Autonomous Agents and Multi-Agent Systems (2022) 36:30
https://doi.org/10.1007/s10458-022-09558-6

Higher-order theory of mind is especially useful in unpredictable negotiations

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

Accepted: 12 April 2022

Abstract

In social interactions, people often reason about the beliefs, goals and intentions of others. This *theory of mind* allows them to interpret the behavior of others, and predict how they will behave in the future. People can also use this ability recursively: they use *higher-order theory of mind* to reason about the theory of mind abilities of others, as in "he thinks that I don't know that he sent me an anonymous letter". Previous agent-based modeling research has shown that the usefulness of higher-order theory of mind reasoning can be useful across competitive, cooperative, and mixed-motive settings. In this paper, we cast a new light on these results by investigating how the predictability of the environment influences the effectiveness of higher-order theory of mind. Our results show that the benefit of (higher-order) theory of mind reasoning is strongly dependent on the predictability of the environment. We consider agent-based simulations in repeated one-shot negotiations in a particular negotiation setting known as Colored Trails. When this environment is highly predictable, agents obtain little benefit from theory of mind reasoning. However, if the environment has more observable features that change over time, agents without the ability to use theory of mind experience more difficulties predicting the behavior of others accurately. This in turn allows theory of mind agents to obtain higher scores in these more dynamic environments. These results suggest that the human-specific ability for higher-order theory of mind reasoning may have evolved to allow us to survive in more complex and unpredictable environments.

Keywords Theory of mind · Mixed-motive situation · Negotiation · Complex environment

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'.

✉ Harmen de Weerd
harmen.de.weerd@rug.nl

[1] Department of Artificial Intelligence, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, Faculty of Science and Engineering, University of Groningen, Groningen, The Netherlands

[2] Research group User-Centered Design, School of Communication, Media & IT, Hanze University of Applied Sciences, Groningen, The Netherlands

Published online: 06 May 2022

1 Introduction

When engaging in social interaction, people often rely on their ability to reason about what others know and believe. *Theory of mind*, the ability to attribute mental content to others [56], helps people to understand why others behave the way that they do, to predict their future behavior, as well as to distinguish between intentional and accidental behavior. People also have the ability to use this theory of mind ability recursively, by reasoning about the theory of mind of others. Second-order theory of mind allows people to form nested beliefs such as "Alice *believes* that Bob does not *know* that Carol is throwing him a surprise party", and use these beliefs to predict that Alice will most likely avoid talking about the surprise party while Bob is around.

The human ability to make use of second-order and even higher orders of theory of mind is well-established experimentally [2, 4, 25, 38, 54, 64, 66]. Moreover, the ability to make use of higher-order theory of mind has been associated with higher social competences [45, 48], better negotiation skills [19], and pro-social behavior [40]. However, while primates [8, 62], dogs [43] and corvids [6, 7, 14] have been shown to be able to take at least some mental content of others into account, only humans have been shown an ability for higher-order theory of mind.

Our goal is to show that within a given setting, theory of mind reasoning is more beneficial when the environment is less predictable. Previous agent-based modeling studies have attempted to explain why the uniquely human ability to make use of higher-order theory of mind may have evolved, by looking at environments in which this ability is particularly beneficial to the individual (see Sect. 6). In this paper, we take a look at a different dimension, and investigate how the predictability of the environment affects the effectiveness of (higher-order) theory of mind. We expect that since theory of mind allows agents to generalize the behavior of others across different scenarios (cf. [19, 21, 42]), this may put theory of mind agents at an advantage over agents that rely only on observable features of the environment to predict the behavior of others. This expected advantage for (first-order) theory of mind reasoning may also allow for additional benefits of higher-order theory of mind reasoning.

We make use of the influential Colored Trails setting, introduced by Grosz, Kraus and colleagues [20, 30, 46, 68][1]. We consider a particular Colored Trails setup with three players that is similar to the one used by Ficici and Pfeffer [24], in which they show that human participants indeed make use of theory of mind when playing this game. This setting is particularly interesting because of the way it separates competitive and cooperative aspects (see Sect. 2 for more details).

To determine how the predictability of the environment influences the benefits of reasoning at higher orders of theory of mind, we perform simulation experiments, in which computational agents of various orders of theory of mind negotiate among one another. To this end, we have extended our agent model for theory of mind reasoning, which we have previously used to determine the effectiveness of theory of mind in two-player alternating-offers negotiations [19], to allow for single-shot negotiations among three players (see Sect. 3).

The remainder of the paper is structured as follows. In Sect. 2, we describe the details of our version of the Colored Trails game and our hypotheses (Sect. 2.1) concerning the

[1] Also see http://coloredtrails.atlassian.net/wiki/display/coloredtrailshome/.

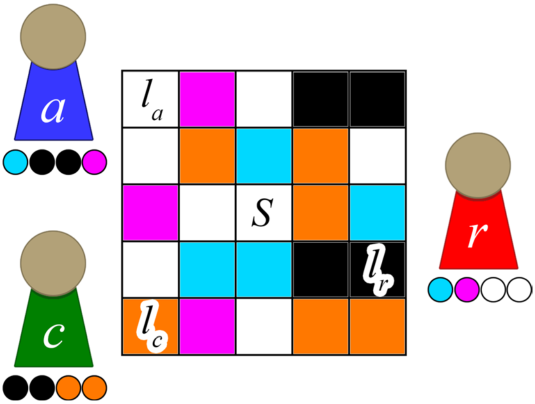

Fig. 1 In our setup, Colored Trails is played by three agents, allocator a, competitor c, and responder r. These agents are initially located at the tile marked S at the center of the board, and are each assigned a goal location. In this example, allocator a, competitor c, and responder r have goal locations l_a, l_c, and l_r, respectively

influence of the predictability of the environment on the effectiveness of higher-order theory of mind. In Sect. 3, we outline how theory of mind may be helpful in this particular setting. This outline is supported by a full mathematical description of our theory of mind agents in Sect. 4. This section also contains several numerical examples of the way agents play Colored Trails. Section 5 describes the setup and the results of the simulation experiments we performed with our theory of mind agents. Finally, Sect. 6 relates our results to existing literature, while Sect. 7 provides a discussion of our results and suggests directions for future research.

2 Game setting

In this paper, we simulate interactions between computational agents playing a particular negotiation game known as Colored Trails [20, 30, 46, 68]. The specific Colored Trails setting we use for our simulations is played by three players on a 5 by 5 board of colored tiles, such as the one depicted in Fig. 1. Each player starts at the center of the board (marked S in Fig. 1), and aims to get as close as possible to their goal location, denoted l_a,l_c, and l_r for agents a, c, and r, respectively.

In addition, each player receives a set of four colored chips, drawn from the same colors as the colors on the board. Players can move from their current tile to an adjacent tile by handing in a chip of the same color as the destination tile. For example, a player who wants to move to the right from the starting location in Fig. 1 would have to hand in an orange chip. If that player would then want to move down, they would also have to hand in an additional black chip.

To quantify performance, we score players based on how close they end up to their goal location. Similar to the scoring in [30], we award 50 points to each player that reaches his goal tile. If a player is unable to reach his goal tile, he pays a penalty of 10 points for each tile in the shortest path from his current location to his goal location. In addition to reaching the goal location, a player can increase his score by owning chips. Chips that have not been used to move on the board increase the score of their owner by 5 points each.

Note that players may not always have the chips needed to reach their goal location. To help them get the chips they need, players are allowed to trade chips with one another. Similar to the Colored Trails setting described by Ficici and Pfeffer [24], the setting we consider here is played by three agents: the *allocator*, the *responder*, and the *competitor*[2]. For clarity, we assign genders to the roles. We will refer to the allocator and the competitor as if they were male, while we will refer to responders as if they were female.

Negotiation takes the form of single-shot bargaining. The allocator and the competitor simultaneously make an offer to exchange any subset of their own chips against any subset of the responder's chips. We consider a fully observable setting, which means that all information about the game board, goal locations, initial locations, and initial sets of chips is observable to each of the agents. Once both allocator and competitor have selected their offers, the offers are revealed to all players. There are no costs associated with making an offer.

Once both offers have been revealed, the responder decides whether or not to accept one of these offers. If the responder does accept one of the offers, the trade is carried out. If the responder chooses not to accept either offer, the initial distribution of chips becomes final. In both situations, the negotiation ends, players move as close to their respective goal locations as possible and are scored accordingly.

Example 1 Consider the situation depicted in Fig. 1. In this scenario, the allocator a can only move one step towards his goal location l_a with his initial set of chips, and would therefore reach a score of −15 (three tiles short of his goal, with three spare chips). With the initial set of chips of the responder r, however, he could reach his goal location. That is, if the allocator a and responder r would exchange chips, allocator a would be able to reach a score of 50. Allocator a therefore offers to exchange his set of chips against the set of chips of the responder. Note that this deal is beneficial to the responder r as well. Like the allocator, responder r can only move one step closer to her goal location l_r with her initial set of chips, resulting in a score of −5 (two steps short of the goal, with three spare chips). If she accepts the offer, she would be able to reach her goal location with one spare chip, resulting in a score of 55.

The competitor c offers to exchange both his black chips against the white and purple chips of the responder r. If this offer would be accepted, the competitor would have the chips needed to reach his goal location l_c with an orange chip to spare, and thereby obtain a score of 55 points. The responder r would be left with only three chips, namely a light blue chip and two black chips. However, this is enough for her to reach her goal location and obtain 50 points.

The competitor and allocator make their offers simultaneously, so that neither of them knows what the other agent has offered the responder before making a choice themselves. Once both offers are made, they are revealed to all agents and the responder r now decides whether to accept one of these offers. Since both offers represent an improvement of the responder's initial score of −5, she decides to accept an offer. Moreover, since the allocator's offer would increase her score by 60, while the competitor's offer would increase her score by only 55, she decides to accept the offer of the allocator.

As the example above illustrates, our current Colored Trails setting separates cooperative and competitive aspects in the form of the responder and competitor agents. The responder represents the cooperative aspect of negotiations. After all, an allocator's offer will only be acceptable if it increases the score of the responder. That is, the

[2] While we differentiate between the allocator and the competitor to facilitate understanding, the only difference between the roles of allocator and competitor is their name.

allocator should find ways to expand the metaphorical pie (cf. [58]) that is being bargained about, so that both the responder and the allocator can get a larger piece.

The competitor, on the other hand, represents the competitive side of negotiations. To convince the responder to choose his offer over the one made by the competitor, the allocator should offer the responder a piece of the pie that is at least as large as the piece offered by the competitor. Through competition with the competitor, the allocator may therefore choose to divide the pie [58] so that it includes a larger piece for the responder, at the expense of his own piece of the pie.

2.1 Unpredictability of the environment

Our Colored Trails setup is very similar to the one used by Ficici and Pfeffer [24], who show that human participants indeed make use of theory of mind when playing this game. We therefore expect that under the right circumstances, higher-order theory of mind agents will outperform agents with more limited theory of mind abilities. In our simulation experiments, we consider how the benefit of higher-order theory of mind varies with the predictability of the environment. To this end, we consider three different types of environment, listed below.

- *Static environment with static goals* Agents repeatedly play Colored Trails on the same board, with always the same set of chips and goal locations;
- *Static environment with dynamic goals* Agents repeatedly play Colored Trails on the same board with always the same set of chips, but with a new randomly drawn goal location at the start of each game;
- *Dynamic environment with dynamic goals* Agents repeatedly play Colored Trails, but each game is played on a new randomly generated board, with randomly drawn chips and with randomly drawn goal locations.

From the perspective of a zero-order theory of mind reasoner, these three settings represent very different environments. In the static environment with static goals, every repeated game gives the zero-order theory of mind agent relevant information about the effectiveness of making certain offers. As a result, in this environment, it is relatively easy to learn the optimal decision to make for a zero-order theory of mind agent.

In the static environment with dynamic goals, negotiation becomes harder for a zero-order theory of mind reasoner. Since agents are randomly assigned a goal locations at the start of each round, the ability of agents to learn from the outcome of a single game is limited. After all, whether the responder will accept a given offer depends on her goal location. However, in the static environment with dynamic goals, the offers that the competitor makes are more relevant to the allocator. In this environment, agents can therefore learn more effectively from the actions of their competitor.

The dynamic environment with dynamic goals represents an environment that is extremely unpredictable. Zero-order theory of mind agents learn little about the effectiveness of making offers from the outcome of a single game. Since every round is played on an entirely new board, agents are very likely to behave differently in every new game. Zero-order theory of mind agents will find it difficult to learn the optimal offer to make in this environment.

2.2 Assumptions in our colored trails setting

While modeling theory of mind in the Colored Trails setting described above, we make use of a number of assumptions. For convenience, these assumptions are listed below.

1. Every player can observe all the characteristics of the game, including the goal location and initial set of chips for each player in the game.
2. The allocator and the competitor make their respective offers simultaneously.
3. Every player can observe the offers that have been made.
4. Players cannot observe the reasoning process of other players.
5. Each player is limited in their ability to reason about the reasoning process of other players (i.e. theory of mind).
6. Players behave rationally with respect to their beliefs.
7. Players make no mistakes in understanding the rules of the game or in performing their desired actions.
8. Players do not consider the possibility that any player would make a mistake in understanding the rules of the games or in executing their desired action.

Note that since we assume that players are limited in their theory of mind ability, these players are unable to achieve *common knowledge of rationality*, in contrast to what is typically assumed of agents in game-theoretic models [33, 63].

3 Theory of mind in colored trails

In this section, we describe the intuition behind the way agents can make use of theory of mind to play Colored Trails. In our setting, the environment is fully observable. However, agents do not know the reasoning capabilities of other players. Instead, agents rely on their theory of mind abilities to predict the behavior of others. Using theory of mind, an allocator can take the perspective of the competitor, and determine what his own decision would have been if the allocator had been in the position of this player. By using his own thought process as a model for the thought process of the competitor, the allocator predicts that the competitor will make the same decision the allocator would have made himself if the roles had been reversed. Due to our assumption that agents behave rationally with respect to their beliefs (see Sect. 2.2), the responder always chooses the option that will yield her the highest possible score. In particular, this means that the responder is not making use of theory of mind. In this section, we therefore describe the theory of mind used by the allocator and the competitor.

In the following subsections, we describe how this process of perspective-taking results in different behavior for agents of different orders of theory of mind playing Colored Trails. We describe this process from the perspective of the allocator. However, the way competitors use theory of mind is completely analogous. The formal description of these theory of mind agents is presented in Sect. 4. In the remainder, we will speak of a ToM_k allocator to indicate an allocator that has the ability to use theory of mind up to and including the kth order, but not beyond.

3.1 Zero-order theory of mind allocator

By convention, a zero-order theory of mind (*ToM* $_0$) allocator is unable to attribute mental content to others. In particular, the *ToM* $_0$ allocator is unable to represent that the competitor and the responder want to reach their respective goal locations, and that the behavior of other agents is consistent with those goals. Although a *ToM* $_0$ allocator cannot make use of theory of mind, such an allocator may make use of an associative learning strategy by constructing zero-order beliefs about the likelihood that a certain offer will be accepted by the responder. The *ToM* $_0$ allocator bases his zero-order beliefs on his observations of the behavior of the responder. For example, through repeated interaction, the *ToM* $_0$ allocator may learn that the responder never accepts an offer that assigns all chips to the allocator himself, and no chips to the responder. Similarly, the *ToM* $_0$ allocator's beliefs may eventually reflect that the responder is more likely to accept offers that assign many chips to her and few to the *ToM* $_0$ allocator, while she is less likely to accept an offer that assigns few chips to the responder and many to the *ToM* $_0$ allocator.

Using these zero-order beliefs, the *ToM* $_0$ allocator can form an expectation about how his score will change if he were to make a particular offer, and select to make the offer that he assigns the highest expected value. This allows the *ToM* $_0$ allocator to play the Colored Trails setting without attributing mental content to others. The zero-order beliefs of the *ToM* $_0$ allocator may eventually reflect that other players have a desire for owning chips, even though the *ToM* $_0$ allocator does not explicitly represent such a desire. Moreover, although the *ToM* $_0$ allocator does not consider the existence of a competitor in his decision process, the zero-order beliefs of a *ToM* $_0$ allocator may eventually reflect information about the behavior of the competitor because of the influence of the offers of the competitor on the behavior of the responder. In addition, the *ToM* $_0$ agent also observes the offers made by the competitor, and can use this information to form beliefs about the likelihood of the responder accepting a certain offer.

3.2 First-order theory of mind allocator

In addition, to the associative learning strategy of the *ToM* $_0$ allocator, a first-order theory of mind (*ToM* $_1$) allocator considers the possibility that other agents have beliefs and goals as well, which determine their behavior. Because of this, a *ToM* $_1$ allocator realizes that in order to get a large piece of pie for himself, his offer should include a large piece of pie for the responder as well. The *ToM* $_1$ allocator is able to consider the game from the perspective of other players, and decide how he would act if he were in the position of that player. The *ToM* $_1$ allocator then considers the possibility that other players in that position may make the same decisions as he would have made himself. That is, by taking the position of another player, a *ToM* $_1$ allocator obtains a prediction of what a *ToM* $_0$ agent might do in that position. This allows the *ToM* $_1$ allocator to make a prediction about the offer that the competitor is going to make, as well as how the responder will choose. The *ToM* $_1$ allocator uses these predictions to decide the offer he should make himself. For example, if a *ToM* $_1$ allocator believes that the competitor is going to make an offer that would increase the score of the responder by 15 points, he also believes that if he were to make an offer that would increase the score of the responder by 20 himself, the responder would certainly choose his offer over the offer made by the competitor.

Note that first-order theory of mind can provide an allocator with a convenient generalization over different games. Even if the ToM_1 allocator finds himself in a novel situation, first-order theory of mind allows the allocator to predict the behavior of other agents. However, although the ToM_1 allocator is able to consider other players as ToM_0 agents, the ToM_1 allocator cannot observe the reasoning process of others (see Sect. 2.2). That is, while a ToM_1 allocator in our setting knows the goal locations of the responder and competitor, the allocator does not know the extent of the reasoning abilities of others. Through repeated interactions, the ToM_1 allocator may learn that his first-order beliefs do not accurately model the behavior of other agents. If this happens, the ToM_1 allocator may choose to play as if he were a ToM_0 allocator and rely on associative learning instead.

3.3 Higher orders of theory of mind agent

Allocators that are capable of using 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. Although responders in our setting do not make use of theory of mind, a higher-order theory of mind allocator can benefit from considering the theory of mind abilities of his competitor. For example, while a ToM_1 allocator believes that the competitor only makes use of zero-order beliefs when making an offer, a ToM_2 allocator considers the possibility that the competitor takes the beliefs and goals of other agents into consideration as well, i.e. that the competitor is a ToM_1 agent.

More generally, for each additional order of theory of mind k, a ToM_k allocator models the competitor as a ToM_{k-1} agent. In our setup, a ToM_k agent is therefore limited in the maximum depth of recursive beliefs he can reason with. For example, while a ToM_3 allocator may believe that the competitor knows that the allocator thinks that the responder prefers blue chips over red chips, a ToM_4 allocator can also consider the possibility that the competitor knows that the allocator believes that the competitor thinks that the responder prefers black chips over purple chips.

Note that if the competitor reasons at $(k-1)$st-order theory of mind, the optimal response of the allocator is to reason at kth-order theory of mind. However, reasoning at kth-order theory of mind may not be optimal when the competitor reasons at any order of theory of mind lower than $k-1$. In this case, the performance of the ToM_k allocator may suffer due to overestimation of the competitor. While agents in our setup know the goal locations of all players, they do not know the extent of the theory of mind abilities of their competitor. Instead, a ToM_k agent forms a hypothesis about the order of theory of mind at which the competitor is currently reasoning by matching the observed behavior of the competitor to behavior predicted by the allocator's theory of mind. This means that a ToM_k allocator may choose to behave as if he were a ToM_n agent ($n < k$) if he believes that this results in more accurate predictions of the competitor's behavior.

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 theory of mind. In this section, we discuss the implementation of computational agents that play according to this intuition. The agent model described in this section is an extension of the theory of mind agent model we previously used to investigate the effectiveness of theory of mind in mixed-motive settings [19]. In our previous work, agents played a sequential negotiation game, where two agents alternated in offering a possible outcome until an agreement was reached. In contrast, the agents in our three-player setting play repeated one-shot negotiation games similar to those presented by Ficici and Pfeffer [24].

In our representation, a Colored Trails game is a tuple $\mathcal{CT} = \langle \mathcal{N}, \mathcal{D}, \pi \rangle$, where:

- $\mathcal{N} = \{a, c, r\}$ is the set of agents, where a is the allocator, c is the competitor, and r is the responder;
- $\mathcal{D} = \mathcal{D}_a \cup \mathcal{D}_c$ is the set of chip distributions that are possible in the game, where $\mathcal{D}_i$ is the set of chip distributions that agent i is allowed to offer to the responder; and
- $\pi = (\pi_a, \pi_c, \pi_r)$ is the set of score functions[3] $\pi_a, \pi_c, \pi_r : \mathcal{D} \to \mathbb{R}$ for allocator a, competitor c, and responder r respectively.

Note that unlike in the alternating offers setting [19], agents may not have the same set of offers they are allowed to make. In addition, agents in our current setting cannot learn the behavior of others through sequential moves in the same game, but also experience no uncertainty about the preferences of other agents.

As in De Weerd et al. [19], this representation focuses on the negotiation aspect of the game, and ignores the task of finding routes between locations. This means that, as mentioned in Sect. 2.2, we assume that agents make no mistakes in finding routes between locations. Instead, the score functions π_i specify the maximum score agent i can achieve given some distribution of chips and given the scoring rules outlined in Sect. 2. However, this does not mean that agents achieve common knowledge about the rules of the game. Rather, agents follow the rules of the game and do not consider the possibility that others would break those rules.

Each game involves two stages. First, the allocator and the competitor simultaneously choose a distribution of chips, $D_a \in \mathcal{D}_a$ and $D_c \in \mathcal{D}_c$ respectively, to offer to the responder. The allocator can make any offer that involves the chips in his own initial set of chips and the initial set of chips of the responder, but leaves the set of chips assigned to the competitor unchanged, and vice versa. Next, the responder chooses at most one of these distributions D_a, D_c to become the final distribution of chips.

Without loss of generality, we assume that the score of each agent in the initial distribution of chips is zero. That is, $\pi_i(D)$ denotes the change in score of agent i if the distribution $D \in \mathcal{D}$ becomes final. This way, the score function π_i for agent i summarizes the game board, as well as the start location, the goal location l_i, and the initial set of chips of agent i. That is, the score function π_i represents all observable features at the start of the game, so that the score of agent i is only dependent on the set of chips in possession of agent i at the end of the game.

4.1 Model of zero-order theory of mind

The decision model for zero-order theory of mind agents we use is identical to the one described in De Weerd et al. [19]. Our *ToM* $_0$ allocator does not form explicit beliefs

[3] Note that this specification of the score function is more general that the board setting presented in Sect. 2. The model presented here can be used for arbitrary preference functions over single-shot negotiation outcomes.

about the mental content of others. Instead, a ToM_0 allocator constructs zero-order beliefs $b^{(0)} : \mathcal{D} \rightarrow [0,1]$, which show that the ToM_0 allocator believes that the probability that a given offer $O \in \mathcal{D}$ will be accepted by the responder is $b^{(0)}(O)$. Over repeated games, a ToM_0 allocator updates these zero-order beliefs to learn which offers O are more likely to be accepted than others. The details of this belief updating are described in Sect. 4.4 below.

Given the ToM_0 allocator's zero-order beliefs $b^{(0)}$, the ToM_0 allocator can calculate the expected change in score as a result of making a given offer O through

$$EV_a^0(O;b^{(0)}) = b^{(0)}(O) \cdot \pi_a(O). \tag{1}$$

As mentioned in Sect. 2.2, we assume that agents act rationally with respect to their beliefs. This means that the ToM_0 allocator chooses to make an offer that maximizes this expected value. We define the set $\mathcal{D}_a^0(b^{(0)}) \subseteq \mathcal{D}_a$ to specify which offers the ToM_0 allocator a believes to maximize his expected value, based on his zero-order beliefs $b^{(0)}$. That is,

$$\mathcal{D}_a^0(b^{(0)}) = \left\{ O \in \mathcal{D}_a \middle| EV_a^0(O;b^{(0)}) = \max_{O' \in \mathcal{D}_a} EV_a^0(O';b^{(0)}) \right\}. \tag{2}$$

Since a ToM_0 allocator a believes that any offer in $\mathcal{D}_a^0(b^{(0)})$ will maximize the expected value, he randomly selects one of the offers in $\mathcal{D}_a^0(b^{(0)})$ to make to the responder. The zero-order theory of mind allocator can therefore be implemented as illustrated by the pseudo-code presented in Algorithm 1.

Algorithm 1 Zero-order theory of mind allocator

```
function EXPECTEDVALUE(offer)
    return $b^{(0)}$[offer] $\times$ $\pi_a$[offer]
end function

function SELECTOFFERS
    offerValues ← list(EXPECTEDVALUE(offer) for each offer in $\mathcal{D}_a$)
    maxValue ← MAX(offerValues)
    return list(offer for each offer in $\mathcal{D}_a$ where EXPECTEDVALUE(offer) = maxValue)
end function

function MAKEOFFER
    acceptableOffers ← SELECTOFFERS
    return acceptableOffers[RANDOMINTEGER(sizeof acceptableOffers)]
end function
```

Example 2 Figure 2 shows an example of a Colored Trails game, in which allocator a is a ToM_0 agent that wants to move from the initial location at the center of the board to goal location l_a in the top left corner. With his initial set of chips, allocator a is left three steps short of his goal with three spare chips, which would yield him a score of -15 points. To decide which offer to make, the ToM_0 allocator calculates the expected value of making each possible offer, and randomly makes one of the offers that maximizes the expected value.

Table 1 shows a number of selected offers that allocator a could make to responder r. For each offer O, the table shows how offer O affects the score of both allocator a and responder r, and the value of the zero-order beliefs $b^{(0)}(O)$ of allocator a, which show what

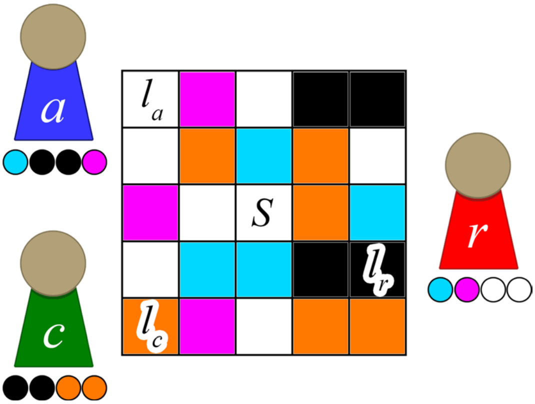

Fig. 2 In Example 2, *ToM* $_0$ allocator a competes with competitor c for the opportunity to trade with responder r. With his initial set of chips, allocator a can only move one step towards his goal location l_a, which would yield a score of 25 points

Table 1 Selected offers that *ToM* $_0$ allocator a considers making to responder r in Example 2. The numbers in parentheses indicate the effect on the scores of allocator a and responder r if one of these offers were to be accepted. Note that in these cases, the score and set of chips owned by competitor c remain unchanged

Offer	Allocator a	Responder r	$b^{(0)}(O)$	$EV_a^0(O; b^{(0)})$
O_1	🔵🔵🟣⚪⚪ (+70 points)	⚫⚫🟣 (-10 points)	0.123	8.61
O_2	🔵🟣🟣⚪⚪ (+70 points)	🔵⚫⚫ (+55 points)	0.123	8.61
O_3	🔵⚫🟣⚪⚪ (+70 points)	🔵⚫🟣 (+0 points)	0.117	8.19
O_4	🔵🟣⚪⚪ (+65 points)	🔵⚫⚫🟣 (+60 points)	0.120	7.80

allocator a believes to be the probability that responder r will accept offer O. Note that a *ToM* $_0$ agent does not consider the goal of the responder. From the perspective of *ToM* $_0$ allocator a, offers O_1 and O_2 are therefore indistinguishable. Both these offers allow allocator a to reach his goal location (50 points) with one chip to spare (5 points), increasing his score from −15 to 55 points (+70 points). Furthermore, allocator a believes that the probability that offers O_1 and O_2 will be accepted is the same. In contrast, although offer O_3 would yield allocator a the same increase in score as offers O_1 and O_2, previous experience leads allocator a to believe that offer O_3 is less likely to be accepted by the responder, as shown by the lower value of the agent's zero-order beliefs in Table 1. As a result, the expected value that allocator a assigns to offer O_3 is lower than the expected value assigned to offers O_1 and O_2.

Note that *ToM* $_0$ allocator a could ask the responder to give him all her chips. Although this offer has the potential to increase the score of allocator a to 70 (+85 points), this offer does not appear in Table 1. Allocator a does not consider making this offer because the allocator has learned that the responder would not accept this offer through earlier interactions.

In this example, both offer O_1 and offer O_2 maximize the expected value for allocator a. The set $\mathcal{D}_a^0(b^{(0)}) = \{O_1, O_2\}$ therefore contains two elements. Allocator a randomly chooses one of these offers to make. However, although allocator a considers offer O_1 and offer O_2 to be equally acceptable, Table 1 shows that offer O_1 would decrease the score of responder r, while offer O_2 would increase her score. That is, responder r would in fact reject offer O_1, but she would accept offer O_2.

4.2 Model of first-order theory of mind

Due to the difference in setting, our ToM_1 agent model deviates strongly from the model presented in De Weerd et al. [19]. Our ToM_1 allocator takes the score of responder r into account when making his decision. In our setting, the responder simply accepts the offer that maximizes her score. That is, the responder is fixed at a zero-order theory of mind strategy. When a ToM_1 allocator a considers making offer $O \in \mathcal{D}_a$, the agent correctly ascribes zero-order theory of mind to the responder and believes that she will not accept this offer if it would decrease her score, or $\pi_r(O) < 0$. Furthermore, the responder would not accept offer O made by the ToM_1 allocator a if competitor c makes an offer $O' \in \mathcal{D}_c$ that would yield her a higher score. Finally, if the offer O of the ToM_1 allocator and offer O' of the competitor would yield the responder the same score, the responder will randomly accept one of the two offers.

This information about the likelihood that an offer O will be accepted by the responder is summarized in the function p. A ToM_1 allocator a who predicts that his competitor c randomly selects an offer from the set $\mathcal{D}$, believes that if he makes the offer $O \in \mathcal{D}_a$ himself, this offer O will be accepted by the responder with probability

$$p(O, \mathcal{D}) = \begin{cases} 0 & \text{if } \pi_r(O) \leq 0 \\ \frac{1}{|\mathcal{D}|}\left(\sum\limits_{\substack{O' \in \mathcal{D} \\ \pi_r(O) > \pi_r(O')}} 1 + \sum\limits_{\substack{O' \in \mathcal{D} \\ \pi_r(O) = \pi_r(O')}} \frac{1}{2} \right) & \text{if } \pi_r(O) > 0 \end{cases}. \tag{3}$$

Note that Eq. (3) describes the behavior of a rational responder r. By attributing rationality to responder r, an allocator a considers the goals of the responder explicitly, and thus engages in first-order theory of mind.

To predict what offer competitor c will be making, ToM_1 allocator a considers the game from the perspective of the competitor. By calculating what offers allocator a might make himself if he were facing the situation competitor c is in, allocator a obtains a prediction of what offers competitor c may make. To do so, the ToM_1 allocator constructs first-order beliefs $b^{(1)} : \mathcal{D} \rightarrow [0, 1]$, which specify what the ToM_1 allocator a believes his zero-order beliefs to have been, if he had been in the position of competitor c. Note that these beliefs do not necessarily reflect the actual beliefs of competitor c. Rather, an agent's first-order beliefs $b^{(1)}$ represent the agent's best guess for the zero-order beliefs $b^{(0)}$ of his competitor. Based on his first-order theory of mind, the expected value that ToM_1 allocator a assigns to making a given offer $O \in \mathcal{D}$ therefore is

$$p(O, \mathcal{D}_c^0(b^{(1)})) \cdot \pi_c(O). \tag{4}$$

Equation (4) represents that allocator a calculates his own optimal decision $\mathcal{D}_c^0(b^{(1)})$ from the perspective of competitor c. That is, allocator a determines what offer he would have made himself if he had been assigned the goal location and initial set of chips that were assigned to competitor c.

Although Eq. (4) describes the way a ToM_1 allocator uses first-order theory of mind, it does not fully describe the behavior of a ToM_1 allocator. Since agents cannot observe the reasoning process of others (Sect. 2.2), agents are not aware of the level of sophistication of others. A ToM_1 allocator has the ability to make use of first-order theory of mind, but such a ToM_1 allocator may come to believe that his first-order theory of mind

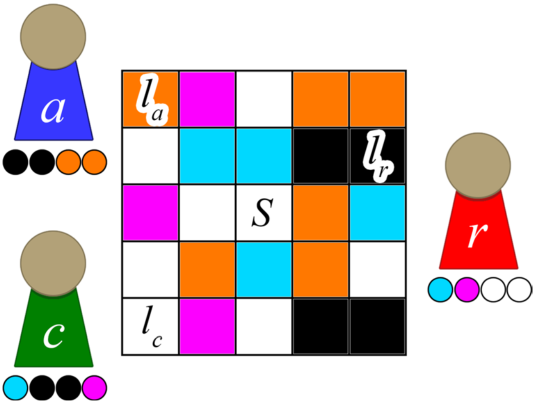

Fig. 3 The *ToM* $_1$ allocator *a* in Example 3 does not only consider his own score, but also considers the score of the responder *r* when deciding what offer to make

does not accurately predict the behavior of the other agents, and that the ToM_1 allocator would be better off following his zero-order beliefs. To this end, the ToM_1 allocator has a confidence variable $c_1 \in [0,1]$, which indicates how much confidence the ToM_1 allocator places on first-order theory of mind over zero-order theory of mind. When deciding on the expected value of making an offer O, the ToM_1 allocator weights the predictions of first-order and zero-order theory of mind accordingly.

In summary, a ToM_1 allocator a calculates the expected value of making a given offer $O \in \mathcal{D}$ through

$$EV_a^1(O;b^{(0)},b^{(1)},c_1) = (1-c_1)\cdot EV_a^0(O;b^{(0)}) + c_1 \cdot p\big(O,\mathcal{D}_c^0(b^{(1)})\big)\cdot \pi_a(O). \quad (5)$$

Similar to the decision process of a *ToM* $_0$ allocator, the *ToM* $_1$ allocator *a* randomly chooses to make one of the offers that maximize the expected value. The set of these offers is given by

$$\mathcal{D}_a^1(b^{(0)},b^{(1)},c_1) = \left\{ O \in \mathcal{D}_a \middle| EV_a^1(O;b^{(0)},b^{(1)},c_1) = \max_{O' \in D} EV_a^1(O';b^{(0)},b^{(1)},c_1) \right\}. \quad (6)$$

Example 3 In this example, we consider the Colored Trails setting depicted in Fig. 3. Note that this setting is similar to the one in Example 2, but with the roles of the allocator and the competitor switched. Hence the board is flipped, and allocator *a* and competitor *c* have switched their initial set of chips. This way, *ToM* $_1$ allocator *a* attributes the reasoning outlined in Example 2 to competitor *c*. That is, *ToM* $_1$ allocator *a* can reason as outlined in Example 2, and make a prediction about the offer competitor *c* is going to make. In this example, we assume that allocator *a* is a *ToM* $_1$ agent that correctly believes that competitor *c* will randomly choose an offer from the set $\mathcal{D}_c^0(b^{(1)}) = \{O_1, O_2\}$.

Since allocator *a* is a *ToM* $_1$ agent, he knows that the responder will not accept his offer *O* if offer *O* would decrease her score, or if competitor *c* makes an offer that would result in a larger increase in her score. Table 2 shows a summary of selected offers that allocator *a* can make. Each of these offers allows both allocator *a* and responder *r* to reach their respective goal locations.

Table 2 Selected offers that *ToM* $_1$ allocator a considers making to responder r in Example 3. The effect of each offer on the scores of allocator a and responder r are given in parentheses

Offer	Allocator a	Responder r	$EV_a^1(O; b^{(0)}, b^{(1)}, 1)$
O_5	🔵🟠🟣○○ (+75 points)	●●🟠 (+55 points)	56.25
O_6	●🟠🟣○○ (+75 points)	🔵●🟠 (+55 points)	56.25
O_7	🟠🟠🟣○○ (+75 points)	🔵●● (+55 points)	56.25
O_8	🔵🟠🟣○ (+70 points)	●●🟠○ (+60 points)	70
O_9	🔵🟠○○ (+70 points)	●●🟠🟣 (+60 points)	70
O_{10}	🟠🟣○○ (+70 points)	🔵●●🟠 (+60 points)	70

Example 2 shows that if competitor c is a *ToM* $_0$ agent, there is a 50% probability that he will make offer O_1, which would be rejected by responder r. However, there is also a 50% probability that competitor c will make offer O_2, which would increase the responder's score by 55 points. If allocator a were to make an offer that increases the score of responder r by 55 points as well, the responder would randomly accept either the offer of allocator a or of competitor c. As a result, allocator a believes that if he were to make an offer that increases the score of the responder by 55 points, there is a 25% probability ($0.5 \cdot 0.5$) that the responder will accept offer O_2 of competitor c instead of his own offer. Based purely on first-order theory of mind, allocator a would therefore assign an expected value of $(1 - 0.25) \cdot 75 = 56.25$ to each of the offers O_5, O_6 and O_7. On the other hand, allocator a is certain that responder would accept any of the offers O_8, O_9 and O_{10}, which allocator a therefore assigns an expected value of 70. These offers also maximize the expected value according to allocator a.

Based on his first-order theory of mind, allocator a would randomly select one of the offers O_8, O_9 and O_{10} to propose to the responder. Depending on his confidence c_1 in first-order theory of mind, allocator a may still decide to select a different offer entirely.

4.3 Model of higher-order theory of mind

For increasingly higher orders of theory of mind, a *ToM* $_k$ allocator considers the possibility that the competitor is increasingly more sophisticated. For example, a *ToM* $_2$ allocator a believes that the competitor c may take into consideration that *ToM* $_2$ allocator a wants to reach his goal. By placing himself in the position of competitor c, a *ToM* $_2$ allocator's second-order theory of mind may predict different behavior for the competitor than predicted by his first-order theory of mind. For each additional order of theory of mind, the *ToM* $_k$ allocator constructs additional beliefs $b^{(k)}$ and confidence c_k. For example, the second-order beliefs $b^{(2)}$ of *ToM* $_2$ allocator a specify what allocator a believes competitor c to believe to be the zero-order beliefs of allocator a.

The expected value that a *ToM* $_k$ allocator a assigns to making an offer $O \in \mathcal{D}_a$ is defined recursively on the equations defined earlier, so that

$$\begin{aligned} EV_a^k(O;b^{(0)},\dots,b^{(k)},c_1,\dots,c_k) &= (1-c_k)EV_a^{k-1}(O;b^{(0)},\dots,b^{(k-1)},c_1,\dots,c_{k-1}) \\ &\quad + c_k \cdot p\big(O, \mathcal{D}_c^{k-1}(b^{(1)},\dots,b^{(k)},1,0,\dots,0)\big) \cdot \pi_a(O). \end{aligned} \tag{7}$$

Note that in the equation above, the *ToM* $_k$ allocator does not attempt to model the confidence in theory of mind $c_1, \dots, c_{k-1}$ of the competitor. Instead, the *ToM* $_k$ allocator models a *ToM* $_{k-1}$ competitor that decides purely on the basis of predictions made by his $(k-1)$st theory of mind. This ensures that a *ToM* $_k$ allocator is always able to model a *ToM* $_{k-1}$ competitor.

Higher orders of theory of mind do not change what the allocator predicts the responder will do. Similar to the decision process of *ToM* $_0$ and *ToM* $_1$ allocators, the *ToM* $_k$ allocator randomly chooses to make an offer from the set of offers that maximize his expected value. This set of offers is given by

$$\mathcal{D}_a^k(b^{(0)},\dots,b^{(k)},c_1,\dots,c_k) = \left\{ O \in \mathcal{D}_a \middle| EV_a^k(O,b^{(0)},\dots,b^{(k)},c_1,\dots,c_k) = \max_{O'\in D} EV_a^k(O',b^{(0)},\dots,b^{(k)},c_1,\dots,c_k) \right\}. \tag{8}$$

Example 4 In this example, we consider the Colored Trails setting depicted in Fig. 2. Note that this setting is the same as in Example 2. This also means that compared to Example 3, the roles of the allocator and the competitor are switched again, so that the board is flipped, and allocator a and competitor c have switched their initial set of chips. This way, *ToM* $_2$ allocator a attributes the reasoning outlined in Example 3 to competitor c.

In this example, we assume that allocator a is a *ToM* $_2$ agent who correctly believes that competitor c will randomly choose an offer from the set $\mathcal{D}_c^1(b^{(2)}, b^{(1)}, 1) = \{O_8, O_9, O_{10}\}$. That is, the allocator believes that the competitor will offer the responder enough chips to reach her goal location with one chip to spare (i.e. increase her score by 60). Note that the *ToM* $_2$ allocator may choose to make offer O_4 (see Table 1), which would increase the score of responder r by 60 points and increase the score of allocator a by 65 points. Allocator a believes that in this case, responder r would be indifferent between the offers of the allocator and the competitor, and that she would therefore accept the allocator's offer with a probability of 50%. The expected value is therefore $EV_a^2(O_4; b^{(2)}, b^{(1)}, b^{(0)}, 1, 0) = 0.5 \cdot 65 = 32.5$.

Alternatively, *ToM* $_2$ allocator a could offer responder r an additional chip. According to allocator a, doing so would guarantee that responder r would accept his offer over the offer of competitor c. After all, this offer would increase the score of the responder by 65 points, while allocator a believes that competitor c will make an offer that would only increase her score by 60 points. However, by leaving himself with only three chips, allocator a would not be able to reach his goal location. At most, he would be able to increase his score by 5 points. That is, the expected value of such an offer O_{11} would be $EV_a^2(O_{11}; b^{(2)}, b^{(1)}, b^{(0)}, 1, 0) = 5$.

Since *ToM* $_2$ allocator a makes the offer with the highest expected value, based on his second-order theory of mind, allocator a decides to make offer O_4. Depending on his confidences c_2 and c_1 in second-order and first-order theory of mind respectively, allocator a may still decide to select a different offer entirely.

4.4 Learning across games

As described above, theory of mind agents construct beliefs about the behavior of others. These beliefs are based on observations of the behavior of others over repeated Colored Trails games. In this section, we describe how this process of belief adjustment occurs.

In this paper, agents construct zero-order beliefs based on observable features of the environment. In the case of our Colored Trails setting, this includes the color of each tile on the game board, the initial sets of chips of each of the agents, and the goal location of the responder. Note that this includes all observable features of the environment, except for the goal locations of the allocator and the competitor. The agent's zero-order beliefs $b^{(0)}$ are calculated as the observed frequency with which offers O has been accepted by the responder in the past, given that the observable features of the environment are the same. This observed frequency is based on the offers made by both the allocator and the competitor.

In a static environment with static goals, all observable characteristics of the environment stay the same. As a result, if a ToM_0 allocator has observed 250 instances in which the offer O has been made, 220 of which have been accepted by the responder, the allocator believes that the probability $b^{(0)}(O)$ that the responder will accept offer O again will be 0.88.

In a static environment with dynamic goals, only the goal locations of the agents change with every repetition of the game. As a result, the ToM_0 agent holds different zero-order beliefs $b^{(0)}$ for each of the 12 possible goal locations of the responder.

In the dynamic environment with dynamic goals, an agent holds different zero-order beliefs for each possible combination of game boards, goal locations, and initial sets of chips. Note, however, that each game is played on a randomly generated 5 by 5 board, where each square is randomly assigned one of five possible colors, resulting in 5^{25} possible game boards. This means that it is unlikely for an agent to encounter the same scenario twice over the course of an experiment. As a result, an agent is in practice unable to use any previous experience with the Colored Trails game to inform his zero-order beliefs.

In addition to updating their zero-order belief $b^{(0)}$, theory of mind agents update their kth-order beliefs the same way as their zero-order beliefs. In our model, an allocator's first-order beliefs $b^{(1)}$ represent what the allocator believes to be the zero-order beliefs of the competitor. Similarly, an allocator's second-order beliefs $b^{(2)}$ represent what the allocator believes the competitor to believe to be the zero-order beliefs of the allocator. Zero-order beliefs record the responses to offers made to the responder, which are independent of who made the offer. In addition, there is no private information in this setting. Because of this, an allocator's zero-order beliefs $b^{(0)}$ are identical to that of his competitor. Moreover, since the allocator knows this, this means that in our Colored Trails setup, a ToM_1 agent's zero-order beliefs $b^{(0)}$ are the same as his first-order beliefs $b^{(1)}$. In fact, since all allocators and competitors construct their beliefs in the same way, the beliefs of these agents are all identical.

In addition to updating his beliefs, a ToM_1 allocator also updates his confidence c_1 in first-order theory of mind after observing the outcome of a game. The ToM_1 allocator does so based on how accurately his first-order theory of mind predicts the behavior of the competitor.

After every game, the ToM_1 allocator observes the offer O_c made by the competitor c. Following first-order theory of mind, the ToM_1 allocator believes that the competitor assigns an expected value $EV_c^0(O_c;b^{(1)})$ to making this offer. Note that this expected value uses the agent's first-order beliefs $b^{(1)}$ about the zero-order beliefs of the competitor rather than his own zero-order beliefs $b^{(0)}$. Moreover, first-order theory of mind also predicts what the maximum expected value is that the competitor assigns to any one offer. The ToM_1 allocator judges the accuracy his first-order theory of mind by comparing these two values. The closer these values are to one another, the more accurate the prediction of first-order theory of mind. To reflect this, the ToM_1 allocator adjusts his confidence c_1 in first-order theory of mind according to

$$c_1 = (1 - \lambda) \cdot c_1 + \lambda \cdot \frac{EV^0(O_c; b^{(1)})}{\max_{O \in \mathcal{D}_c} EV^0(O; b^{(1)})}. \tag{9}$$

This update, based on adaptive expectations [9, 28], ensures that the confidence c_1 remains in the range [0, 1]. The agent-specific learning speed $\lambda \in [0, 1]$ determines how quickly an allocator changes his opinion about the theory of mind abilities of the competitor.

For higher orders of theory of mind, an allocator adjusts his confidence c_k in kth-order theory of mind analogously. The general formula for the adjustment of an allocator's confidence c_k in kth-order theory of mind after observing that competitor c has made offer O_c is given by

$$c_k = (1 - \lambda) \cdot c_k + \lambda \cdot \frac{EV^{k-1}(O_c; b^{(1)}, \dots, b^{(k)}, 1, 0, \dots, 0)}{\max_{O \in \mathcal{D}_c} EV^{k-1}(O; b^{(1)}, \dots, b^{(k)}, 1, 0, \dots, 0)}, \tag{10}$$

where $\lambda \in [0, 1]$ is an agent-specific learning speed. The theory of mind agents we describe adjust their confidence in theory of mind through adaptive expectations rather than Bayesian updating. This is because agents model the order of theory of mind at which the competitor is reasoning while the competitor is doing the same. That is, although a ToM_k agent cannot use orders of theory of mind beyond kth-order theory of mind, the order of theory of mind at which the agent is reasoning is not fixed, but may change over time.

5 Results of simulation experiments

We implemented the theory of mind agents described in Sect. 3 in Java[4] and performed simulations in which agents played repeated one-shot Colored Trails games. The reason for using an agent-based model is that, apart from very specific instances, the behavior of the agents in our setting is poorly described as an equilibrium. Allocators continuously adjust their behavior to the actions of their competitor, while the competitor adjusts his behavior to the actions of the allocator. Using agent-based modeling, we can analyze how the interaction between these complex decision-making processes plays out.

In each simulation, groups consisting of an allocator agent, a competitor agent, and a responder agent played consecutive negotiation games. At the start of each simulation, all beliefs and confidences of allocator and competitor were set to 1. This means that a ToM_0 agent initially believes that any offer will be accepted by the responder, including an "offer" that consists of requesting the responder's full set of chips. Since this initialization results in particularly poor performance for zero-order theory of mind agents, the first 1000 games were considered to be a setup phase for the zero-order theory of mind agent and were not included in our analysis. After the lead time of 1000 games, the allocator, competitor, and responder played one experimental game.

Furthermore, to account for the effect of different game settings, the results were averaged over 5000 runs.[5] To increase the likelihood that agents had an incentive to negotiate to increase their score, games in which some agent could reach his or her goal location

[4] The Java sources of our implementation are available from the OpenABM Computational Model library at https://www.comses.net/users/2498/.

[5] We averaged over 5,000 runs to ensure that a 1 point difference in average allocator score would result in a significant difference, using a conservative estimate for the standard deviation of allocator scores of 35. This allows us to focus on the relevance of our results, rather than on their statistical significance.

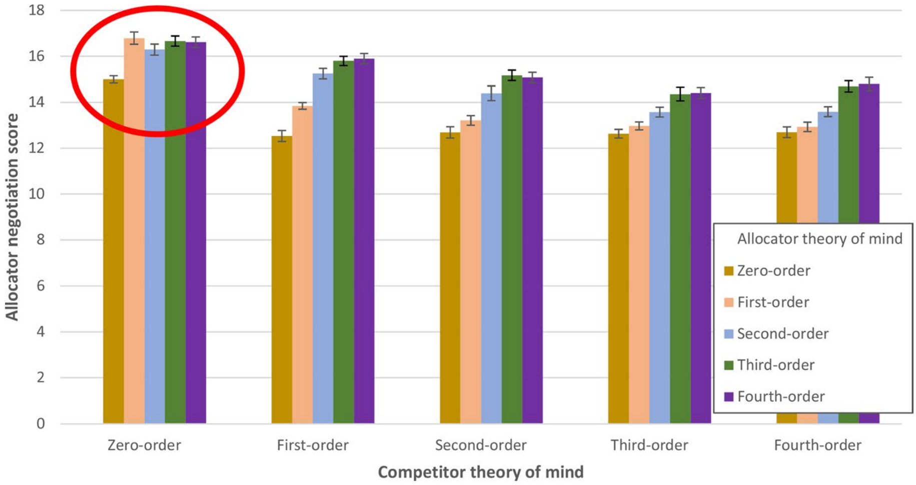

Fig. 4 Average negotiation score of an allocator in the static Colored Trails game with static goals. Results are shown for different combinations of orders of theory of mind of both the allocator and the competitor. Brackets indicate standard errors of the results

with the initial set of chips without any trading were excluded from analysis. Due to this rule, approximately 40% of the generated scenarios were rejected.

In the subsections below, we consider the results for the different environments described in Sect. 2:

- Static environment with static goals
- Static environment with dynamic goals
- Dynamic environment with dynamic goals

For each of these environments, we recorded the average performance of a ToM_i allocator in the presence of a ToM_j competitor ($i, j \in \{0, 1, 2, 3, 4\}$), which is calculated as the average difference between the allocator's final score after negotiation ended and his initial score at the start of negotiation. For the simulations, the learning speed λ was set to 0.1 for both allocator and competitor. Additional simulations with different values for the learning speed showed similar results as the ones reported here.

In the following subsections, we present the results for each of the three different environments separately. Differences in average scores are tested using two-sample *t*-tests, assuming unequal variances, at the 5% significance level. In Figs. 4, 5, 6, 7, 8, 9, 10, 11 and 12, error brackets indicate one standard error. Due to the high number of replications, the standard errors are much smaller than the standard deviations, which were typically around 25 points for the allocator. These standard deviations are rather high compared to the average negotiation scores of an allocator since the responder can only trade with either the allocator or the competitor, but not with both. An allocator is therefore expected to

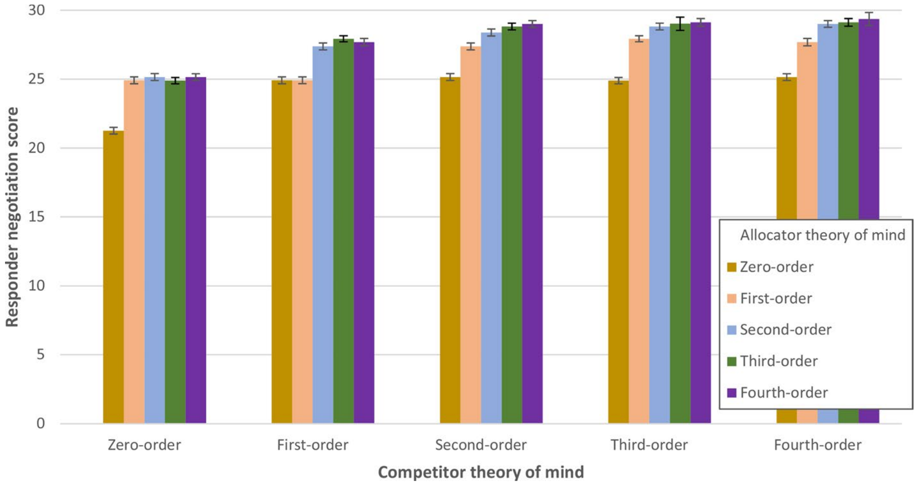

Fig. 5 Average negotiation score of the responder in the static Colored Trails game with static goals. Results are shown for different combinations of orders of theory of mind of both the allocator and the competitor agent. Brackets indicate standard errors of the results

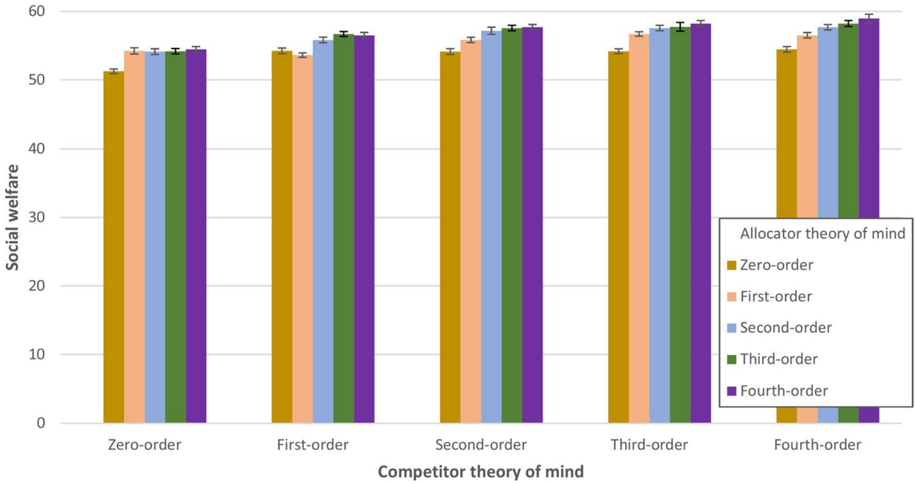

Fig. 6 Average social welfare in the static Colored Trails game with static goals. Results are shown for different combinations of orders of theory of mind of both the allocator and the competitor agent. Brackets indicate standard errors of the results

receive a negotiation score of 0 in half the negotiations. Whenever the error brackets of two individual bars do not overlap, the difference is statistically significant at the 5% level. For the outcomes of sensitivity analysis on several key parameters, see the Supplementary Information.

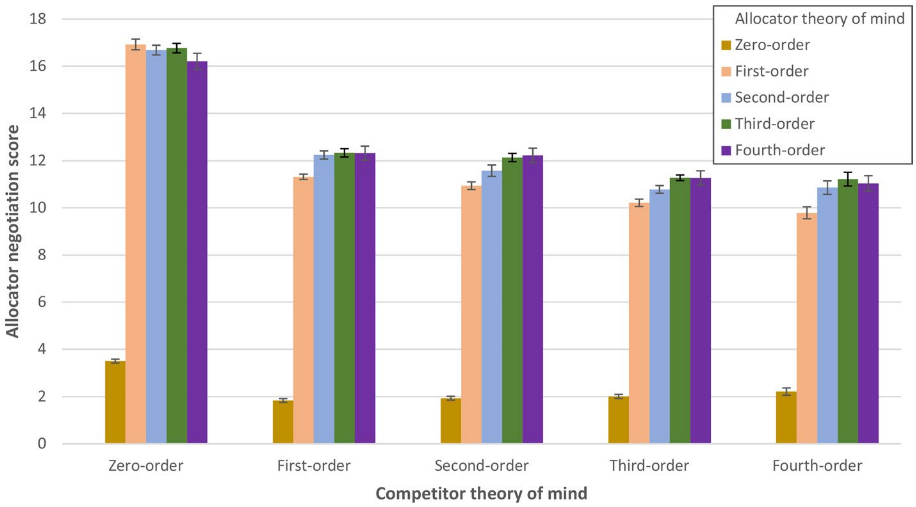

Fig. 7 Average negotiation score of an allocator in the static Colored Trails game with dynamic goals. Results are shown for different combinations of orders of theory of mind of both the allocator and the competitor. Brackets indicate standard errors of the results

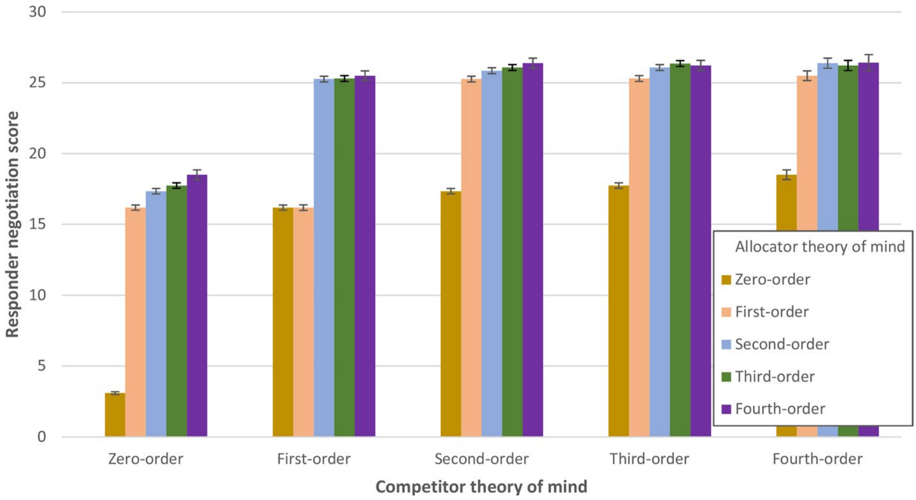

Fig. 8 Average negotiation score of the responder in the static Colored Trails game with dynamic goals. Results are shown for different combinations of orders of theory of mind of both the allocator and the competitor agent. Brackets indicate standard errors of the results

5.1 Static environment with static goals

In the static environment with static goals, agents played each of the 1001 games in a run on the same board with the same initial sets of chips and goal locations for each agent. This means that each observation an allocator makes is in the exact same scenario. Recall that allocators observe the reaction of the responder to their own offer as well as to the offer of

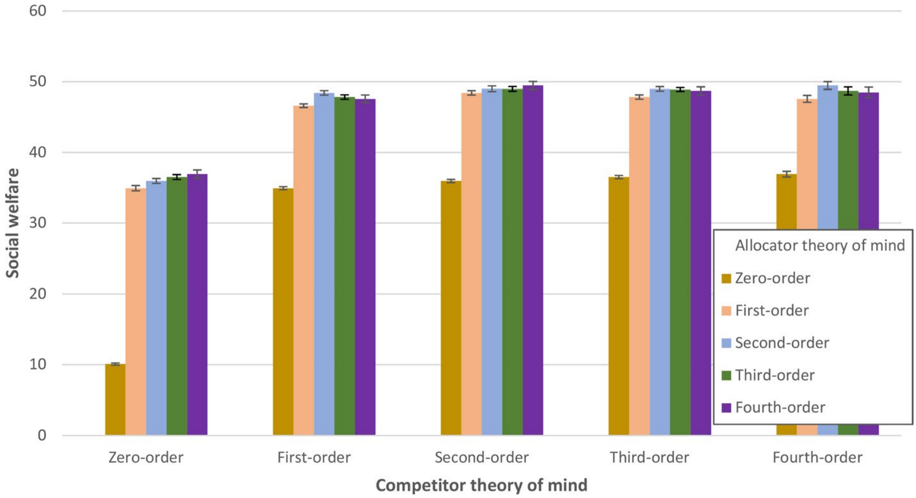

Fig. 9 Average social welfare in the static Colored Trails game with dynamic goals. Results are shown for different combinations of orders of theory of mind of both the allocator and the competitor agent. Brackets indicate standard errors of the results

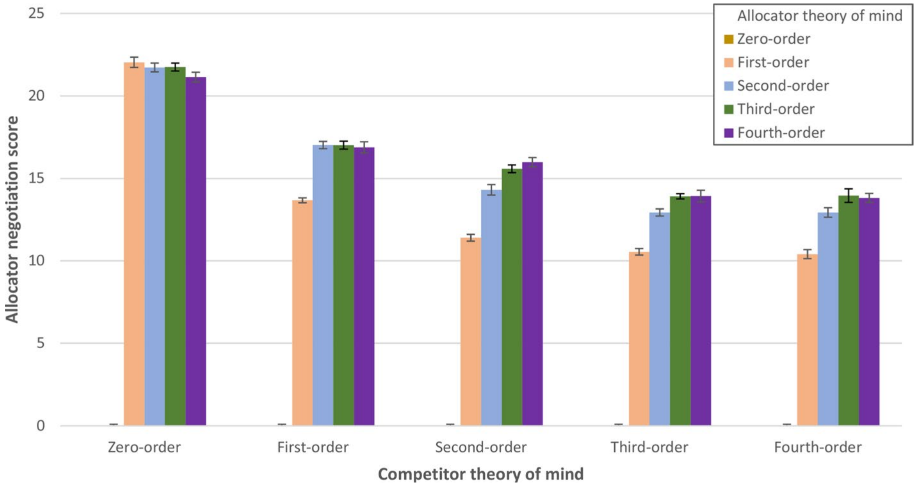

Fig. 10 Average negotiation score of an allocator in the dynamic Colored Trails game with dynamic goals. Results are shown for different combinations of orders of theory of mind of both the allocator and the competitor. Brackets indicate standard errors of the results. Note the extremely low bars for ToM_0 allocators, representing a negotiation score of zero

the competitor. At the end of each game, an allocator therefore makes two observations. However, since the competitor will generally need different chips to get to his goal location than the allocator, the offer made by the competitor is not likely one that the allocator would want to make. That is, although the allocator technically makes two observations of the responder's behavior per round of play, it is likely that only one of these observations is actually useful to the ToM_0 allocator.

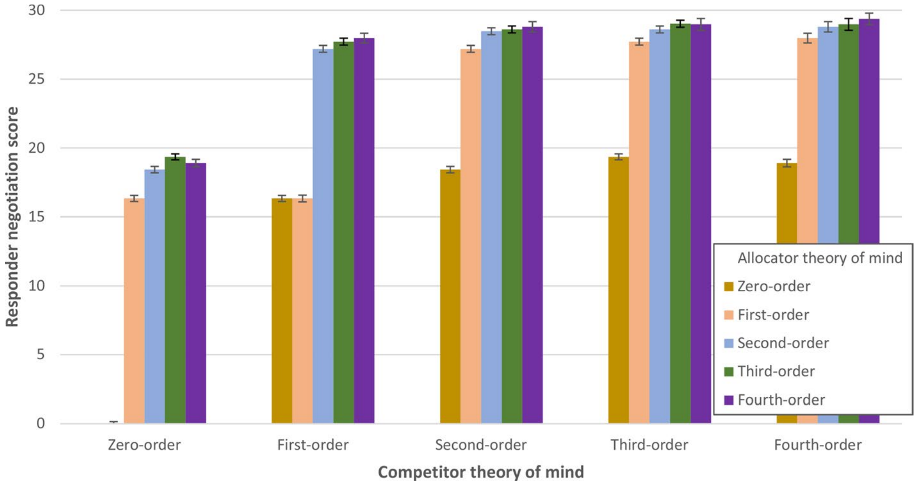

Fig. 11 Average negotiation score of the responder in the dynamic Colored Trails game with dynamic goals. Results are shown for different combinations of orders of theory of mind of both the allocator and the competitor agent. Brackets indicate standard errors of the results. Note the extremely low first bar, representing a negotiation score of zero

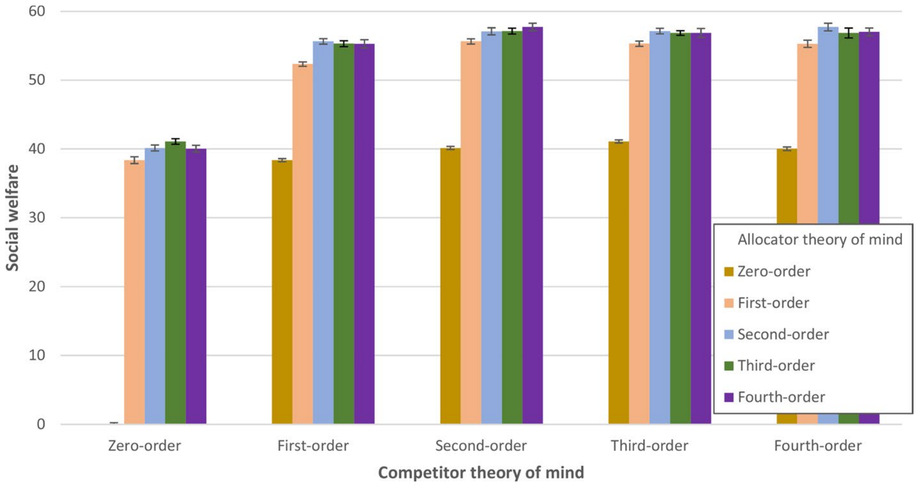

Fig. 12 Average social welfare in the dynamic Colored Trails game with dynamic goals. Results are shown for different combinations of orders of theory of mind of both the allocator and the competitor agent. Brackets indicate standard errors of the results. Note the extremely low first bar, representing a social welfare score of zero

Figure 4 shows the increase in score due to negotiation of a ToM_i allocator in the presence of a ToM_j competitor in the final game of a run, averaged over 5000 runs. Each bar shows the average difference between the score of the allocator before and after negotiation as a function of the theory of mind abilities of both himself and the competitor. Each group of bars corresponds to a particular theory of mind level of the competitor. Within each group, separate bars show the increase in score due to negotiation that allocators of different orders of theory of mind achieve, given the order of theory of mind of the competitor.

For example, the first group of bars (highlighted by the red circle) shows the average negotiation score of allocators that play against a ToM_0 competitor. The figure shows that while a ToM_0 allocator is capable of negotiating for an average 14.8 point increase, a ToM_1 allocator would increase his score by 16.8 points on average by negotiating. When playing against a ToM_0 competitor, allocators do not obtain additional benefits for reasoning at orders of theory of mind higher than the first. In Fig. 4, this is shown by the overlapping brackets of the bars associated with the ToM_1, ToM_2, ToM_3, and ToM_4 allocators.

Since the competitor and the allocator perform similar roles in our setting, Fig. 4 also shows the score of the competitor. A ToM_1 competitor playing against a ToM_0 allocator would be able to obtain the same average negotiation score of 16.8 points as a ToM_1 allocator playing a ToM_0 competitor is able to achieve. Figure 4 therefore indicates how an agent's competitive success varies with his theory of mind abilities by showing how large a piece of pie agents of different orders of theory of mind are able to negotiate for themselves.

Although the ToM_0 allocator is unable to reason about the goals and beliefs of the responder and the competitor, Fig. 4 shows that the ToM_0 allocator is still able to negotiate effectively. Even when his competitor is more sophisticated and can make use of theory of mind, the ToM_0 allocator can negotiate a significant increase in his score on average. Despite the effectiveness of the ToM_0 allocator, Fig. 4 shows that for each group of bars, the bar of the ToM_1 allocator is higher than the bar of the ToM_0 allocator. That is, ToM_1 allocators obtain a higher negotiation score than ToM_0 allocators, irrespective of the theory of mind abilities of the competitor. Where the error brackets of the two bars do not overlap, the difference is statistically significant.

Figure 4 also shows that the bar corresponding to the ToM_2 allocator is higher than the bar corresponding to the ToM_1 allocator next to it for each group of bars, except for the first group of bars. That is, ToM_2 allocators outperform ToM_1 allocators when the competitor makes use of theory of mind. Similarly, ToM_3 allocators outperform ToM_2 allocators whenever the competitor can make use of theory of mind. However, the results show no significant differences between the performance of ToM_3 allocators and ToM_4 allocators, suggesting that fourth-order theory of mind does not yield any additional advantages for allocators. These results therefore suggest that in this particular Colored Trails setup, there is no competitive advantage for orders of theory of mind beyond the third.

Although the responder simply selects the offer that is most beneficial to her, the responder also increases her score through the negotiation process. Figure 5 summarizes this increase in score due to negotiation for the responder in the static environment with static goals as a function of the theory of mind abilities of both the allocator and the competitor agent. That is, Fig. 5 shows the average size increase of the pie that the responder accepts from either the allocator or the competitor. Note that the figure shows symmetry in the sense that the increase in score of the responder in the presence of a ToM_0 allocator and a ToM_1 competitor is the same as the increase in score of the responder in the presence of a ToM_1 allocator and a ToM_0 competitor.

Figure 5 shows that an allocator's first-order theory of mind is beneficial to the responder. When either the allocator or the competitor is a ToM_1 agent, the responder obtains a higher increase in score than when both of them are ToM_0 agents (leftmost bar, only 21.4 points). That is, ToM_1 allocators make offers that are more generous towards the responder than ToM_0 allocators. Similarly, ToM_2 allocators make offers that are more generous towards the responder than ToM_1 allocators. Interestingly, however, ToM_2 allocators fail to do so when the competitor is incapable of reasoning at first-order theory of mind. In Fig. 5, this is apparent from the first group of bars. While the second bar is larger than the first, the remaining bars are no higher than the second one. That is, ToM_2 agents will only offer the responder a larger piece of the pie if they believe the competitor may offer a large piece of pie to the responder as well. Furthermore, theory of mind abilities beyond the second order do not seem to benefit the responder. In each group of bars, the height of the bars does not increase past the third bar.

Interestingly, the responder benefits more from the theory of mind abilities of an allocator than the allocator himself does. When both allocator and competitor are ToM_0 agents, each of them increases his score through negotiation by 15.0 points on average (leftmost bar in Fig. 4). The average increase in the score of the responder in this case is 21.4 points (leftmost bar in Fig. 5). When both allocator and competitor are ToM_2 agents, negotiation increases the score of each of them by an average of 14.4 points, while the responder receives an increase in score of 28.1 points on average. While higher-order theory of mind agents successfully negotiate a larger pie to share with the responder, they do so at the expense of increasing their own piece of the pie. The additional pie that higher-order theory of mind agents negotiate for ends up with the responder.

Figure 6 shows the increase in social welfare in the static Colored Trails environment with static goals due to negotiation as a function of the theory of mind abilities of the allocator and the competitor. That is, Fig. 6 shows the total size of the pie the agents end up sharing. The leftmost bar in Fig. 6 shows that even when both allocator and competitor are ToM_0 agents and do not take the score of the responder into account when making an offer, social welfare improves as a result of negotiation. As expected, the presence of a ToM_1 allocator has a small but positive influence on social welfare. That is, ToM_1 allocators succeed not only in obtaining a larger piece of pie than ToM_0 allocators, but also in enlarging the total size of the pie that is being shared by the agents.

Similarly, ToM_2 allocators have a stronger positive effect on social welfare than ToM_1 allocators. However, whether the allocator has the ability to make use of orders of theory of mind beyond the second does not appear to significantly increase social welfare any further.

Summing up, performance results in the static environment with static goals show that agents obtain additional competitive advantages from the ability to make use of first-order, second-order, and third-order theory of mind, while there are no additional competitive advantages for fourth-order theory of mind reasoning. We find that the use of theory of mind also results in a benefit for the responder. Interestingly, however, this ‘cooperative’ advantage only extends to second-order theory of mind. The presence of third-order theory of mind agents does not increase the share of the pie received by the responder any further. Moreover, the additional pie offered to the responder is not an altruistic act on the part of the theory of mind agents. Second-order theory of mind agents will only offer a larger piece of the pie to the responder when they believe their competitor to be capable of first-order theory of mind.

5.2 Static environment with dynamic goals

In the static environment with dynamic goals, agents played each of the 1001 games in a run on the same board with the same initial sets of chips. However unlike the simulation experiment in the static environment with static goals, at the start of each game, each agent was randomly assigned one of the 12 possible goals locations. Recall that allocators form zero-order beliefs for each possible goal location of the responder. This means that during the experimental round, an allocator in the static environment with dynamic goals only has an expected 83 previous observations of the responders reaction to his offers for the same goal location of the responder. However, unlike allocators in the static environment with static goals, allocators in the static environment with dynamic goals may also obtain relevant information from the offers made by their competitor. In particular, ToM_0 allocators may learn to make better offers by copying the behavior of their higher-order theory of mind competitor.

Figure 7 shows the increase in score due to negotiation of a ToM_i allocator in the presence of a ToM_j competitor in the final game of a run, averaged over 5000 runs. Note that even though the capabilities of agent have not changed, compared to the situation with static goals, ToM_0 agents in the static environment with dynamic goals are only able to negotiate for a small increase in their score. As a result, the advantage of applying first-order theory of mind is greatly increased. For each group of bars in Fig. 7, the second bar is much higher than the first bar.

Similar to the results in the static environment with static goals, there are additional advantages for reasoning at second-order and for reasoning at third-order theory of mind in the static environment with dynamic goals. As Fig. 8 shows, the additional advantages for second-order and third-order theory of mind reasoning are much smaller than the advantage of first-order theory of mind reasoning. However, this is mainly due to the poor performance of the ToM_0 allocator in this case.

Figure 8 shows the negotiation score of responders in the static environment with dynamic goals. The results show a similar pattern as the one described for the static environment with static goals. Higher-order theory of mind allocators make an offer that is just beneficial enough for the responder to convince her to choose the allocator's offer. However, there is an interesting difference to the static environment with static goals. In the first set of bars, corresponding to a zero-order theory of mind competitor, the negotiation score of the responder becomes higher for increasingly higher orders of theory of mind allocators. This may be caused by the ToM_0 competitor learning from the offers made by the allocator.

Figure 9 shows the increase in social welfare in the static environment with dynamic goals as a function of the theory of mind abilities of the allocator and the competitor. Irrespective of the theory of mind abilities of the competitor, the ability of the allocator to reason using first-order theory of mind greatly improves social welfare. The presence of theory of mind agents has a stronger effect in the static environment with dynamic goals than in the static environment with static goals. The presence of a ToM_2 allocator increases social welfare even further while reasoning using even higher orders of theory mind does not appear to increase social welfare any further. However, note that the social welfare achieved in the static environment with dynamic goals is fairly low. None of the bars in Fig. 9 reach a social welfare of 50, while the social welfare scores in the static environment with static goals depicted in Fig. 6 are all above 50.

In summary, as expected, theory of mind reasoning is more beneficial in the static environment with dynamic goals than it is in the static environment with static goals. Although the capabilities of the ToM_0 agent are the same across the two environments, added unpredictability in the environment decreases the ability of ToM_0 agents to learn what offer to make. As a result, the competitive advantage afforded by first-order theory of mind is greatly increased in this setting.

5.3 Dynamic environment with dynamic goals

In the dynamic environment with dynamic goals, agents played each game on a new randomly generated board with new randomly generated sets of chips and goal locations for each agent. Recall that *ToM* $_0$ agents form zero-order beliefs based on observable features in the environment, including the colors of each of the tiles on the board. Since there are 5^{25} possible game boards (five possible colors for each of the 25 tiles on the board), *ToM* $_0$ agent will not be able to learn the optimal response on any given game board in the 1000 training rounds. Indeed, Fig. 10 shows that *ToM* $_0$ allocators obtain a negotiation score of zero (cf. extremely low bars for *ToM* $_0$ allocators). Since they are unable to learn the responder's behavior, their 'offer' to the responder assigns all available chips to the *ToM* $_0$ allocator himself, and none to the responder. After all, agents choose the offer that maximizes their expected gain, and owning spare chips increases an agent's score.

Unsurprisingly given this fact, the advantage of reasoning about the mental content of others is more beneficial in this setting than in the previous settings. In fact, the scaling of the vertical axis in Fig. 10 has been adjusted to accommodate the first group of bars. Where theory of mind reasoning would increase an allocator's score by no more than 2 points on average in the static environment with static goals, a *ToM* $_1$ agent competing with a *ToM* $_0$ competitor obtains a negotiation score of 22.0 while the competitor obtains a negotiation score of 0, despite the fact that agents have exactly the same capabilities across environments.

Figure 10 shows additional advantages for second-order and third-order theory of mind reasoning. As in the environments discussed previously, fourth-order theory of mind reasoning does not appear to give a significant additional increase to the allocator's score in the dynamic environment with dynamic goals.

The responder's negotiation score in the dynamic environment with dynamic goals is depicted in Fig. 11. Note that this figure shows a pattern that is almost identical to the one observed in the static environment with dynamic goals (Fig. 8). In the dynamic environment, however, responders obtain a higher negotiation score than in the static environment with dynamic goals. The same effect can be seen in the results for social welfare (Fig. 12). Again, as in the previous environments, the presence of a first-order theory of mind agent significantly increases the responder's negotiation score and social welfare. The presence of a second-order theory of mind agent increases the responder's negotiation score and social welfare even further. As before, there does not appear to be a social welfare advantage for third-order theory of mind reasoning.

Summing up, the results in the dynamic environment with dynamic goals show the largest benefit for first-order theory of mind reasoning among the three environments we consider, which is mainly due to the inability of zero-order theory of mind agents to negotiate effectively in this environment. In addition, we find additional benefits for second-order and third-order theory of mind reasoning. The additional benefits for second-order theory of mind reasoning influence the negotiation score of both the allocator and the responder, while third-order reasoning only affects the score of the allocator.

6 Related work on models of theory of mind

In this paper, we describe a recursive approach to modeling theory of mind agents. However, our approach is not the only way to model theory of mind. In this section, we describe a number of methodologies similar to the one we have chosen here. A helpful, recent survey of agent models of theory of mind is provided by Albrecht and Stone [1].

Recursive and hierarchical approaches have been used to formally model the behavior of human participants [3, 11, 12, 15, 23, 37, 44, 47, 49, 50, 60, 61]. Level-k reasoning [15, 60], quantal response equilibria [47], cognitive hierarchies [11], and noisy introspection models [37] measure the level of sophistication of agents by the maximum number of steps of iterated reasoning the agent is capable of considering. Camerer et al. [11, 12] estimate the distribution of the level of sophistication used by human participants over a range of non-repeated one-shot games such as the p-beauty contest and the traveler's dilemma, and find that cognitive hierarchies both fit behavioral data well and are consistent with verbal reports, response times, and visual fixations. Camerer et al. find an average of 1.5 steps of iterated reasoning, although Wright and Leyton-Brown [70] estimate the average level of strategic reasoning to be closer to 0.5 steps. Part of the participants were found to use more than two steps of iterated reasoning, which roughly corresponds to the use of second-order theory of mind. Typically, however, only few players were found to be well-described as higher-level agents [69, 70].

In contrast to the models described above, our theory of mind agents learn from the behavior of others and adjust their level of theory of mind reasoning accordingly, similar to models such as recursive opponent modeling [35, 36], I-POMDP [34, 52], networks of influence diagrams [29], dynamic level-k models [39], experience-weighted attraction [10], and Game Theory of Mind [71]. Similar to our approach, I-POMDP agents construct nested behavioral models, but cannot observe the level of sophistication of other agents directly. Instead, agents infer the order of theory of mind at which other agents reason by matching observed behavior of others to the behavior predicted by the application of theory of mind. A ToM_k agent can reason about other agents that make use of orders of theory of mind up to and including $(k-1)$st-order theory of mind. This means that when a ToM_4 agent believes that his trading partner is a ToM_1 agent, he may decide to behave as if he were a ToM_2 agent. When agents mutually engage in modeling the order of theory of mind at which the other is reasoning, this may influence the effectiveness of higher orders of theory of mind. Through simulation, these effects are taken into account.

Our approach differs from previous work in that the behavior of our agents changes based on the observed behavior of others. Previous models of theory of mind typically assume that the most basic agent responds optimally under the assumption that other agents act according to a known non-strategic policy [11, 15, 37, 47, 49, 60, 71]. Instead, our zero-order theory of mind agents attempt to learn the optimal behavior through heuristics and associative learning, allowing the zero-order theory of mind agent to produce sophisticated policies (cf. [13]), since the exact behavior of a ToM_0 agent therefore depends on the behavior of others. For example, a zero-order theory of mind allocator in Colored Trails can learn from the offers made by more sophisticated agents. This way, a zero-order theory of mind agent may learn to behave as if it were using theory of mind, without the need of actually engaging in theory of mind itself. That is, the zero-order theory of mind allocator appears to take the goal of the responder into account, even though he is unable to conceive of the idea that others have unobservable mental content such as a goal.

Note that the goal of our agent model is not to accurately describe the way humans make use of theory of mind through agent-based modeling tools such as PsychSim [57, 59]. Rather, our goal is to explain the evolution of higher-order theory of mind abilities by determining under what circumstances the use of this cognitively demanding ability presents individuals with advantages over individuals that rely on simple heuristics [31, 32, 41].

Previous agent-based modeling studies have attempted to explain why higher-order theory of mind may have evolved by investigating environments in which this ability is particularly beneficial to the individual. Such research has shown that the use of higher-order theory of mind can result in an advantage over opponents in competitive environments [16, 17, 26, 27, 51], in more efficient cooperation [16, 18, 71], and in stability in negotiation settings [19]. While this previous research shows that higher-order theory of mind reasoning is effective in certain settings, in our current work, we argue that the effectiveness of higher-order theory of mind reasoning is not specific to a setting , but is influenced by the unpredictability of the environment.

In contrast to the computational models discussed above, theory of mind can also be modeled using formal logic [22, 23, 67]. For example, Felli et al. [23] construct human-like theory of mind through formal logic which includes both *stereotypical reasoning*, in which the agent uses simple social rules to predict the behavior of others, as well as *empathetic reasoning*, in which an agent attributes its own reasoning process to others to determine what behaviors of others are plausible.

7 Conclusion

Experimental evidence suggests that people make use of higher-order theory of mind, while other animals do not appear to have this ability. Agent-based modeling research shows that in competitive settings [16, 17, 26], cooperative settings [16, 18, 71], as well as in mixed-motive settings [19], agents can benefit from higher-order theory of mind reasoning. However, while this shows that higher-order theory of mind can be beneficial in some settings, it does not explain *why* these settings foster theory of mind reasoning.

In the current work, we show that the unpredictability of the environment contributes to the evolutionary advantages of higher-order theory of mind reasoning. We have placed computational agents in a simulated one-shot negotiation setting to show that the ability to make use of theory of mind can indeed lead agents to negotiate more effectively, resulting in a higher average score for both the agent himself and his trading partner. By varying the predictability of the environment, we show that the benefit of theory of mind reasoning is more pronounced in environments that are more dynamic (i.e. have more observable features that vary each round).

We investigated a particular mixed-motive setting known as Colored Trails, a board game that has been used as a research test-bed to investigate decision-making in groups of people and computer agents across a range of situations (e.g. [20, 24, 30, 46, 53, 68]). In our setting, an allocator and his competitor simultaneously offer a trade to a responder, who in turn chooses whether or not to accept one of these offers. This setup is very similar to the one used by Ficici and Pfeffer [24], who show that humans make use of theory of mind in these negotiation games. We hypothesized that agents that are unable to make use of theory of mind reasoning would be more successful when repeated negotiation games were played on the same game board than when every game was played on an new, unfamiliar game board.

Our results show that across the environments we considered, there is an advantage for first-order theory of mind reasoning. Furthermore, we find additional advantages for second-order and for third-order theory of mind reasoning, while fourth-order theory of mind reasoning does not appear to yield additional benefits. Interestingly, while the benefits of first-order and second-order theory of mind reasoning include a higher score for the reasoner himself as well as for his trading partner, third-order theory of mind reasoning only benefits the reasoner himself. That is, whereas first-order and second-order theory of mind help a reasoner to increase the size of the pie to share with a trading partner, the benefit of third-order theory of mind mainly consists of finding opportunities to share pie with a trading partner.

In addition, we find that as the environment becomes more unpredictable, agents incapable of theory of mind reasoning have more difficulties negotiating effectively, while theory of mind agents are unaffected. As a result, the benefits of theory of mind reasoning are more pronounced in unpredictable environments than they are in static environments, especially for first-order theory of mind reasoning. These results confirm that the diminishing returns to more complex opponent models observed by Ficici and Pfeffer [24] are not only due to limitations of human reasoning.

At first glance, the conclusion that complex reasoning strategies such as theory of mind may be especially beneficial in more complex environments may seem to contradict previous work that shows that heuristics are especially useful in complex environments (cf. [31, 32]). However, while theory of mind certainly is a complex reasoning strategy, it is also a heuristic in which the reasoner makes use of information that is not available in the environment to predict the behavior of others. That is, while theory of mind is not fast, it is a heuristic in the sense that it is frugal: it makes use of relatively little information to make a prediction (see also [5, 55]). Our results suggest that in complex environments, a mid-range heuristic such as theory of mind may outperform simpler heuristics.

Previous findings show that the effectiveness of higher-order theory of mind reasoning is typically less pronounced in cooperative settings than it is in competitive settings [16–18]. Our current results offer an explanation for this variation in the effectiveness of higher-order theory of mind reasoning across settings. In competitive games, players have an incentive to be unpredictable to opponents. That is, players in a competitive game aim to increase the unpredictability of the environments for their zero-order theory of mind competitors, which according to our current results benefits higher-order theory of mind reasoners. In cooperative settings, however, players have an incentive to be predictable for others. Players therefore aim to reduce the unpredictability of the environment in these settings, which would suggest that the effectiveness of higher-order theory of mind is diminished.

We previously performed agent-based simulations in an incomplete information Colored Trails environment [19] where two agents alternate in proposing a redistribution of chips until either an offer is accepted or one of the agents withdraws from negotiation. In contrast to our current findings in repeated single-shot negotiation among three agents, increasingly higher orders of theory of mind reasoning did not lead to higher negotiation scores in this alternating offers setting. Our current results suggest that this is at least partially due to the negotiation setting. When agents alternate in making offers on the same board (i.e. a static environment), agents have more opportunity to learn the optimal offer to make. Our results suggest that this benefits zero-order theory of mind agents, and decreases the potential benefit of theory of mind reasoning.

Agent-based modeling is typically performed using agents that rely on fixed strategies that are computationally inexpensive. Simulations with more complex reasoning strategies can provide additional insights (see, for example, also [65]), and are necessary to allow for human-agent teamwork [21, 42]. Our results show how agent-based modeling with agents that make use of complex reasoning strategies such as theory of mind can provide new insights into the benefits of those reasoning strategies. These agents may even be used to train people in the application of their theory of mind and negotiation skills ([19]). In future work, we aim to determine whether training people by having them negotiate with artificial agents in the Colored Trails setting improves their negotiation skills in more concrete settings such as negotiating the details of a smoking ban or negotiations between representatives of universities and student unions.

References

1. Albrecht, S. V., & Stone, P. (2018). Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence, 258*, 66–95.
2. Apperly, I. A. (2010). *Mindreaders: The Cognitive Basis of "Theory of Mind.''*. Hove, UK: Psychology Press.
3. Arad, A., & Rubinstein, A. (2012). The 11–20 money request game: A level-k reasoning study. *American Economic Review, 102*(7), 3561–3573.
4. Arslan, B., Verbrugge, R., Taatgen, N., & Hollebrandse, B. (2018). Accelerating the development of second-order false belief reasoning: A training study with different feedback methods. *Child Development*. https://doi.org/10.1111/cdev.13186 in press.
5. Bobadilla-Suarez, S., & Love, B. C. (2018). Fast or frugal, but not both: Decision heuristics under time pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(1), 24.
6. Bugnyar, T., & Kotrschal, K. (2002). Observational learning and the raiding of food caches in ravens, Corvus corax: Is it 'tactical' deception? *Animal Behaviour, 64*(2), 185–195. https://doi.org/10.1006/anbe.2002.3056
7. Bugnyar, T., Reber, S. A., & Buckner, C. (2016). Ravens attribute visual access to unseen competitors. *Nature Communications, 7*, 10506.
8. Burkart, J. M., & Heschl, A. (2007). Understanding visual access in common marmosets, callithrix jacchus: Perspective taking or behaviour reading? *Animal Behaviour, 73*(3), 457–469.
9. Cagan, P. (1956). The monetary dynamics of hyper-inflation. In M. Friedman (Ed.), *Studies in the quantity theory of money* (pp. 25–117). Chicago: The University of Chicago Press.
10. Camerer, C., & Hua Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica, 67*(4), 827–874.
11. Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics, 119*(3), 861–898.
12. Camerer, C. F., Ho, T. H., & Chong, J. K. (2015). A psychological approach to strategic thinking in games. *Current Opinion in Behavioral Sciences, 3*, 157–162.
13. Chandrasekaran, M., Doshi, P., Zeng, Y., & Chen, Y. (2017). Can bounded and self-interested agents be teammates? Application to planning in ad hoc teams. *Autonomous Agents and Multi-Agent Systems, 31*(4), 821–860.

14. Clayton, N. S., Dally, J. M., & Emery, N. J. (2007). Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1480), 507–522. https://doi.org/10.1098/rstb.2006.1992
15. Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. *Econometrica, 69*(5), 1193–1235. https://doi.org/10.1111/1468-0262.00239
16. Devaine, M., Hollard, G., & Daunizeau, J. (2014). Theory of mind: Did evolution fool us? *PloS ONE, 9*(2), e87619. https://doi.org/10.1371/journal.pone.0087619
17. de Weerd, H., Verbrugge, R., & Verheij, B. (2013). How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence, 199–200*, 67–92. https://doi.org/10.1016/j.artint.2013.05.004
18. de Weerd, H., Verbrugge, R., & Verheij, B. (2014). Higher-order theory of mind in the Tacit Communication Game. *Biologically Inspired Cognitive Architectures, 11*, 10–21. https://doi.org/10.1016/j.bica.2014.11.010
19. de Weerd, H., Verbrugge, R., & Verheij, B. (2017). Negotiating with other minds: The role of recursive theory of mind in negotiation with incomplete information. *Autonomous Agents and Multi-Agent Systems, 31*(2), 250–287. https://doi.org/10.1007/s10458-015-9317-1
20. de Jong, S., Hennes, D., Tuyls, K., & Gal, Y.K. (2011). Metastrategies in the Colored Trails game. In L. Sonenberg, P. Stone, K. Tumer, P. Yolum (Eds.). *Proceedings of the 10th international conference on autonomous agents and multiagent systems, international foundation for autonomous agents and multiagent systems,* (vol 2, pp. 551–558).
21. Dignum, F., Prada, R., & Hofstede, G.J. (2014) . From autistic to social agents. In *Proceedings of the 2014 international conference on autonomous agents and multi-agent systems, international foundation for autonomous agents and multiagent systems,* (pp. 1161–1164).
22. Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. (2004). *Reasoning about knowledge* (2nd ed.). Cambridge: MIT Press.
23. Felli, P., Miller, T., Muise, CJ., Pearce, AR., & Sonenberg, L .(2015). Computing social behaviours using agent models. In Q. Yang , M. Wooldridge (Eds.) *Proceedings of the twenty-fourth international joint conference on artificial intelligence,* (pp. 2978–2984).
24. Ficici, S.G., & Pfeffer, A. (2008b) . Modeling how humans reason about others with partial information. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems,* IFAAMAS, (pp. 315–322).
25. Flobbe, L., Verbrugge, R., Hendriks, P., & Krämer, I. (2008). Children's application of theory of mind in reasoning and language. *Journal of Logic, Language and Information, 17*(4), 417–442.
26. Franke, M., & Galeazzi, P. (2014). On the evolution of choice principles. In J. Szymanik, R. Verbrugge (Eds.) *Proceedings of the second workshop reasoning about other minds: logical and cognitive perspectives, co-located with advances in modal logic,* Groningen, The Netherlands, CEUR Workshop Proceedings, (vol 1208, pp. 11–15).
27. Frey, S., & Goldstone, R. L. (2018). Cognitive mechanisms for human flocking dynamics. *Journal of Computational Social Science, 1*(2), 349–375. https://doi.org/10.1007/s42001-018-0017-x
28. Friedman, M. (1957). *A theory of the consumption function*. New Jersey: Princeton University Press.
29. Gal, Y., & Pfeffer, A. (2003). A language for modeling agents' decision making processes in games. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems,* (pp. 265–272), ACM
30. Gal, Y., Grosz, B., Kraus, S., Pfeffer, A., & Shieber, S. (2010). Agent decision-making in open mixed networks. *Artificial Intelligence, 174*(18), 1460–1480. https://doi.org/10.1016/j.artint.2010.09.002
31. Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart, evolution and cognition*. Oxford: Oxford University Press.
32. Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: the foundations of adaptive behavior*. Oxford: Oxford University Press.
33. Gintis, H. (2009). *The bounds of reason: game theory and the unification of the behavioral sciences*. Princeton, NJ: Princeton University Press.
34. Gmytrasiewicz, P. J., & Doshi, P. (2005). A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research, 24*, 49–79. https://doi.org/10.1613/jair.1579
35. Gmytrasiewicz, P.J. & Durfee, E.H. (1995). A rigorous, operational formalization of recursive modeling. In V.R. Lesser (ed) *Proceedings of the first international conference on autonomous agents and multiagent systems,* (pp. 125–132).
36. Gmytrasiewicz, P. J., Noh, S., & Kellogg, T. (1998). Bayesian update of recursive agent models. *User Modeling and User-Adapted Interaction, 8*(1–2), 49–69. https://doi.org/10.1023/A:1008269427670

37. Goeree, J. K., & Holt, C. A. (2004). A model of noisy introspection. *Games and Economic Behavior, 46*(2), 365–382. https://doi.org/10.1016/S0899-8256(03)00145-3
38. Hedden, T., & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition, 85*(1), 1–36. https://doi.org/10.1016/S0010-0277(02)00054-9
39. Ho, T. H., & Su, X. (2013). A dynamic level-k model in sequential games. *Management Science, 59*(2), 452–469.
40. Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review. *Developmental Psychology, 52*(8), 1192.
41. Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux, New York.
42. Kaminka, G.A. (2013). Curing robot autism: A challenge. In *Proceedings of the 2013 international conference on autonomous agents and multi-agent systems, international foundation for autonomous agents and multiagent systems,* (pp. 801–804).
43. Kaminski, J., Bräuer, J., Call, J., & Tomasello, M. (2009). Domestic dogs are sensitive to a human's perspective. *Behaviour, 146*(7), 979–998. https://doi.org/10.1163/156853908X395530
44. Kawagoe, T., & Takizawa, H. (2012). Level-k analysis of experimental centipede games. *Journal Of Economic Behavior & Organization, 82*(2–3), 548–566.
45. Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology, 4*(3–4), 231–244.
46. Lin, R., Kraus, S., Wilkenfeld, J., & Barry, J. (2008). Negotiating with bounded rational agents in environments with incomplete information using an automated agent. *Artificial Intelligence, 172*(6–7), 823–851. https://doi.org/10.1016/j.artint.2007.09.007
47. McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior, 10*(1), 6–38. https://doi.org/10.1006/game.1995.1023
48. Miller, T., Pearce, A. R., & Sonenberg, L. (2018). Social planning for trusted autonomy. In H. A. Abbass, J. Scholz, & D. J. Reid (Eds.), *Foundations of trusted autonomy* (pp. 67–86). Switzerland: Springer, Cham.
49. Mohlin, E. (2012). Evolution of theories of mind. *Games and Economic Behavior, 75*(1), 299–318. https://doi.org/10.1016/j.geb.2011.11.009
50. Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review, 85*(5), 1313–1326.
51. von der Osten, FB., Kirley, M., & Miller, T. (2017). The minds of many: Opponent modelling in a stochastic game. In *Proceedings of the 25th international joint conference on artificial intelligence,* (pp. 3845–3851).
52. Panella, A., Gmytrasiewicz, P., Panella. (2017). Interactive POMDPs with finite-state models of other agents. *Autonomous Agents and Multi-Agent Systems, 31*(4), 861–904.
53. Peled, N., Kraus, S., et al. (2015). A study of computational and human strategies in revelation games. *Autonomous Agents and Multi-Agent Systems, 29*(1), 73–97.
54. Perner, J., & Wimmer, H. (1985). John *thinks* that Mary *thinks* that...'' attribution of second-order beliefs by 5 to 10 year old children. *Journal of Experimental Child Psychology, 39*(3), 437–471. https://doi.org/10.1016/0022-0965(85)90051-7
55. Pöppel, J. & Kopp, S. (2018). Satisficing models of Bayesian Theory of Mind for explaining behavior of differently uncertain agents: Socially interactive agents track. In *Proceedings of the 17th international conference on autonomous agents and multiagent systems, international foundation for autonomous agents and multiagent systems,* (pp. 470–478).
56. Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(04), 515–526. https://doi.org/10.1017/S0140525X00076512
57. Pynadath, D.V. & Marsella, S.C. (2005). PsychSim: Modeling theory of mind with decision-theoretic agents. In L.P. Kaelbling, A. Saffiotti (Eds.) *IJCAI-05: nineteenth international joint conference on artificial intelligence,* (pp. 1181–1186).
58. Raiffa, H., Richardson, J., & Metcalfe, D. (2002). *Negotiation analysis: the science and art of collaborative decision making*. Cambridge, MA: Belknap Press.
59. Si, M., Marsella, S. C., & Pynadath, D. V. (2010). Modeling appraisal in theory of mind reasoning. *Autonomous Agents and Multi-Agent Systems, 20*(1), 14.
60. Stahl, D. O., & Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior, 10*(1), 218–254. https://doi.org/10.1006/game.1995.1031
61. Stuhlmüller, A., & Goodman, N. D. (2014). Reasoning about reasoning by nested conditioning: Modeling theory of mind with probabilistic programs. *Cognitive Systems Research, 28*, 80–99. https://doi.org/10.1016/j.cogsys.2013.07.003
62. Tomasello, M. (2009). *Why we cooperate*. Cambridge, MA: MIT Press.

63. Verbrugge, R. (2009). Logic and social cognition: The facts matter, and so do computational models. *Journal of Philosophical Logic, 38*, 649–680. https://doi.org/10.1007/s10992-009-9115-9
64. Verbrugge, R., Meijering, B., Wierda, S., van Rijn, H., & Taatgen, N. (2018). Stepwise training supports strategic second-order theory of mind in turn-taking games. *Judgment and Decision Making, 13*(1), 79–98.
65. Wijermans, N., Jorna, R., Jager, W., van Vliet, T., & Adang, O. (2013). Cross: Modelling crowd behaviour with social-cognitive agents. *Journal of Artificial Societies and Social Simulation 16*(4), 1, https://doi.org/10.18564/jasss.2114, http://jasss.soc.surrey.ac.uk/16/4/1.html
66. Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128.
67. van Ditmarsch, H., van der Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic* (Vol. 337). Berlin: Springer.
68. van Wissen, A., Gal, Y., Kamphorst, B. A., & Dignum, M. V. (2012). Human-agent teamwork in dynamic environments. *Computers in Human Behavior, 28*(1), 23–33. https://doi.org/10.1016/j.chb.2011.08.006
69. Wright, JR., & Leyton-Brown, K . (2010) . Beyond equilibrium: Predicting human behavior in normal-form games. In *Proceedings of the 24th conference on artificial intelligence,* (pp. 901–907).
70. Wright, J.R. & Leyton-Brown, K. (2012). Behavioral game theoretic models: A Bayesian framework for parameter analysis. In *Proceedings of the 11th international conference on autonomous agents and multiagent systems-volume 2, International foundation for autonomous agents and multiagent systems,* (pp. 921–930).
71. Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game theory of mind. *PLoS Computational Biology, 4*(12), e1000254. https://doi.org/10.1371/journal.pcbi.1000254