If you know what I mean

de Weerd, Hermanes Albertus

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
de Weerd, H. A. (2015). If you know what I mean: Agent-based models for understanding the function of higher-order theory of mind [Groningen]: University of Groningen

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Negotiating with other minds

The role of recursive theory of mind in negotiation with incomplete information

Abstract

Theory of mind refers to the ability to reason explicitly about unobservable mental content of others, such as beliefs, goals, and intentions. People often use this ability to understand the behavior of others as well as to predict future behavior. People even take this ability a step further, and use higher-order theory of mind by reasoning about the way others make use of theory of mind and in turn attribute mental states to different agents. One of the possible explanations for the emergence of the cognitively demanding ability of higher-order theory of mind suggests that it is needed to deal with mixed-motive situations. Such mixed-motive situations involve partially overlapping goals, so that both cooperation and competition play a role. In this paper, we consider a particular mixed-motive situation known as Colored Trails, in which computational agents negotiate using alternating offers with incomplete information about the preferences of their trading partner. In this setting, we 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 can communicate their interests.

This chapter has been submitted for review at a journal. Parts of this chapter have previously been published in *PRIMA 2013: Principles and Practice of Multi-Agent Systems*. 

147
6.1 Introduction

In social settings, people often make predictions of the behavior of others by making use of their \textit{theory of mind}; they reason about unobservable mental content such as beliefs, desires, and intentions of others. Without this theory of mind, an individual is limited to reasoning only about behavior, such as in the sentence “Mary is looking in the drawer”. Such individuals are said to have a \textit{zero-order theory of mind}. \textit{First-order theory of mind} allows agents to reason about unobservable mental content of others as well, and understand sentence like “Mary is looking in the drawer because she \textit{believes} that there is chocolate in the drawer”. People are also capable of taking this theory of mind ability a step further, and reason about way others are using theory of mind. Using \textit{second-order theory of mind}, people understand sentences such as “Alice \textit{believes} that Bob \textit{knows} that Carol is throwing him a surprise party”, and reason about the way Alice is reasoning about Bob’s knowledge.

Behavioral experiments have demonstrated the human ability to make use of higher-order (i.e. at least second-order) theory of mind, both through tasks that require explicit reasoning about second-order belief attributions (Apperly, 2010; Perner & Wimmer, 1985; Wimmer & Perner, 1983; Miller, 2009; Arslan, Hohenberger, & Verbrugge, 2012), as well as through strategic games (Hedden & Zhang, 2002; Meijering et al., 2011; Qu et al., 2012; Goodie et al., 2012; Zhang et al., 2012; De Weerd, Broers, & Verbrugge, 2015). In contrast, the extent to which non-human species are able to use theory of mind of any kind is under debate. Although many non-human species have been proposed to be able to reason about the mental content of others (Tomasello, 2009; Kaminski et al., 2008; Schmelz et al., 2011; Burkart & Heschl, 2007; Kaminski et al., 2006; Clayton et al., 2007; Kaminski, Bräuer, Call, & Tomasello, 2009), the results of theory of mind experiments in these species are criticized for being inconclusive (Penn & Povinelli, 2007; Carruthers, 2008; Van der Vaart et al., 2012; Van der Vaart & Hemelrijk, 2014). Nonetheless, even primates such as chimpanzees fail to show convincingly that they understand that others may hold false beliefs (Call & Tomasello, 1999; Kaminski et al., 2008; Krachun et al., 2009, 2010), which suggests that they do not have full use of first-order theory of mind.

The difference in theory of mind abilities between humans and other animals raises the issue on why humans can make recursive use of theory of mind, while other animals do not appear to have this ability. Furthermore, although humans can make use of higher orders of theory of mind, theory of mind is only used up to a certain point (Flobbe et al., 2008; Hedden & Zhang, 2002; Verbrugge, 2009). This suggests that there may be settings in which the ability to make use of higher orders of theory of mind presents individuals with an evolutionary advantage that explains why humans make use of higher-order theory of mind despite the high
6.1. Introduction

cognitive demands of such a system.

According to the Machiavellian intelligence hypothesis (Byrne & Whiten, 1988; Whiten & Byrne, 1997), the emergence of social cognition, which includes theory of mind, can be explained through a competitive advantage. According to this theory, social cognition allows an individual to deceive and manipulate others more effectively. If increasingly higher orders of theory of mind allow an individual to predict the behavior of his opponents more accurately, individuals that make use of higher-order theory of mind are at an evolutionary advantage over others without such abilities.

Our earlier research using agent-based models has shown that there are indeed competitive settings in which individuals benefit greatly from both first-order and second-order theory of mind, while the additional advantage for reasoning at even higher orders of theory of mind is limited (De Weerd et al., 2013b). However, behavior that appears to be deceptive may not always be an indicator for the use of theory of mind. For example, it has been suggested that deceptive behavior reported in primates (Byrne & Whiten, 1992) and corvids (Bugnyar & Kotrschal, 2002) may be a result of associative learning (Penn & Povinelli, 2007) or factors such as stress (Van der Vaart et al., 2012) rather than reasoning about the minds of others.

Alternatively, the Vygotskian intelligence hypothesis (Vygotsky, 1978; Moll & Tomasello, 2007) suggests that the emergence of theory of mind can be explained through social cooperation. The use of higher-order theory of mind would allow individuals to form shared intentionality. This implies that each of the participants in a coordinated effort knows the role each of their partners is intending to fulfill, and believes that each of their partners is committed to the common goal (Bratman, 1992; Tomasello, 2009; Dunin-Keplicz & Verbrugge, 2010; Gärdenfors, 2012). This kind of social cooperation would allow for the emergence of culture, collaboration, communication, and teaching.

The Vygotskian intelligence hypothesis would explain both the human capacity for theory of mind and the capacity to engage in altruistic cooperative action (Tomasello, 2009; Gintis, 2009; Bowles & Gintis, 2011). Our results from agent-based simulations in a communication game show that higher-order theory of mind can indeed help to reach a cooperative solution more quickly (De Weerd, Verbrugge, & Verheij, 2015). However, computational models have shown that many forms of cooperation can also emerge through simple mechanisms, without need for a cognitively demanding ability such as theory of mind (Nowak, 2006; De Weerd & Verbrugge, 2011; Van der Post et al., 2013).

A third hypothesis that explains the emergence of theory of mind is the social brain hypothesis (Dunbar, 1998b; Dunbar & Shultz, 2007). According to the social brain hypothesis, human intelligence evolved to deal with increasing complexity of social life. The social brain hypothesis explains why the size of the neocortex as a proportion of total brain size grows with many indicators of social
complexity, such as group size (Dunbar, 1992), grooming clique or social network size (Kudo & Dunbar, 2001), the occurrence of social play (Lewis, 2000), and the formation of longterm pair bonds (Shultz & Dunbar, 2010). These indicators do not relate directly to a competitive or cooperative advantage for individuals, but rather prevent the collapse of groups under the weight of competitive pressure by strengthening the bonds between its members.

Finally, a fourth hypothesis that specifically concerns higher-order theory of mind states that higher orders of theory of mind may be needed for mixed-motive interactions (Verbrugge, 2009). Such mixed-motive interactions involve both cooperative and competitive elements, such as in negotiations or crisis management (Verbrugge, 2009; Van Santen et al., 2009). Individuals interacting in the setting of a mixed-motive situation cooperate with each other by searching for outcomes that are mutually beneficial. At the same time, individuals also compete with each other to reach a mutually beneficial outcome that benefits themselves as much as possible. Mixed-motive interactions can be understood as the task of sharing a pie (Raiffa et al., 2002). When individuals cooperate, they find ways to enlarge the pie they are sharing, while they also compete to obtain as large a piece of pie as possible for themselves. Theory of mind allows individuals to reason explicitly about the goals and beliefs of others. This ability may be crucial for an individual to balance cooperative and competitive goals in order to successfully negotiate a larger pie to share, which includes a larger piece of pie for the individual himself.

In this paper, we focus on the last hypothesis, and determine whether theory of mind allows agents to achieve better outcomes in mixed-motive interactions. Our earlier research into the effectiveness of higher-order theory of mind shows that in a one-shot version of the negotiation game Colored Trails used in this paper, individuals benefit from the ability to make use of theory of mind (De Weerd et al., 2014a). Agents that made use of first-order and second-order theory of mind managed to negotiate a larger piece of pie for themselves than agents of a lower order of theory of mind. In addition, such agents succeeded in negotiating a pie of a larger total size than agents of a lower order of theory of mind. Third-order theory of mind did not provide an agent with benefits beyond those of second-order theory of mind. Surprisingly, fourth-order theory of mind agents could obtain a competitive advantage over competing negotiators.

Although these results show that there are mixed-motive situations in which the ability to make use of higher orders of theory of mind is advantageous to computational agents, the results are based on repeated single-shot negotiation games. In the current paper, we extend the agent model and investigate more realistic mixed-motive settings in which individuals engage in multiple rounds of negotiation. Using agent-based computational models, we simulate agents that alternate in making offers until an agreement is reached. We study mixed-motive situations through the influential Colored Trails setting, introduced by Grosz, Kraus and colleagues (Lin et al., 2008; Gal et al., 2010; De Jong et al., 2011; Van Wissen
6.2. Colored Trails

To determine to what extent reasoning at higher orders of theory of mind results in better offers, we compare performance of computational agents that negotiate in the setting of Colored Trails. Colored Trails is a board game designed as a research test-bed for investigating decision-making in groups of people and computer agents (Lin et al., 2008; Gal et al., 2010; Van Wissen et al., 2012). Our specific setting is similar to the one we used previously to test the effectiveness of higher-order theory of mind in single-shot negotiations (De Weerd et al., 2014a). The game is played by two players on a square board consisting of 25 colored tiles, as depicted in Figure 6.1. At the start of the game, each player receives a set of four colored chips, selected randomly from the same five possible colors as those on the board. Each player is initially located on the center tile of the board, indicated with the letter S in Figure 6.1a. Players can move to a tile adjacent to their current location by handing in a chip of the same color as the destination tile. Figure 6.1b shows an example of a Colored Trails board as well as a possible path across the board. A player following the path from location A to the white tile marked B would have to hand in one black chip, one gray chip, and one white chip.

Each player is also assigned a goal location, which is randomly drawn from the board tiles that are at least three steps away from the initial location (striped tiles in Figure 6.1a). The goal of each player is to approach the goal as closely as possible. To reach that goal, players are allowed to trade chips among each

---

1Also see http://coloredtrails.atlassian.net/wiki/display/coloredtrailshome/.
6.2. Colored Trails

Figure 6.1: The Colored Trails game is played on a 5 by 5 board. (a) Each player starts at the central tile $S$ and receives a goal location drawn randomly from the striped tiles. (b) To follow the path from location $A$ to location $B$, a player needs to hand in one black, one gray, and one white chip.

other. This trading of chips in the Colored Trails setting represents a multi-issue bargaining situation, in which every color represents a different issue or task to overcome. Different paths from the initial location to the goal location on the board represent different ways of achieving the same goal, while each chip represents the means to complete a task or resolve an issue.

In Colored Trails, players are scored based on their success in reaching their goal location. For each step a player takes towards his or her goal, the player receives 100 points. Any player that succeeds in reaching their goal receives an additional 500 points. Finally, any chip that has not been used to move around the board is worth an additional 50 points to its owner. This scoring ensures that players have the highest incentive to reach their goal location, but that they are also motivated to compete over control of unused chips.

Although players are scored based on how closely they approach their own goal, Colored Trails is not a strictly competitive game. Since a player may need a different set of chips to achieve his goal than his trading partner, there may be an opportunity for a cooperative trade, which allows both players to obtain a higher score. That is, although the score of a player is not influenced by how closely his trading partner reaches his or her goal location, players may still benefit from taking into account the goal of their trading partner. However, agents in our Colored Trails setup do not know the goal location of their trading partner from the start.

Trading among players takes the form of a sequence of alternating offers. The initiator makes an initial offer for a redistribution of chips. His trading partner then decides whether or not to accept this offer. If the offer is accepted, the proposed distribution of chips becomes final, the players move as close to their respective goal locations as possible, and the game ends. Alternatively, the trad-
ing partner may decide to withdraw from negotiations, which makes the initial distribution final. Finally, the trading partner may also decide to continue the game by rejecting the offer, and make his own offer for a redistribution of chips.

There are no restrictions on the offers that players can make. For example, a player is allowed to repeat an offer that has been previously rejected by his trading partner, or make an offer that he has previously rejected himself. However, both players pay a 1 point penalty for each round of play. That is, when negotiations end after a total of five offers have been made, the final score of each player is reduced by five points.

In this paper, we investigate to what extent higher orders of theory of mind allow computational agents to make better offers. Based on our previous results in the one-shot variation of Colored Trails (De Weerd et al., 2014a), we expect that theory of mind will provide agents with significant advantages over agents that are more limited in their theory of mind abilities. More specifically, we expect agents that are capable of a higher orders of theory of mind to be able to manipulate the beliefs of their trading partner in order to achieve higher individual scores than agents that are more limited in their theory of mind abilities. Additionally, we also expect that the presence of higher-order theory of mind agents in the negotiation has a positive effect on social welfare, as measured by the sum of the scores of the two negotiating agents. These expectations are captured by hypotheses $H_1$ and $H_2$.

**Hypothesis $H_1$:** First-order and second-order theory of mind agents obtain a higher score than agents that are more restricted in their ability to make use of theory of mind.

**Hypothesis $H_2$:** Social welfare, measured as the sum of the scores over the two agents, increases when the theory of mind abilities of either agent increases.

In Section 6.3, we describe the way computational agents play Colored Trails, and how the ability to reason about the goals of others influences the choices agents make. The formal description of these agents can be found in Section 6.4.

### 6.3 Theory of mind in Colored Trails

In this section, we describe the way theory of mind agents play Colored Trails. The theory of mind agents described here make use of simulation-theory of mind (Davies, 1994; Nichols & Stich, 2003; Hurley, 2008), in which the agent takes the perspective of his trading partner, and determines what his own decision would be if the agent had been in the position faced by his trading partner. Using the implicit assumption that his trading partner’s thought process can be accurately modeled by his own thought process, the agent then predicts that his trading
partner will make the same decision the agent would have made if the roles had been reversed.

In the remainder of this section, we describe how this process of perspective-taking results in different behavior for agents of different orders of theory of mind playing Colored Trails. The formal description of these theory of mind agents is presented in Section 6.4. We will use the shorthand $ToM_k$ agent to indicate an agent that has the ability to use theory of mind up to and including the $k$-th order, but not beyond.

### 6.3.1 Zero-order theory of mind agent

The zero-order theory of mind ($ToM_0$) agent can model the behavior of his trading partner, but the $ToM_0$ agent is unable to attribute mental content to others. In particular, the $ToM_0$ agent is unable to represent that his trading partner wants to reach a certain goal location, and that the behavior of the trading partner is consistent with that desire. A $ToM_0$ agent is essentially fixated on his own piece of pie, and does not consider the piece of pie of other agents at all. Instead, the $ToM_0$ agent constructs zero-order beliefs about the likelihood that his trading partner will accept a certain offer. The $ToM_0$ agent bases these zero-order beliefs on observations of the behavior of the trading partner. For example, through repeated interaction, the $ToM_0$ agent will learn that offers that assign many chips to the trading partner and few to the $ToM_0$ agent are more likely to be accepted, while offers that assign few chips to the trading partner and many to the $ToM_0$ agent himself are more likely to be rejected.

Using these zero-order beliefs, the $ToM_0$ agent can form an expectation about how his score will change if he were to make a particular offer, and select the offer that he assigns the highest expected value. This allows the $ToM_0$ agent to play the Colored Trails setting without attributing mental content to others. That is, although the zero-order beliefs of the $ToM_0$ agent will eventually reflect that his trading partner has a desire for owning chips, the $ToM_0$ agent does not explicitly represent such a desire.

The $ToM_0$ agent engages purely in positional bargaining (Fisher & Ury, 1981), by only reasoning about specific offers that he believes his trading partner will accept, and that he is willing to accept himself. Because the $ToM_0$ agent has no theory of mind, he is unable to represent that his trading partner has interests that underlie the offers that his trading partner is willing to accept.

### 6.3.2 First-order theory of mind agent

In addition to his zero-order beliefs, a first-order theory of mind ($ToM_1$) agent considers the possibility that his trading partner has beliefs and goals as well, which determine whether or not his trading partner will accept an offer. A $ToM_1$
agent therefore realizes that in order to get a large piece of pie for himself, it is essential to enlarge the pie as a whole. The ToM₁ agent is able to consider the game from the perspective of his trading partner, and decide what his action would be if he were in the position of that player.

Since each player wants to increase his own score through negotiation, each player reveals information about his goal location whenever he makes an offer. Although a ToM₁ agent does not know the goal location of his trading partner, the agent can learn the goal location of his trading partner through the offers he receives. For example, consider the situation shown in Figure 6.2, in which agent j offers to trade the black chip owned by agent i against the gray chip owned by agent j. A ToM₁ agent would conclude that owning the black chip allows agent j to move closer to his goal location. Also, since this trade would leave agent j without any gray chips, agent i believes that agent j does not need any gray chips to reach his goal location.

Although the ToM₁ agent is able to consider his trading partner as a ToM₀ agent, the ToM₁ agent does not know the extent of the theory of mind abilities of his trading partner with certainty. Through repeated interactions, the ToM₁ agent may learn that his first-order beliefs fail to accurately model the behavior of his trading partner. If this happens, the ToM₁ agent may choose to play as if he were a ToM₀ agent.

6.3.3 Higher orders of theory of mind agent

Agents that are able to use orders of theory of mind beyond the first order consider the possibility that other agents take into account that others have beliefs and goals as well. A higher-order theory of mind agent reasons about the way his offers influence what his trading partner believes about his goal location. Such an agent may choose to select the offer that provides his trading partner with as
much information as possible about his goal location. This way, a \( \text{ToM}_2 \) agent can inform his trading partner that he prefers a small piece of cherry pie over a large piece of chocolate pie, so that his trading partner can take this into consideration when making an offer.

For example, suppose agent \( i \) in Figure 6.2 is a \( \text{ToM}_2 \) agent with goal location \( G \). Through second-order theory of mind, agent \( i \) knows that if he makes an offer in which the black chip is assigned to agent \( j \), his trading partner will know that agent \( i \) does not need the black chip to reach his goal location. This information may allow agent \( j \) to construct an offer that would be acceptable to agent \( i \), even if agent \( i \) does not know the goal location of his trading partner. That is, higher orders of theory of mind allow agents to communicate their interest to their trading partner through the offers they make, and engage in interest-based negotiation (Fisher & Ury, 1981).

Higher orders of theory of mind also allow agents to manipulate the beliefs of trading partners of a lower order of theory of mind. For example, a higher-order theory of mind agent may construct an offer that gives his trading partner an incorrect impression of his goal location. Such an agent may exaggerate the value of the chips he already possesses and downplay the value of other chips in order to get a better deal. Whether a higher-order theory of mind agent decides to reveal his true goal location or attempt to manipulate the beliefs of his trading partner depends on what the agent believes to result in the highest score for himself.

As with the \( \text{ToM}_1 \) agent, a higher-order theory of mind agent does not know the extent of the theory of mind abilities of his trading partner. Instead, a \( \text{ToM}_k \) agent has \( k+1 \) hypotheses about the future behavior of his trading partner. While playing the Colored Trails game, the \( \text{ToM}_k \) agent continuously updates his beliefs concerning which of these hypotheses best fits the actual behavior of the trading partner.

### 6.4 Mathematical model of theory of mind

In the previous section, we presented the intuition behind theory of mind agents negotiating in a Colored Trails setting using simulation-theory of mind. In this section, we discuss the implementation of computational agents that play according to this intuition. The agents described in this section is inspired by the theory of mind agents we used in De Weerd et al. (2013b) to investigate the effectiveness of theory of mind in competitive settings. This model is extended to allow for generalization over different stage games, and to allow for sequential games.

In our representation, a Colored Trails game is a tuple \( \mathcal{CT} = \langle \mathcal{N}, D, L, \pi_i, \pi_j \rangle \), where:

- \( \mathcal{N} = \{i, j\} \) is the set of agents;
6.4. Mathematical model of theory of mind

• \( \mathcal{D} \) is the set of possible distributions of chips;
• \( \mathcal{L} \) is the set of possible goal locations; and
• \( \pi_i, \pi_j : \mathcal{L} \times \mathcal{D} \rightarrow \mathbb{R} \) are the score functions for agents \( i \) and \( j \) respectively, such that \( \pi_i(l, D) \) denotes the score of agent \( i \) when his goal location is \( l \in \mathcal{L} \) and the chips are distributed according to distribution \( D \in \mathcal{D} \).

In addition, each agent \( i \) knows his own goal location \( l_i \) from the start of the game, while agent \( i \) does not know the goal location \( l_j \) of trading partner \( j \). In the setting we describe, we therefore assume that each agent knows the set of possible offers \( \mathcal{D} \), but has incomplete information about the preferences of his trading partner.

In our representation of the Colored Trails game, we focus on the negotiation aspect and ignore the task of finding routes between locations. This is reflected in the score functions \( \pi_i \) and \( \pi_j \), which specify the maximum score agents \( i \) and \( j \) can achieve given some distribution of chips. This means that we assume that agents make no mistakes in finding routes between locations and that agents do not consider the possibility that their trading partner would make any mistake in finding these routes. Note that these assumptions do not imply that the agents have common knowledge about any aspect of the game. Rather, we assume that our theory of mind agents have no beliefs that contradict a common knowledge about such aspects of the game.

Without loss of generality, we assume that the score of each agent in the initial distribution is zero. That is, \( \pi_i(l, D) \) denotes the change in score of agent \( i \) if his goal location is \( l \in \mathcal{L} \) and the distribution \( D \in \mathcal{D} \) becomes final. In the model description below, we will show formulas from the point of view of agent \( i \). Formulas from the point of view of trading partner \( j \) are analogous.

Over the course of the game, agents alternate in making offers, which results in a sequence of offers \( \{O_0, O_1, \ldots \} \). After the initial offer \( O_0 \) has been made, the agent who received the last offer \( O^t \) decides whether to accept the offer \( O^t \), withdraw from negotiations, or make an offer \( O^{t+1} \) of his own.

In the following subsections, we describe in detail how theory of mind agents play the Colored Trails game. Section 6.4.1, Section 6.4.2, and Section 6.4.3 describe the decision-making process of a \( \text{ToM}_0 \) agent, a \( \text{ToM}_1 \) agent, and a \( \text{ToM}_k \) agent \((k \geq 2)\), respectively. Section 6.4.4 outlines how agents learn within the context of a single game, while Section 6.4.5 describes how agents learn across different games.

6.4.1 Model of zero-order theory of mind

The \( \text{ToM}_0 \) agent described in Section 6.3.1 does not form explicit beliefs about the mental content of others. Instead, the \( \text{ToM}_0 \) agent constructs zero-order beliefs \( b^{(0)} : \mathcal{D} \rightarrow [0, 1] \) about the likelihood \( b^{(0)}(O) \) that a certain offer \( O \) will be
accepted by his trading partner. Using these zero-order beliefs, the ToM$_0$ agent can estimate the value of continuing negotiations by making an offer $O$. Since each round of negotiating costs each agent 1 point, the expected value the ToM$_0$ agent assigns to making offer $O$ is

$$ EV^0_i(O; l_i, b(0)) = b(0)(O) \cdot \pi_i(l_i, O) - 1. \quad (6.1) $$

If the ToM$_0$ agent were to choose to continue negotiation, he would therefore randomly select an offer $O_t \in D$ that he assigns the highest expected value. That is, he selects $O_t$ such that

$$ EV^0_i(O_t; l_i, b(0)) = \max_{O \in D} EV^0_i(O; l_i, b(0)). \quad (6.2) $$

However, making a counteroffer is not the only option available to the ToM$_0$ agent. After the initial offer of a game, an agent can also accept the previous offer $O_{t-1}$ made by his trading partner. Finally, an agent may also decide to withdraw from negotiations, in which case the initial distribution becomes final.

The ToM$_0$ agent rationally decides which of the three options outlined above he will take. That is, the ToM$_0$ agent selects the option that he believes will yield him the highest score, as described in the ToM$_0$ response function

$$ ToM_0i(O_{t-1}; l_i, b(0)) = \begin{cases} O_t & \text{if } EV^0_i(O_t; l_i, b(0)) > 0 \text{ and } EV^0_i(O_t; l_i, b(0)) > \pi_i(l_i, O_{t-1}) \\ \text{accept} & \text{if } \pi_i(l_i, O_{t-1}) > 0 \text{ and } \pi_i(l_i, O_{t-1}) \geq EV^0_i(O_t; l_i, b(0)) \\ \text{withdraw} & \text{otherwise.} \end{cases} \quad (6.3) $$

Equation 6.3 shows that if the ToM$_0$ agent $i$ believes that there is an offer $O_t$ that he expects to yield him a positive score that is strictly higher than the score he would get when accepting the offer $O_{t-1}$, the agent will reject offer $O_{t-1}$ and make counteroffer $O_t$. Alternatively, if the agent believes that there is no such offer $O_t$, but accepting the offer $O_{t-1}$ would give him a positive score, the ToM$_0$ agent accepts the offer $O_{t-1}$. In all other cases, the ToM$_0$ agent withdraws from negotiation.

Note that although the ToM$_0$ agent observes the entire sequence of offers $\{O_0, O_1, \ldots \}$, the agent does not explicitly use the entire sequence of offers to make a decision. Instead, the ToM$_0$ agent decides purely on the basis of his zero-order beliefs $b(0)$, which describe his beliefs about the future behavior of the trading partner.

### 6.4.2 Model of first-order theory of mind

The use of first-order theory of mind allows a ToM$_1$ agent to put himself in the position of his trading partner to consider an offer from the perspective of his
6.4. Mathematical model of theory of mind

To do so, the ToM₁ agent forms first-order beliefs \( b^{(1)} : \mathcal{D} \rightarrow [0, 1] \) that represent what the ToM₁ agent believes to be the zero-order beliefs of his trading partner. That is, the ToM₁ agent believes that his trading partner believes that the probability of the ToM₁ agent accepting a given offer \( O \in \mathcal{D} \) is \( b^{(1)}(O) \).

A ToM₁ agent uses his first-order beliefs to predict his trading partner’s behavior as using the ToM₀ response function described in Equation 6.3. Given the goal location \( l_j \) of his trading partner, the ToM₁ agent calculates the expected value of making offer \( O \in \mathcal{D} \) as

\[
EV^{(1)}_i(l_j, O; l_i, b^{(1)}) = \begin{cases} 
-1 & \text{if } \text{ToM}_0(O; l_j, b^{(1)}) = \text{withdraw}, \\
\pi_i(l_i, O) - 1 & \text{if } \text{ToM}_0(O; l_j, b^{(1)}) = \text{accept}, \\
\max \left\{ \pi_i(l_i, \text{ToM}_0(O; l_j, b^{(1)})) , 0 \right\} - 2 & \text{otherwise.}
\end{cases}
\] (6.4)

Equation 6.4 shows that the ToM₁ agent looks further ahead into the negotiation process than the ToM₀ agent. The ToM₀ agent only forms beliefs about whether or not his trading partner will accept an offer, the ToM₁ agent can also make a prediction about what counteroffer his trading partner could make, and how the ToM₁ agent himself would react to this counteroffer. Since the game is sequential, this results in the ToM₁ agent looking one step further ahead into the negotiation process.

The sequential nature of the game also means that a player may change his beliefs after receiving an offer \( O_{t-1} \), but before deciding whether or not to make a counteroffer \( O_t \). In our agent model, the ToM₁ agent takes this belief update, which will be described in detail in Section 6.4.4, into account. When deciding on whether to make offer \( O \in \mathcal{D} \), the ToM₁ agent determines how making this offer \( O \) would change his zero-order beliefs if he had been in the position of his trading partner, and makes further calculations using the adjusted first-order beliefs \( U(b^{(1)}, O) \) (see Equation 6.10).

Since agents do not know the goal location \( l_j \) of their trading partner from the start, a ToM₁ agent cannot calculate Equation 6.4 directly. Instead, the ToM₁ agent forms beliefs about the goal location of his trading partner in the form of a probability distribution \( p^{(1)} : \mathcal{L} \rightarrow [0, 1] \), so that the ToM₁ agent believes that the likelihood of his trading partner having goal location \( l \in \mathcal{L} \) is \( p^{(1)}(l) \).

Furthermore, although the ToM₁ agent is capable of using theory of mind, he may decide that first-order theory of mind does not accurately describe the behavior of his trading partner. In this case, the ToM₁ agent may decide to rely on his zero-order beliefs \( b^{(0)} \) instead. The ToM₁ agent has a confidence variable \( c_1 \in [0, 1] \), which indicates how much confidence the ToM₁ agent places in the predictions of first-order theory of mind. When deciding on the expected value of making an offer \( O \), the ToM₁ agents weighs the predictions of first-order and
zero-order theory of mind accordingly. In summary, a ToM$_1$ agent $i$ calculates the expected value of making a given offer $O \in \mathcal{D}$ through

$$EV_i^{(1)}(O; l_i, b^{(0)}, b^{(1)}, p^{(1)}, c_1) = (1 - c_1) \cdot EV_i^{(0)}(O; l_i, b^{(0)}) + c_1 \cdot \sum_{l \in L} p^{(1)}(l) \cdot EV_i^{(1)}(l, O; l_i, U(b^{(1)}, O)).$$

(6.5)

Given these values, the ToM$_1$ agent randomly selects an offer $O_t \in \mathcal{D}$ that maximizes his expected value as a suitable counteroffer. Similar to the way the ToM$_0$ agent decides, the ToM$_1$ agent decides to accept, withdraw, or make a counteroffer based on what he expects will yield him the highest score.

$$ToM_i(O_{t-1}; l_i, b^{(0)}, b^{(1)}, p^{(1)}) = \begin{cases} O_t & \text{if } EV_i^{(1)}(O_t) > 0 \text{ and } EV_i^{(1)}(O_t) > \pi_i(l_i, O_{t-1}) \\ \text{accept} & \text{if } \pi_i(l_i, O_{t-1}) > 0 \text{ and } \pi_i(l_i, O_{t-1}) \geq EV_i^{(1)}(O_t) \\ \text{withdraw} & \text{otherwise.} \end{cases}$$

(6.6)

First-order theory of mind benefits the ToM$_1$ agent in two ways. Firstly, theory of mind allows the agent to gain information about the goal location of his trading partner through the offers he receives. He does so by determining how consistent a possible goal location is with the offer his trading partner has made (see Section 6.4.4). Secondly, the ToM$_1$ agent takes into account that making an offer $O$ changes the beliefs and behavior of his trading partner. This may allow the ToM$_1$ agent to make an offer $O_t$ that he expects his trading partner to reject, with the intention of causing his trading partner to make an offer $O_{t+1}$ that the ToM$_1$ agent is willing to accept.

### 6.4.3 Higher-order theory of mind agents

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 agent can use his understanding of first-order theory of mind agents to create an offer that signals his goal location to the trading partner as clearly as possible. Alternatively, the ToM$_2$ agent can adjust his offer to give his trading partner false information about his goal location.

Each additional order of theory of mind allows an agent to consider an additional model of opponent behavior. These models are constructed analogously to first-order theory of mind, and include additional beliefs $b^{(k)}$, location beliefs $p^{(k)}$, and a confidence $c_k$ in $k$th-order theory of mind. Based on the application
of \( k \)th-order theory of mind, the \( ToM_k \) agent formulates the expected value of making an offer \( O \in D \), given that his trading partner has goal location \( l \), as

\[
EV_i^{(k)}(l, O) = \begin{cases} 
-1 & \text{if } ToM^{(k-1)}_{j}(O) = \text{withdraw}, \\
\pi_i(l, O) - 1 & \text{if } ToM^{(k-1)}_{j}(O) = \text{accept}, \\
\max \{\pi_i(l, ToM^{(k-1)}_{j}(O)), 0\} - 2 & \text{otherwise}.
\end{cases}
\] (6.7)

These expected values are then combined with the expected values of lower orders of theory of mind according to

\[
EV_i^{(k)}(O) = (1 - c_k) \cdot EV_i^{(k-1)}(O) + c_k \cdot \sum_{l \in L} p^{(k)}(l) \cdot EV_i^{(k)}(l, O)).
\] (6.8)

This yields the \( k \)th-order theory of mind response function

\[
ToM_{ki}(O_{t-1}) = \begin{cases} 
O_t & \text{if } EV_i^{(k)}(O_t) > 0 \text{ and } EV_i^{(k)}(O_t) > \pi_i(l_t, O_{t-1}) \\
\text{accept} & \text{if } \pi_i(l_t, O_{t-1}) > 0 \text{ and } \\
& \pi_i(l_t, O_{t-1}) \geq EV_i^{(k)}(O_t) \\
\text{withdraw} & \text{otherwise}.
\end{cases}
\] (6.9)

### 6.4.4 Learning from observations

Agents form beliefs about the likelihood that their trading partner will accept a given offer. During negotiation, agents update these beliefs based on the offers their trading partner makes. An agent’s zero-order belief \( b^{(0)} \) specifies that the agent believes that the probability that his trading partner will accept a given offer \( O \in D \) is \( b^{(0)}(O) \). Whenever he receives an offer \( O_{t-1} \) from his trading partner, the \( ToM_0 \) agent updates his beliefs to reflect that he considers it less likely that his trading partner would accept certain offers. More precisely, the \( ToM_0 \) agent decreases his belief that his trading partner will accept an offer \( O \) when offer \( O \) assigns more chips of some color \( c \) to the agent himself than offer \( O_{t-1} \) does. For example, suppose that the trading partner makes an offer \( O_{t-1} \) that assigns 4 blue chips to agent \( i \). Agent \( i \) then decreases his belief that the trading partner will accept any offer that assigns 5 or more blue chips to agent \( i \) himself.

The belief update as a result of receiving an offer \( O_{t-1} \) from the trading partner is represented by \( U(b^{(0)}, O_{t-1}) \), which is defined as

\[
U(b^{(0)}, O_{t-1})(O) = (1 - \lambda)^m \cdot b^{(0)}(O),
\] (6.10)

where \( m \) is the number of colors for which the offer \( O \) assigns fewer chips to the trading partner than the offer \( O_{t-1} \), and \( \lambda \in [0, 1] \) is an agent-specific learning speed.
A similar update takes place when the trading partner rejects an offer made by agent \( i \). When the offer \( O_t \) made by agent \( i \) is rejected, the agent updates his beliefs to reflect that he believes it to be less likely that his trading partner will accept an offer \( O \) that assigns at least as many chips of a given color \( c \) to the agent as offer \( O_t \) does. The belief update as a result of receiving an offer \( O_{t-1} \) from the trading partner is represented by \( U(b^{(0)}, O_{t-1}) \), which is defined as

\[
U_R(b^{(0)}, O_t)(O) = (1 - \lambda)m' \cdot b^{(0)}(O),
\]

where \( m' \) is the number of colors for which the offer \( O \) assigns at least as many chips to agent \( i \) as the offer \( O_t \), and \( \lambda \in [0, 1] \) is an agent-specific learning speed.

Agents that make use of theory of mind also update their beliefs concerning the goal location of their trading partner in response to receiving an offer \( O_{t-1} \) from their trading partner. By putting himself in the position of his trading partner, a ToM\(_k\) agent believes it to be impossible that his trading partner has a goal location \( l \in L \) for which \( \pi(l, O_{t-1}) \leq 0 \). For otherwise, this would mean that the trading partner has made an offer that decreases his score. For other locations \( l \in L \), the agent adjusts his beliefs based on the expected increase in the score of the trading partner if the offer \( O_{t-1} \) would be accepted. That is, after observing the offer \( O_{t-1} \) from his trading partner, the ToM\(_k\) agent updates his location beliefs \( p^{(k)} \) so that

\[
p^{(k)}(l) := \begin{cases} 
0 & \text{if } \pi_j(l, O_{t-1}) \leq 0 \\
\beta \cdot p^{(k)}(l) \cdot \frac{1 + EV^{(k-1)}_i(O_{t-1})}{1 + \max_{O \in D} EV^{(k-1)}_i(O)} & \text{otherwise},
\end{cases}
\]

where \( \beta \) is a normalizing constant. This update increases the beliefs assigned to locations for which the offer \( O_{t-1} \) made by the trading partner receives a high expected value. These are offers that are closer to the offer that the ToM\(_k\) agent would have made himself if he had been a ToM\(_{k-1}\) agent in the position of his trading partner.

**Example 6.1.** Figure 6.3 shows an example of the process of updating location beliefs for a ToM\(_1\) agent. In this example, agent \( j \) offers to trade the black chip owned by agent \( i \) against the gray chip owned by agent \( j \). Agent \( i \) interprets this offer by calculating the change in score for agent \( j \) if agent \( i \) were to accept the offer, for each possible goal location of agent \( j \). In Figure 6.3, these changes in scores are shown on the corresponding locations. For example, if the goal location of agent \( j \) is the tile in the bottom right corner, the score of agent \( j \) would increase by 600 points if agent \( i \) were to accept the offer.

Since making an offer decreases the score of each agent by 1 point, agent \( i \) only makes offers that would increase his own score. By putting himself
6.4. Mathematical model of theory of mind

Figure 6.3: Example of a Colored Trails game in which agent $j$ offers to trade his gray chip against the black chip of agent $i$. Using first-order theory of mind, agent $i$ calculates in what way accepting this offer will affect the score of agent $j$, for each possible goal location. The higher the increase in score, the more likely agent $i$ considers the location to be the goal location of agent $j$.

in the position of his trading partner, agent $i$ therefore believes that agent $j$ also only makes offers that increase the score of agent $j$. Agent $i$ considers it impossible for any location with zero or negative score to be the goal location of his trading partner. That is, when agent $i$ receives offer $O$, for each possible location $l \in L$ with $\pi_j(l, O) \leq 0$, agent $i$ sets $p^{(1)}(l) = 0$.

For the remaining locations $l$, agent $i$ determines what offer $O'$ he would have made himself, and compares how this relates to the offer $O$ that his trading partner actually made. For example, if the goal location of agent $j$ is the bottom left tile, accepting the offer of agent $j$ would only increase his score by 50. However, for this goal location, agent $j$ could have made a better offer. If agent $j$ would have offered to exchange a white chip for the black chip of agent $i$, he could have increased the score of agent $j$ by 150. As a result, agent $i$ believes that it is unlikely that the goal location of agent $j$ is the bottom left tile. On the other hand, agent $i$ considers it very likely that the goal location of his trading partner is one of the locations with a number higher than 500, such as the bottom right tile. If agent $i$ were to accept the offer made by agent $j$, agent $j$ would be able to reach any of these locations.

At the same time as updating his location beliefs, the $ToM_k$ agent also updates his confidence in $k$th-order theory of mind $c_k$ to reflect how well the agent feels the $k$th-order theory of mind model fits the behavior of his trading partner. This is achieved through

$$c_1 := (1 - \lambda) \cdot c_k + \lambda \cdot \sum_{l \in L} p^{(k)}(l) \cdot \frac{1 + EV_i^{(k-1)}(O_{i-1})}{1 + \max_{O \in D} EV_i^{(k)}(O)}.$$  

(6.13)
Using this update, the agent therefore assigns a higher confidence to orders of theory of mind that assign a high expected value to the offer $O_{t-1}$ made by the agent’s trading partner compared to the offer that the agent would have selected himself if he had been a $ToM_{k-1}$ agent in the position of his trading partner. Unlike the way location beliefs are updated, confidences are updated using adaptive expectations. This is because agents may change the order of theory of mind at which they reason over the course of a negotiation, while they are unable to change their goal location.

Many of the belief updates described in this section make use of learning speed parameter $\lambda$. The agent’s learning speed is a fixed parameter that represents the degree to which the agent adjusts his beliefs in response to behavior of his trading partner. For example, a $ToM_0$ agent with a high learning speed strongly believes that his trading partner is unwilling to accept any offers other than the one the trading partner makes himself. Such a $ToM_0$ agent is less likely to make a counteroffer and more likely to withdraw from negotiations or accept the offer of his trading partner. On the other hand, a $ToM_0$ agent with learning speed $\lambda = 0$ does not adjust his behavior at all. Instead, such an agent will keep making the same offer, and expects that this will eventually lead to a successful trade.

Following De Weerd et al. (2013b), theory of mind agents do not attempt to model the learning speed $\lambda$ of other agents. Instead, an agent makes use of his own learning speed when updating the beliefs he assigns to his trading partner. As a result, the beliefs that a theory of mind agent attributes to his trading partner are generally incorrect, unless both agents have the same learning speed.

### 6.4.5 Learning across games

Theory of mind allows agents to view the game from the perspective of their trading partner. This provides theory of mind agents with a way to generalize the behavior of the trading partner across games; theory of mind agents believe that their trading partner plays the game similarly to the way they play the game themselves. However, $ToM_0$ agents do not have this ability. In order to be able to use observations of the behavior of the trading partner in one particular game to determine the likelihood that the trading partner will accept a given offer $O$ in a different game situation, the $ToM_0$ agent therefore needs to generalize across games. Note that learning discussed in this section determines how the $ToM_0$ agent constructs his zero-order beliefs at the start of a negotiation. Over the course of a negotiation, agents perform additional belief updates as described in Section 6.4.4.

In our setting, the $ToM_0$ agent generalizes across games by classifying offers by the number of chips that are transferred from the agent to his trading partner, and the number of chips that are transferred from the trading partner to the agent himself. This allows agents to distinguish, for example, between an offer
that trades one red chip for one blue chip and an offer that trades two red chips for
two blue chips. However, across different games, the agent does not distinguish
between an offer that trades one red chip for one blue chip and an offer that trades
one green chip for one yellow chip. Since agents in our setting own an initial set
of four chips, this generalization causes the $ToM_0$ agent to distinguish 25 classes
of offers. Nevertheless, a separate pilot study indicated that this simple heuristic
allowed agents to make mutually beneficial offers after a short learning period.

At the start of each game, the agent’s zero-order beliefs $b^{(0)}(O)$ about the
probability that the trading partner will accept a given offer $O \in D$ is set to the
observed frequency with which offers that transfer the same number of chips from
the agent to the trading partner and the same number of chips from the trading
partner to the agent have been accepted by the trading partner in the past. For
example, if a $ToM_0$ agent has made 250 offers in which two chips were transferred
to the trading partner and one chip to the agent, of which 220 have been accepted
by the trading partner, the $ToM_0$ agent assigns a probability of 88% that his
trading partner will accept an offer to trade two green chips owned by the agent
against one red chip owned by the trading partner at the start of the game. Over
the course of the negotiation process, this belief can still change, as described in
Section 6.4.4.

6.5 Simulation results

We performed simulations where the theory of mind agents described in Section 6.4
played the Colored Trails setting described in Section 6.2. Pairs of agents played
repeated negotiation games, where each individual game is played on a different
board with different sets of initial chips and different goal locations. In each new
game, agents started by reasoning at the highest order of theory of mind available
to them. For example, a $ToM_2$ agent always started the game by taking into
account the beliefs his trading partner might have about his own beliefs.

To ensure that all agents had an incentive to negotiate to increase their score,
games in which some agent could reach his goal location with the initial set of
chips without trading were excluded from analysis. Additionally, the first 200
games were considered to be a setup phase for the zero-order beliefs of agents,
which were initialized at 1. That is, at the start of a simulation, an agent believes
that any offer will be accepted. During the first 200 games, agents may learn,
for example, that an “offer” that consists of requesting the trading partner’s full
set of chips is unlikely to be accepted. The results from these 200 training games
were not included in analysis.

The figures in this section show the average change in score due to negotiation,
which is calculated as the average difference between an agent’s final score after
negotiation ended and his initial score at the start of negotiation. Although agents
never accept an offer that decrease their score, negative scores are possible when agents take many rounds. In these cases, the cost of negotiation can outweigh the benefits of a mutually beneficial trade. Negotiation scores were averaged over 1,000 consecutive Colored Trails games. Although negotiations could take infinitely long in theory, games that continued for more than 100 rounds of offers 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, and negotiated as if this limit did not exist.

Although the agents alternate in making offers, so that both agents make offers to their trading partner, previous research into negotiation suggests that the opening bid of a negotiation can serve as an anchor for the entire negotiation process, making the first bid of a game especially influential (Raiffa et al., 2002; Van Poucke & Buelens, 2002; Rosette, Kopelman, & Abbott, 2013). Because of this, we differentiate between results for initiators, who make the first offer in every game, and responders.

In the following subsections, we separate results for the competitive and cooperative aspects of negotiation in Colored Trails. In Section 6.5.1, we present the individual performance of agents, which shows how well agents compete. Section 6.5.2 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.

### 6.5.1 Individual performance results

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

Figure 6.4 shows the average negotiation scores of a ToM₀ initiator and a ToM₀ responder negotiating with each other, as a function of the learning speeds of the two agents. In these figures, a lighter color indicates a higher score. As a visual aid, the plane of zero performance appears as a semi-transparent plane in these figures. Figure 6.4 shows that even though ToM₀ agents are unable to reason explicitly about the goals and desires of their trading partner, they are often able to increase their score through negotiation. The ToM₀ agents only fail to reach a positive negotiation score when both agents have learning speed $\lambda = 0$. An agent with zero learning speed does not adjust his behavior in response to his trading partner, but repeats the same offer until his trading partner makes an acceptable offer. That is, an agent with zero learning speed expects his trading partner to adjust to his position, while the agent is unwilling to offer any alternatives himself. When both ToM₀ agents follow this strategy and neither is willing to accept the
6.5. Simulation results

Figure 6.4: Average negotiation score of a ToM₀ initiator (left) and a ToM₀ responder (right) as a function of their respective learning speeds.

Figure 6.5: Average negotiation score of a ToM₀ initiator (left) and a ToM₁ responder (right) as a function of their respective learning speeds.

initial offer of their trading partner, they will be unable to reach an agreement and only carry the burden of a failed negotiation.

The results in Figure 6.4 also show that for ToM₀ agents, negotiation score increases as their own learning speed decreases unless the trading partner has learning speed \( \lambda = 0 \). This means that there is an evolutionary pressure on ToM₀ agents to decrease their learning speed, and adjust the offers they make as slowly
as possible. However, this eventually results in the worst possible outcome in which negotiation fails.

Negotiation failure does not occur when a ToM\(_0\) initiator negotiates with a ToM\(_1\) responder. Figure 6.5 shows that for every combination of learning speeds of the ToM\(_0\) initiator and the ToM\(_1\) responder, both agents receive a positive score on average. However, the evolutionary pressure to reduce learning speed still exists for the ToM\(_0\) initiator. A ToM\(_0\) initiator receives the largest piece of pie when his learning speed is \(\lambda = 0\), in which case he leaves only a small piece of pie for the ToM\(_1\) responder. This means that although the presence of the ToM\(_1\) responder prevents negotiation failure, the ToM\(_0\) initiator benefits most.

Figure 6.6 shows a similar pattern when the roles are reversed, so that a ToM\(_1\) initiator and a ToM\(_0\) responder play Colored Trails. The ToM\(_0\) responder performs best when his learning speed is zero, while the ToM\(_1\) initiator prefers a higher learning speed. This allows the agents to negotiate successfully, with the ToM\(_0\) responder receiving the most benefit. The presence of a ToM\(_1\) agent yields a larger pie for the negotiating agents to share, but it is the ToM\(_0\) agent that receives the largest piece.

Figure 6.6 also shows that when the ToM\(_1\) initiator has a learning \(\lambda \leq 0.4\), the score of both agents is zero. In these cases, the ToM\(_1\) initiator withdraws from negotiation instead of making an 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. This causes the ToM\(_1\) agent to believe that the likelihood of finding a mutually beneficial trade is not worth the cost of negotiating. As a result, a ToM\(_1\) initiator with learning speed \(\lambda = 0\) withdraws from negotiation before making the initial offer.

Figure 6.7 shows the negotiation scores of a ToM\(_1\) initiator and a ToM\(_1\) responder negotiating in Colored Trails. The figures show symmetry around the line of equal learning speeds that indicates that the ToM\(_1\) agent with the lower learning speed generally receives the largest piece of the pie. A ToM\(_1\) agent with a higher learning speed attributes this learning speed to his trading partner and expects that his offers will influence the behavior of his trading partner more strongly. This also leads a ToM\(_1\) agent to believe 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 he believes to be more beneficial for his trading partner at the expense of his own score. This puts evolutionary pressure on ToM\(_1\) agents to lower their learning speed. However, since ToM\(_1\) agents perform poorly when their learning speed falls below \(\lambda = 0.2\), the evolutionary pressure on ToM\(_1\) agents is to have learning speeds close to \(\lambda = 0.2\).

Figure 6.8 and Figure 6.9 show the performance of ToM\(_1\) agents and ToM\(_2\) agents that negotiate with each other. The graphs show that the ToM\(_2\) agent
6.5. Simulation results

Figure 6.6: Average negotiation score of a $ToM_1$ initiator (left) and a $ToM_0$ responder (right) as a function of their respective learning speeds.

(a) $ToM_1$ initiator score  
(b) $ToM_0$ responder score

Figure 6.7: Average negotiation score of a $ToM_1$ initiator (left) and a $ToM_1$ responder (right) as a function of their respective learning speeds.

(a) $ToM_1$ initiator score  
(b) $ToM_1$ responder score

typically has a higher score than his $ToM_1$ trading partner, irrespective of the roles and learning speeds of the agents. That is, the $ToM_2$ agent is highly effective in obtaining a larger piece of the pie than his trading partner.

The negotiation scores of two $ToM_2$ agents negotiating with each other are shown in Figure 6.10. Interestingly, the performance of the $ToM_2$ initiator in Figure 6.10a is quite similar to the performance of the $ToM_2$ responder shown in
6.5. Simulation results

Figure 6.8: Average negotiation score of a ToM$_2$ initiator (left) and a ToM$_1$ responder (right) as a function of their respective learning speeds.

Figure 6.9: Average negotiation score of a ToM$_1$ initiator (left) and a ToM$_2$ responder (right) as a function of their respective learning speeds.

Figure 6.10b. Whereas results from lower orders of theory of mind show many opportunities to divide the pie in one large piece and one small piece, ToM$_2$ agents generally divide the pie in two pieces that are similar in size. As a result, the graphs in Figure 6.10 show little variation in color. Nevertheless, ToM$_2$ agents that have a positive learning speed that is close to zero tend to do slightly better than ToM$_2$ agents that have a different learning speed.
6.5. Simulation results

In summary, the results in this section show that the ability to make use of theory of mind can help individuals to negotiate better, although they do not show the same pattern as found for competitive games (De Weerd et al., 2013b) as predicted by hypothesis $H_1$ in Section 6.2. Even though the presence of $ToM_1$ agents prevent negotiation failure in our simulations, the $ToM_1$ agent does not have a direct competitive advantage over a $ToM_0$ agent. Instead, the $ToM_1$ agent suffers the cost for achieving a cooperative solution, which leaves the $ToM_0$ agent with the larger piece of pie.

The $ToM_2$ agent, on the other hand, does outperform the $ToM_1$ agent as expected by hypothesis $H_1$. Our results show that the $ToM_2$ agent can negotiate successfully with a $ToM_1$ trading partner, resulting in a pie that includes a large piece for the $ToM_2$ agent. When two $ToM_2$ agents negotiate with each other, the resulting pie is divided into pieces of a fairly similar size. In the next subsection, we take a closer look at the cooperative abilities of these theory of mind agents.

### 6.5.2 Social welfare results

In the previous section, we compared the individual competitive performance of agents of various orders of theory of mind negotiating in Colored Trails. In this section, we show how theory of mind affects the cooperative ability of agents, by looking at the social welfare that theory of mind agents achieve through negotiation, where social welfare is measured by the sum of the scores of the initiator and the responder. Figure 6.11 and Figure 6.12 show the increase in social welfare for
6.5. Simulation results

Figure 6.11 shows the average combined negotiation score of ToM\(_0\) and ToM\(_1\) agents playing Colored Trails. Figure 6.11a shows that ToM\(_0\) agents can cooperate surprisingly well. In the best-case scenario, both the ToM\(_0\) initiator and the ToM\(_0\) responder have a learning speed of around \(\lambda = 0.2\), which results in an average increase in the social welfare of over 400 points. However, due to the competitive element of Colored Trails described in Section 6.5.1, cooperation among ToM\(_0\) agents is not stable. The ToM\(_0\) agents experience an evolutionary pressure towards zero learning speed, which can eventually lead to negotiation failure.
6.5. Simulation results

Although Section 6.5.1 shows that the presence of a ToM₁ agent can ensure that negotiation failure does not occur, Figure 6.11 shows that this does not imply a higher social welfare. Figure 6.11b and Figure 6.11c do not show an improvement over the performance of ToM₀ agents shown in Figure 6.11a.

Figure 6.11d shows that when two ToM₁ agents play Colored Trails together, they achieve the highest social welfare when both agents have a learning speed as high as possible. However, the competitive element in Colored Trails puts an evolutionary pressure on ToM₁ agents to lower their learning speed to a value of
6.6 Summary and discussion

We have investigated the claim that the human ability for higher-order theory of mind may have arisen because of the existence of mixed-motive settings in which the use of higher-order theory of mind presents individuals with an evolutionary advantage (Verbrugge, 2009). For that purpose, we have simulated interactions

\[ \lambda = 0.2. \] Although this does not lead to a breakdown of negotiation like in the case of \( ToM_0 \) agents, social welfare suffers from the lower learning speed of \( ToM_1 \) agents. The individual desire of \( ToM_1 \) agents to obtain as large a piece of pie as possible results in a smaller pie to share.

Although the increase in social welfare depends greatly on the learning speed of \( ToM_0 \) agents and \( ToM_1 \) agents, Figure 6.12a and Figure 6.12b show that the learning speed of a \( ToM_2 \) agent has little effect on social welfare when a \( ToM_2 \) agent and a \( ToM_1 \) agent negotiate in Colored Trails. Instead, how negotiation affects social welfare in these cases is determined mostly by the learning speed of the \( ToM_1 \) agent.

When two \( ToM_2 \) agents negotiate, the highest social welfare is achieved when both agents have a low but positive learning speed, as shown in Figure 6.12c. Note that this learning speed also yields them the highest score individually. That is, when two \( ToM_2 \) agents negotiate, the learning speed that would yield an agent the largest piece of pie is also the learning speed that would yield the largest pie to share. However, note that the highest social welfare that the \( ToM_2 \) agents achieve is not as high as the highest social welfare achieved by \( ToM_0 \) agents.

In summary, contrary to hypothesis \( H_2 \) formulated in Section 6.2, our simulation results do not provide any evidence to support that theory of mind directly increases social welfare. However, we find that theory of mind can help to stabilize negotiation. While \( ToM_0 \) agents have the potential to negotiate a high social welfare, natural selection would favor those \( ToM_0 \) agents that increase their individual score at the expense of social welfare. These \( ToM_0 \) agents therefore face a social dilemma similar to the prisoner’s dilemma, which leads \( ToM_0 \) agents to a situation in which negotiation breaks down.

A \( ToM_1 \) agent is able to avoid complete breakdown of negotiation by recognizing the need for cooperation. However, \( ToM_1 \) agents face a similar social dilemma in which the individual desire to obtain as large a piece of pie as possible leads to a smaller pie to share. Interestingly, \( ToM_2 \) agents do not face the same social dilemma. When a \( ToM_2 \) agent negotiates with a \( ToM_1 \) or \( ToM_2 \) trading partner, the individual goal to obtain as large a piece of pie as possible leads to a pie for which the total size is as large as possible as well. In this sense, higher-order theory of mind can benefit social welfare in negotiation settings such as Colored Trails.
between computational agents to show how higher orders of theory of mind can help in obtaining better outcomes in negotiation.

We investigated a particular mixed-motive setting known as Colored Trails (Lin et al., 2008; Gal et al., 2010; Van Wissen et al., 2012), in which agents of various orders of theory of mind alternate in offering a redistribution of chips under incomplete information about the preferences of their trading partner. Zero-order theory of mind agents only negotiate in terms of the positions they are willing to accept without any regard for the goal of their trading partner. First-order theory of mind agents recognize the need for a mutually beneficial outcome, and reason explicitly about the goals of others. Second-order theory of mind allows agents to communicate their own goals and engage in interest-based negotiation (Fisher & Ury, 1981) to uncover mutually beneficial solutions.

We found that under the right conditions, agents could successfully negotiate a mutually beneficial trade in Colored Trails without any theory of mind ability. However, agents experience an evolutionary pressure to increase their own score at the expense of their trading partner. Through natural selection, this eventually leads to a situation in which all negotiation fails. For zero-order theory of mind agents, negotiation in Colored Trails is similar to the prisoner’s dilemma, where the competitive aspect of getting as much of the pie as possible overshadows the cooperative aspect of negotiation to the point where there no longer is any pie to share.

By reasoning explicitly about the goals of the trading partner, first-order theory of mind agents prevent a complete breakdown in negotiation. First-order theory of mind thus allows agents to better balance the cooperative and competitive aspects of negotiation. However, first-order theory of mind agents still have an incentive to increase their own piece of pie at the expense of the total size of the pie. That is, for first-order theory of mind agents, the competitive aspect of negotiation continues to be a hindrance to efficient cooperation.

Although first-order theory of mind has a limited effectiveness in the negotiation setting we describe, second-order theory of mind benefits agents greatly. When a second-order theory of mind agent negotiates with another agent capable of theory of mind, the second-order theory of mind agent typically receives the larger share of the pie. Additionally, neither agent has an incentive to deviate from the outcome that maximizes total pie size. That is, second-order theory of mind agents balance cooperative and competitive goals to the point where agents that succeed in negotiating the largest total pie possible could not have received a larger piece of pie for themselves by changing their behavior. These results show that in mixed-motive settings such as negotiations, agents can benefit from the use of higher-order theory of mind.

In existing literature, there are several approaches to bounded rationality and recursive modeling of the behavior of others that are similar to our theory of mind agents. In behavioral economics, recursive modeling of the behavior of others is
modeled through iterated best-response models such as level-$n$ theory (Stahl & Wilson, 1995; Bacharach & Stahl, 2000; Costa-Gomes et al., 2001), cognitive hierarchies (Camerer et al., 2004), quantal response equilibria (McKelvey & Palfrey, 1995), and noisy introspection models (Goeree & Holt, 2004). In these models, an agent’s level of sophistication is measured by the maximum number of steps of iterated reasoning the agent is capable of considering. In terms of our theory of mind agents, an agent that only makes one reasoning step corresponds roughly to zero-order theory of mind. The goal of these approaches is to describe the level of sophistication of human participants. Over a range of one-shot non-repeated games such as the $p$-beauty contest and the traveler’s dilemma, participants are estimated to use an average of 1.5 steps of iterated reasoning (Camerer et al., 2004), while only few players were found to be well-described as higher-level agents (Wright & Leyton-Brown, 2010).

In the models described above, a level-$n$ agent assumes that all other agents are exactly one level of sophistication lower than himself, or that the distribution of lower level agents can be described with a fixed probability distribution. However, in repeated game settings, such assumptions can be detrimental to an agent (Hu & Wellman, 1998). The theory of mind agents we describe are more similar to models of recursive opponent modeling (Gmytrasiewicz & Durfee, 1995; Gmytrasiewicz et al., 1998), interactive POMDPs (Gmytrasiewicz & Doshi, 2005), and game theory of mind (Yoshida et al., 2008). In all these approaches, agents adjust their level of recursive reasoning in reaction to the behavior of others. An agent of level $k$ can consider others as being agents of any level up to and including level $k - 1$. Such an agent does not observe the level of sophistication of others directly, but forms beliefs concerning the level of sophistication of others based on observed behavior.

In contrast to previous work, in which the most basic agent typically assumes that the behavior of others can be modeled as noise, our zero-order theory of mind agent continues to actively models the behavior of other agents. This means that our zero-order theory of mind agent reacts to the actions of other agents by adjusting his behavior accordingly. In particular, a zero-order agent may learn more sophisticated behavior by observing the behavior of higher-order theory of mind agents.

In addition, the goal of our work is to identify settings in which there is an evolutionary incentive to reason using higher orders of theory of mind which could explain the emergence of human-like theory of mind abilities. However, although we model human-like theory of mind abilities, our goal is not to replicate actual human social behavior, as agent-based simulation tools such as PsychSim (Pynadath & Marsella, 2005). Rather, we explicitly compare simple learning strategies that rely solely on modeling the behavior of others with more complex strategies that include theory of mind to determine the extent of their effectiveness. In this sense, our work also differs from formal methods such as dynamic epistemic logic.
6.7 Conclusion

Behavioral experiments show that people can use higher-order theory of mind to reason about the way others reason about mental content, while other animals do not appear to have this ability. One of the possible explanations for the emergence of the human ability to make use of higher-order theory of mind predicts that this ability is especially useful in mixed-motive situations, where both cooperative and competitive aspects play a role. Through agent-based simulations, we have shown that higher orders of theory of mind indeed help agents to better balance cooperative and competitive goals.

In our earlier research into the effectiveness of higher-order theory of mind, we found that first-order and second-order theory of mind allow agents to agree on a cooperative solution faster than agents without theory of mind abilities (De Weerd, Verbrugge, & Verheij, 2015). However, in strictly cooperative settings, theory of mind is not needed to maintain cooperation once a cooperative solution has been found; a zero-order agent can simply copy the strategy to stabilize cooperation. We find that higher-order theory of mind does convey a cooperative advantage in the mixed-motive setting we investigate in the current work. When both cooperative and competitive elements play a role, higher-order theory of mind can help to balance these elements to stabilize the cooperative solution. Interestingly, this cooperative solution is achieved purely through calculated selfishness; second-order theory of mind agents behave cooperatively, not because (Fagin et al., 1995; Van Ditmarsch et al., 2007), which are used to study recursive reasoning about the knowledge of others from a prescriptive perspective. In contrast, our theory of mind agents typically construct an incorrect model of the beliefs of others. We determine to what extent reasoning at increasingly higher orders of theory of mind remains effective, even under these conditions.

Our work is also related to research into automated agents in negotiation applications (Lin, Gal, Kraus, & Mazliah, 2014; Rosenfeld, Zuckerman, Segal-Halevi, Drein, & Kraus, 2014; Fatima, Kraus, & Wooldridge, 2014; Ficici & Pfeffer, 2008; Lin et al., 2008; Kraus, 1997). Previous work in continuous double auctions suggests that higher levels of sophistication may not always be beneficial for agents in terms of their payoff (Hu & Wellman, 1998; Wellman & Hu, 1998; Park, Durfee, & Birmingham, 2004). In contrast, our results suggest that in repeated Colored Trails games, in which it is difficult to generalize information across different game setups, automated agents may benefit from several rounds of recursive modeling, even when agents are faced with incomplete information about the preferences of their trading partner. This may be especially useful in situations where computational agents interact directly with humans, who are known to make use of theory of mind.
they have an innate sense of fairness, but because they believe that it will result in a better outcome for themselves.

Our earlier results also show that in strictly competitive settings, both first-order and second-order theory of mind agents outperform opponents of a lower order of theory of mind, while the advantage of applying even higher orders of theory of mind are limited (De Weerd et al., 2013b). In contrast, although second-order theory of mind agents outperform first-order theory of mind agents in our Colored Trails setting, we find that first-order theory of mind agents do not outperform zero-order theory of mind agents. The reason for this difference is that in strictly competitive situations, an agent that increases his own score does so at the expense of his opponent. In a mixed-motive situation, increasing your own score may increase the score of your trading partner as well. In Colored Trails, the first-order theory of mind agent increases his own score by preventing negotiation failure. However, this increases the score of his trading partner even more. As a result, the zero-order theory of mind agent obtains a larger piece of the pie. Still, from the perspective of the first-order theory of mind agent, obtaining a small piece of pie is preferable to no pie at all.

In current work, we aim to let theory of mind agents negotiate directly with human participants. This way, we can determine the extent to which participants make use of higher-order theory of mind during negotiations. Additionally, theory of mind agents may help participants to develop their negotiation skills. By confronting participants with agents capable of using increasingly higher orders of theory of mind, they may be able to train their higher-order theory of mind to reach better negotiation outcomes.

Acknowledgments

This work was supported by the Netherlands Organisation for Scientific Research (NWO) Vici grant NWO 277-80-001, awarded to Rineke Verbrugge for the project ‘Cognitive systems in interaction: Logical and computational models of higher-order social cognition’. 