM_0$ initiator performs best when his learning speed is close to zero. Since a $ToM_1$ agent will adjust his offers to take into account the score of his trading partner, the trading partner has no incentive to make any concessions. That is, even though the presence of a $ToM_1$ agent prevents negotiation from failing, the $ToM_0$ agent benefits more from this outcome than the $ToM_1$ agent.

Figure 3d shows the negotiation score of a $ToM_1$ initiator playing Colored Trails with a $ToM_1$ responder. The figure shows a ridge along the line of equal learning speeds, where the agent with the lower learning speed generally obtains the higher score. A $ToM_1$ agent with a high learning speed attributes this learning speed to his trading partner and expects that his offers strongly influence the behaviour of his trading partner. Such a $ToM_1$ agent believes that his trading partner is quick to conclude that a negotiation will be unsuccessful. To prevent his trading partner from withdrawing from negotiations, the $ToM_1$ agent makes offers that benefit his trading partner more at the expense of his own score.

However, unlike $ToM_0$ agents, $ToM_1$ agents suffer when their learning speed becomes too low. When a $ToM_1$ agent has a learning speed below $\lambda = 0.2$, his performance does not increase as his learning speed goes down. Because of this, the evolutionary pressure on $ToM_1$ agents to have low learning speeds does not lead to a situation in which negotiation fails as it does for $ToM_0$ agents.

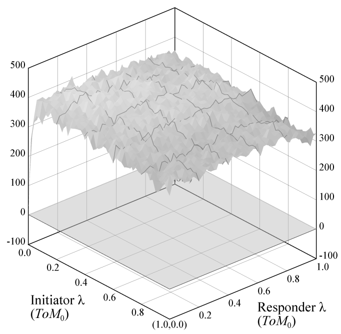
Figure 3e and Figure 3e show the performance of a $ToM_1$ and a $ToM_2$ agent negotiating in the Colored Trails setting. The figures show that the $ToM_2$ initiator is more effective in obtaining a large piece of the pie when negotiating with a $ToM_1$ responder than vice versa, irrespective of the learning speeds of the agents. Compared to results of lower orders of theory of mind, Figure 3e also shows a fairly flat surface, indicating that the score of the $ToM_2$ initiator is less dependent on learning speeds of the initiator and responder.

Figure 4 shows that when a $ToM_2$ initiator negotiates with a $ToM_2$ responder, his score is not influenced greatly by the learning speeds of either agent. Nonetheless, the $ToM_2$ initiator performs best when his learning speed is low, but greater than zero.

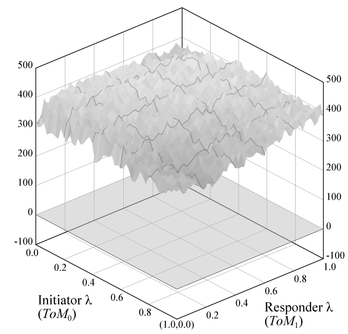
The results in this section show that although a $ToM_1$ agent succeeds in securing a larger pie when negotiating with a $ToM_0$ trading partner than a $ToM_0$ agent, the $ToM_1$ agent only obtains a small piece of this pie. The $ToM_2$ agent negotiates successfully with other theory of mind agents as well, but also ensures that he receives a large piece of pie for himself. In the next subsection, we take a closer look at the cooperative abilities of these theory of mind agents.

### 4.2 Social welfare results

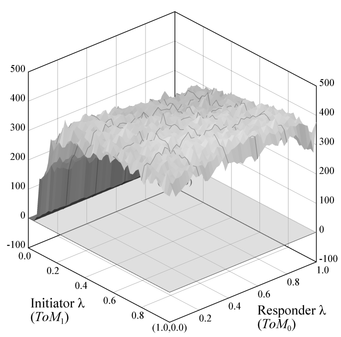
In the previous section, we compared the individual competitive performance of agents of several orders of theory of mind negotiating in Colored Trails. In this section, we show how theory of mind affects the cooperative ability of agents. To this end, we take a closer look at the increase in social welfare that theory of mind agents achieve, where social welfare is measured by the sum of the scores of the initiator and the responder. Figure 5 and Figure 6 show the increase in social welfare for different combinations of theory of mind initiators and responders.



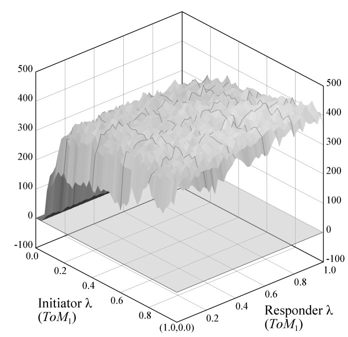

(a) $ToM_0$ agent, $ToM_0$ partner



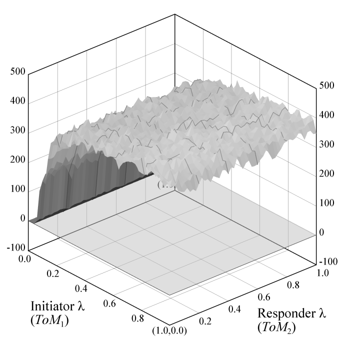

(b) $ToM_0$ agent, $ToM_1$ partner



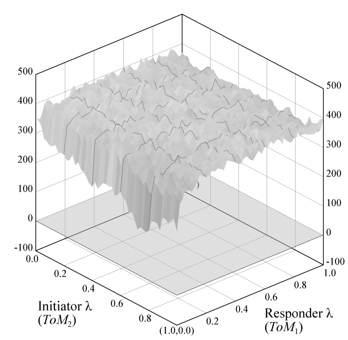

(c) $ToM_1$ agent, $ToM_0$ partner




(d) $ToM_1$ agent, $ToM_1$ partner




(e) $ToM_1$ agent, $ToM_2$ partner




(f) $ToM_2$ agent, $ToM_1$ partner

Fig. 5: Average combined negotiation score of theory of mind agents playing Colored Trails.

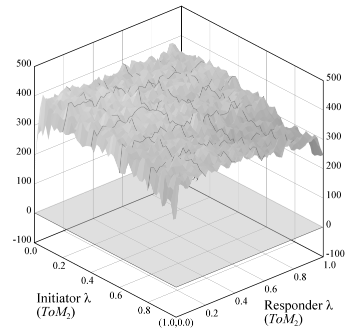



Fig. 6: Average combined negotiation score of two $ToM_2$ agents negotiating.

Figure 5a shows that $ToM_0$ agents can cooperate surprisingly well. However, cooperation between $ToM_0$ agents is not stable due to the competitive element of Colored Trails. The $ToM_0$ agents experience an evolutionary pressure towards zero learning speed, which can eventually lead to negotiation failure.

Although Section 4.1 shows that the presence of a $ToM_1$ agent can stabilize cooperation in Colored Trails, Figure 5 shows that this does not imply higher social welfare. Figures 5b and 5c do not show an improvement over the performance of $ToM_0$ agents shown in Figure 5a. When two $ToM_1$ agents play Colored Trails together, they achieve the highest social welfare when both agents have the maximum learning speed $\lambda = 1$. However, the competitive element in Colored Trails puts an evolutionary pressure on $ToM_1$ agents to lower their learning speed. Although this does not lead to a breakdown of negotiation as it does for $ToM_0$ agents, social welfare suffers from the lower learning speed of $ToM_1$ agents. That is, the individual desire of $ToM_1$ agents to obtain as large a piece of pie as possible results in a smaller pie to share.

In contrast, Figures 5e and 5f show that the ability of a $ToM_2$ agent and a $ToM_1$ agent to achieve a high social welfare depends mostly on the learning speed of the $ToM_1$ agent. Interestingly, the learning speed of the $ToM_1$ agent that results in a higher social welfare also yields the $ToM_1$ agent the best individual score. That is, the learning speed that would yield the $ToM_1$ agent the largest piece of pie when negotiating with a $ToM_2$ agent also yields the largest total pie.

Figure 6 shows a similar effect when two $ToM_2$ agent negotiate in Colored Trails. The maximum social welfare is achieved when both agents have a low non-zero learning speed, which also yields them individually the highest score.

## 5 Discussion

We have used agent-based models to show how theory of mind can present individuals with an advantage over others that lack such an ability in the negotiation setting Colored Trails. Our results show how theory of mind can indeed present individuals with evolutionary advantages, by facilitating successful negotiation with others. Although agents without a theory of mind are very successful under optimal circumstances, zero-order theory of mind agents face a cooperative dilemma where agents can defect on cooperation by not making any concessions to their trading partner. When this strategy is adopted by all agents, these agents are unable to negotiate successfully. In terms of sharing a pie, the individual goals of agents to receive as large a piece of pie as possible increases competition until there no longer is any pie to share.

Agents negotiate more successfully when the ability to make use of theory of mind is introduced. By reasoning explicitly about the goals of their trading partner, first-order theory of mind agents can recognize the need for a mutually beneficial outcome. However, this does not solve the cooperative dilemma completely. The attempts of first-order theory of mind agents to obtain as large a piece as possible comes at the expense of the size of the pie as a whole.

Although first-order theory of mind has a limited effectiveness in the negotiation setting we describe, second-order theory of mind greatly benefits agents. When a second-order theory of mind agent negotiates with another agent capable of theory of mind, neither agent has an incentive to deviate from the outcome that maximizes social welfare. That is, agents that succeed in negotiating the largest possible pie could not have received a larger piece of pie for themselves by changing their behaviour. Moreover, second-order theory of mind also allows an agent to obtain a larger piece of the pie for himself. In future research, we aim to determine whether the extent to which theory of mind is effective in mixed-motive settings goes beyond second-order theory of mind.

The setting we investigate is similar to [16] in that we model bounded rational agents negotiating with incomplete information. However, the second-order theory of mind agents that we model also take into account that their trading partner has incomplete information concerning their own goals as well. In future work, these agents of [16] can be compared directly to theory of mind agents.

The effectiveness of theory of mind in our setup can be understood in terms of interest-based bargaining [11, 22]. Without theory of mind, agents can only negotiate in terms of positions, by making offers without regard for the interests of others. First-order theory of mind allows an agent to identify the interests of his trading partner, while a second-order theory of mind agent is able to communicate his interests through his choice of offers. That is, theory of mind allows agents to engage in interest-based bargaining, where agents reveal their interests to uncover mutually beneficial solutions.

**Acknowledgments**

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

## References

1. Premack, D., Woodruff, G.: Does the chimpanzee have a theory of mind? Behav. Brain Sci. **1**(4) (1978) 515–526

2. Apperly, I.: Mindreaders: The Cognitive Basis of “Theory of Mind”. Psychology Press, Hove, UK (2011)
3. Perner, J., Wimmer, H.: “John thinks that Mary thinks that...”. Attribution of second-order beliefs by 5 to 10 year old children. J. Exp. Child Psychol. **39**(3) (1985) 437–471
4. Hedden, T., Zhang, J.: What do you think I think you think?: Strategic reasoning in matrix games. Cognition **85**(1) (2002) 1–36
5. Meijering, B., van Rijn, H., Taatgen, N.A., Verbrugge, R.: I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. In: CogSci, Cognitive Science Society (2011) 2486–2491
6. Penn, D.C., Povinelli, D.J.: On the lack of evidence that non-human animals possess anything remotely resembling a ‘theory of mind’. Philos. T. R. Soc. B **362**(1480) (2007) 731–744
7. Tomasello, M.: Why we Cooperate. MIT Press, Cambridge, MA (2009)
8. van der Vaart, E., Verbrugge, R., Hemelrijk, C.K.: Corvid re-caching without ‘theory of mind’: A model. PLoS ONE **7**(3) (2012) e32904
9. Verbrugge, R.: Logic and social cognition: The facts matter, and so do computational models. J. Philos. Logic **38** (2009) 649–680
10. van Santen, W., Jonker, C., Wijngaards, N.: Crisis decision making through a shared integrative negotiation mental model. Int. J. Emerg. Manag. **6**(3) (2009) 342–355
11. Raiffa, H., Richardson, J., Metcalfe, D.: Negotiation Analysis: The Science and Art of Collaborative Decision Making. Belknap Press (2002)
12. Epstein, J.M.: Generative Social Science: Studies in Agent-based Computational Modeling. Princeton University Press, Princeton (NJ) (2006)
13. Endriss, U., Pacuit, E.: Modal logics of negotiation and preference. In Fisher, M., Hoek, W., Konev, B., Lisitsa, A., eds.: Logics in Artif. Intell. Volume 4160 of Lect. Notes Comp. Science. Springer, Berlin (2006) 138–150
14. Wooldridge, M., Parsons, S.: On the use of logic in negotiation. In: Proc. Workshop Agent Commun. Lang. (2000)
15. Fabregues, A., Sierra, C.: Hana: a human-aware negotiation architecture. Decision Support Systems (2013)
16. Lin, R., Kraus, S., Wilkenfeld, J., Barry, J.: Negotiating with bounded rational agents in environments with incomplete information using an automated agent. Artif. Intell. **172**(6) (2008) 823–851
17. Gal, Y., Grosz, B., Kraus, S., Pfeffer, A., Shieber, S.: Agent decision-making in open mixed networks. Artif. Intell. **174**(18) (2010) 1460–1480
18. van Wissen, A., Gal, Y., Kamphorst, B., Dignum, M.V.: Human–agent teamwork in dynamic environments. Computers Human Behav. **28** (2012) 23–33
19. de Weerd, H., Verbrugge, R., Verheij, B.: Agent-based models for higher-order theory of mind. In: ESSA. (2013) In press.
20. de Weerd, H., Verbrugge, R., Verheij, B.: How much does it help to know what she knows you know? An agent-based simulation study. Artif. Intell. **199–200** (2013) 67–92
21. Van Poucke, D., Buelens, M.: Predicting the outcome of a two-party price negotiation: Contribution of reservation price, aspiration price and opening offer. J. Econ. Psychol. **23**(1) (2002) 67–76
22. Rahwan, I., Sonenberg, L., Dignum, F.P.M.: On interest-based negotiation. In Dignum, F., ed.: Adv. Agent Commun. Lect. Notes Artif. Intell. Springer, Berlin (2004) 383–401