Estimating the Use of Higher-Order Theory of Mind Using Computational Agents

3 authors, including:

Harmen de Weerd
Hanzehogeschool Groningen
20 PUBLICATIONS   128 CITATIONS

SEE PROFILE

Rineke Verbrugge
University of Groningen
209 PUBLICATIONS   1,439 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Efficient Metamathematics View project

Cognitive systems in interaction: Logical and computational models of higher-order social cognition View project

Research Article

Harmen de Weerd*, Denny Diepgrond, and Rineke Verbrugge

Estimating the use of higher-order theory of mind using computational agents

Abstract: When people make decisions in a social context, they often make use of *theory of mind*, by reasoning about unobservable mental content of others. For example, the behavior of a pedestrian who wants to cross the street depends on whether or not he believes that the driver of an oncoming car has seen him or not. People can also reason about the theory of mind abilities of others, leading to recursive thinking of the sort 'I think that you think that I think...'. Previous research suggests that this ability may be especially effective in simple competitive settings. In this paper, we use a combination of computational agents and Bayesian model selection to determine to what extent people make use of higher-order theory of mind reasoning in a particular competitive game known as matching pennies. We find that while many children and adults appear to make use of theory of mind, participants are also often classified as using a simpler reactive strategy based only on the actions of the directly preceding round. This may indicate that human reasoners do not primarily use their theory of mind abilities to compete with others.

1 Introduction

In social interactions, people often reason about the beliefs, goals, and intentions of others. People use this so-called *theory of mind* [39] or *mentalizing* to understand why others behave the way they do, as well as to predict the future behavior of others. People can even use their theory of mind to reason about the way others make use of theory of mind. For example, people make use of *second-order theory of mind* to understand a sentence such as "Alice *knows* that

*Corresponding Author: Harmen de Weerd, Denny Diepgrond, Rineke Verbrugge: Institute of Artificial Intelligence, University of Groningen

Bob *knows* that Carol is throwing him a birthday party", by reasoning about what Alice knows about what Bob knows.

The human ability to make use of higher-order theory of mind is especially apparent in story comprehension tasks. Adults perform much better than chance on story comprehension questions that explicitly involve theory of mind reasoning up to the fourth order [32, 46]. Interestingly and contrary to predictions of traditional game theory and complexity theory [47], experimental evidence shows that people have more difficulty applying their theory of mind abilities in strategic games. In these settings, individuals are typically found to reason at low orders of theory of mind and are slow to adjust their level of theory of mind reasoning to more sophisticated opponents [5, 26, 28, 49]. However, some empirical research suggests that the use of theory of mind by participants can be facilitated by context [14, 34], setting [11, 13, 26], and training [35].

Results from an empirical study by Goodie et al. [26] suggest that participants may be particularly encouraged to make use of higher-order theory of mind in simple and strictly competitive settings. Simulation studies show that the ability to make use of higher-order theory of mind can indeed be particularly effective in such settings [9, 10, 15, 17, 18]. However, in these simple competitive settings, it is difficult to distinguish between participants who make use of theory of mind and participants who rely on simpler, behavior-based strategies. In addition, participants may vary in their strategy use [see, for example, 21, 27, 36, 40]. This may cause estimation methods based on population data to yield unreliable results.

In this paper, we use a combination of computational agents and Bayesian model selection, introduced by [45], to estimate strategy use of individual participants in simple strategic games. This method allows us to consider a population that differs in their use of theory of mind, in contrast to the existing estimation methods in the behavioral economics literature, which determine which level of theory of mind reasoning best describes the population as a whole [5, 49]. Our goal is to test the effectiveness of the Bayesian estimation procedure, as well as to determine to what extent human participants make use of theory of mind in simple competitive games. To do so, we apply this method on two empirical studies in which participants play a simple game known as matching pennies.

In the first study, Devaine, Hollard, and Daunizeau [14] let participants play against Bayesian theory of mind opponents. Participants faced these opponents in the setting of a hide-and-seek task and in the setting of a casino task. In the hide-and-seek task, participants were asked to search for their opponent, who was hidden in one of two possible hiding locations. In the casino task, participants had to choose one of two slot machines to play against. Importantly, the only difference between these two tasks was the cover story presented to

the participants. Both games had the exact same structure, namely 'matching pennies', and the exact same computer opponent.

In the second study, Sher, Koenig, and Rustichini [43] let children between the ages of 3 and 9 play a sender-receiver game. This game involved two boxes, one of which hid a piece of candy, while the other box hid a rock. Only the sender was told which box contained the candy. The sender was then asked to point at one of the boxes, after which the receiver could choose one of the two locations. If the receiver selected the box with the candy, the receiver could keep the candy. Otherwise, the sender would get the candy.

In both these studies, human participants played against an opponent that followed a known and fixed strategy. In Devaine et al. [14], participants played against software agents that followed a theory of mind strategy, while Sher et al. [43] let children play against a confederate who always selected the action that would have won in the last round. In this paper, we estimate the level of theory of mind reasoning of both the participants as well as their pre-'programmed' opponents. This allows us to both validate the estimation method, by comparing our estimation results to the known strategies of the computer opponents, and estimate the extent of human theory of mind use in simple competitive games.

The remainder of this paper is structured as follows. In Section 2, we give an overview of related work on theory of mind in behavioral economics. In Section 3, we present the details of the matching pennies game that participants play in the studies of Devaine and colleagues and Sher and colleagues. Section 4 outlines the estimation method and agent strategies that we consider in our estimation. The results of the estimation are presented in Section 5. In Section 6, we discuss these results and suggest directions for future research.

2 Related work

The behavioral economics literature contains several approaches to bounded rationality and recursive modeling of the behavior of others that are related to the theory of mind agents we present in this study. Similar to our theory of mind model, the level of sophistication of an agent based on iterated best-response models such as level-n theory [1, 2, 7, 38, 44], cognitive hierarchies [5], quantal response equilibria [33], and noisy introspection models [25], is measured by the maximum number of steps of recursive reasoning that the agent can consider. When these models are used to estimate the level of iterated reasoning of a player, human participants are typically found to use low levels of iterated reasoning. Camerer et al. [5], for example, find that over various non-repeated

single-shot games such as the p-beauty contest and the traveler's dilemma, participants were estimated to use an average 1.5 steps of recursive reasoning, which corresponds to first-order theory of mind reasoning. In a meta-analysis of these types of games, Wright and Leyton-Brown [49] find evidence of participant behavior that is consistent with higher-order theory of mind reasoning. However, few players were found to be well-described as higher-level agents.

One limitation of the iterated reasoning models described above is that a level-n agent assumes that all other agents are exactly one level of sophistication lower than himself, or that the distribution of lower level agents can be described with a fixed probability distribution. However, in repeated game settings, such assumptions can be detrimental to an agent [30]. The theory of mind agents we describe in the rest of this article are more similar to dynamic models of theory of mind, such as experience-weighted attraction learning [6], recursive opponent modeling [22, 24], interactive POMDPs [23], and game theory of mind [50]. In these approaches, agents adjust their level of recursive reasoning in reaction to the behavior of others. An agent of level k can consider others as being agents of any level up to and including level $k - 1$. Such an agent does not observe the level of sophistication of others directly, but forms beliefs concerning the level of sophistication of others based on observed behavior.

These dynamic models of theory of mind reasoning show that over repeated trials, human participants can successfully increase their level of theory of mind reasoning. For example, Doshi et al. [16] use adjusted interactive POMDPs to model human behavior in repeated competitive single-shot games. They find that although humans generally reason at low levels of theory of mind, participants exhibit higher levels of reasoning in simpler settings. Yoshida et al. [50] evaluate the behavior of human participants in a sequential game variation on the cooperative Stag Hunt game. Using game theory of mind, they find evidence that participants make use of higher-order theory of mind reasoning.

When computational models are used in the literature to estimate the level of theory of mind reasoning of human players, the models are typically fitted on aggregate data. That is, the estimation method implicitly assumes that there is a *single* model that best explains the behavior of *all* participants. In contrast, we make use of random-effects Bayesian model selection, which estimates what *distribution* of strategies best explains participant data. This modeling approach allows for participants to differ in the strategy they employ. Therefore, our approach is more realistic, corresponding to the large variety of strategies among individual human players [21, 27, 36, 40].
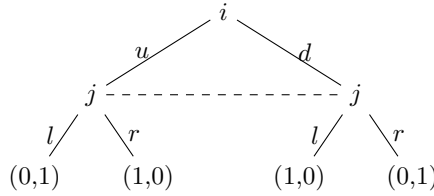
Fig. 1. The hide-and-seek game [14] and the sender-receiver game [43] share a common underlying structure. First, player i chooses to perform either action u or action d. Without knowing what choice player i has made, player j then chooses to perform either action l or r. This results in an outcome (π^i, π^j), where player i obtains payoff π^i and player j obtains outcome π^j.

3 Matching pennies

Matching pennies is a simple two-player game in which two players, i and j, independently select one of two possible actions. Figure 1 shows an extensive form representation of the game. In this figure, the actions resulting in the outcomes (u, l) and (d, r) are considered 'matched', which results in player j winning. In the remaining outcomes (u, r) and (d, l), the actions are 'mismatched', and player i wins.

The hide-and-seek game [14] and the sender-receiver game [43] share the underlying structure of matching pennies. Both games start with player i selecting to perform either action u or action d. Afterwards, without knowing the choice of player i, player j decides whether to perform action l or action r. In the hide-and-seek game [14], player i is the hider that hides at location 0 (action u) or at location 1 (action d). Player j is the seeker that can either search location 0 (action l) or location 1 (action r). In the sender-receiver game [43], player i is the sender who points at the hidden candy (action u) or at the hidden rock (action d). Player j is the receiver who either selects the box pointed out by the sender (action l) or the other box (action r).

The unique Nash equilibrium in matching pennies is for both players to randomly select one of their actions to play [see, for example, 29]. That is, unless both players play randomly, at least one of the players has an incentive to change their behavior. Human participants are known to have difficulties generating random sequences [41, 48]. As a result, human participants playing repeated matching pennies games typically deviate from playing the Nash equilibrium. This may encourage participants to try and take advantage of their opponent's deviations. In Section 4, we describe the method we use to determine what strategies participants use.

4 Estimation method

To determine to what extent human participants make use of theory of mind when playing matching pennies, we make use of a technique known as group-level random-effects Bayesian model comparison (RFX-BMS), introduced by Stephan et al. [45]. This technique models participants as individuals who can differ in the strategy they use while playing matching pennies. Random-effects Bayesian model selection treats strategies as random effects that can vary among participants, and which occur with fixed and unknown population frequencies. That is, unlike fixed-effects Bayesian model selection, we do not assume that there is one strategy that best describes the actions of all participants. Instead, we define a number of strategies that participants may use, including both theory of mind strategies and simpler behavior-based strategies. Each of these strategies s generates pieces of evidence $p(y|s)$ representing the probability that choosing actions according to strategy s will result in some observed data y. In addition, participants are assumed to be a random sample of the population of strategies. Using the model evidences and the participant data, random-effects Bayesian model selection estimates the relative frequencies of strategies in the general population.

To test to what extent participants make use of theory of mind while playing matching pennies, we compare the observed behavior of participants with the predicted behavior of computational agents following different strategies. That is, the model evidence $p(y|s)$ generated by a given model is the probability that that model will perform the same action as the participant, given the history of moves observed by the participant. In addition to theory of mind strategies, we consider a number of reactive strategies. These reactive strategies do not make use of an unobservable internal state. Instead, these strategies only react to the most recently observed behavior. This technique has previously been used by Devaine et al. [14] to estimate the effect of framing on theory of mind use in participants. Instead, we compare theory of mind use across different opponent strategies. We also attempt to identify known strategies through Bayesian RFX-BMS estimation. This allows us to determine to what extent Bayesian RFX-BMS estimation accurately distinguishes between theory of mind strategies and behavior-based reactive strategies.

In the following subsections, we discuss each of the strategies included in the analysis in detail. To avoid confusion, for the remainder of this paper we refer to player i as if she were female, while we refer to player j as if he were male.

4.1 Reactive strategies

To determine whether participant behavior is better explained by simple reactive strategies or by the use of theory of mind, we consider three reactive strategies. The reactive strategies we describe here do not make use of an internal state. Instead, these strategies respond to actions observed in the previous round of play only. Each of these strategies is parameterized with a single parameter $\lambda \in [0, 1]$, which we will refer to as the learning speed. Note that although we use the same symbol for each strategy, λ has a different interpretation in each strategy.

The first reactive strategy we consider is the *biased* strategy. This strategy does not react to the behavior of either player, but instead selects each action with a fixed probability. Specifically, a biased player i selects action u with some individual bias probability λ and action d with probability $(1 - \lambda)$. Similarly, when player j follows a biased strategy, he selects action l with probability λ and action r with probability $(1 - \lambda)$.

We also consider an *other-regarding* strategy. A player following this strategy selects with probability λ the action that would have resulted in the player winning the previous round. As a result, an other-regarding agent selects the action that would have lost the previous round with probability $(1 - \lambda)$. For example, suppose that in the previous round of the game, player i has chosen action u and player j has selected action l. Since this means that player j has won the previous round, an other-regarding player i would select action d with probability λ, while an other-regarding player j would repeat his previous choice l with probability λ.

Finally, we consider a *self-regarding* strategy. A self-regarding player repeats the action performed in the previous round with probability λ and selects the other action with probability $(1 - \lambda)$. For example, if a self-regarding player i chooses action u, her next action has probability λ of being u and probability $(1 - \lambda)$ of being d. This strategy is intended to model players that attempt to play randomly, but either switch between actions too often ($\lambda < 0.5$) or switch too little ($\lambda > 0.5$).

Note that each of these three reactive strategies is equivalent to the Nash equilibrium strategy when the parameter λ equals 0.5. In this case, each strategy randomly chooses one of the two options to play. In addition to these three reactive strategies, we also include the Nash equilibrium strategy as a separate parameter-less strategy in our analysis of player strategies.
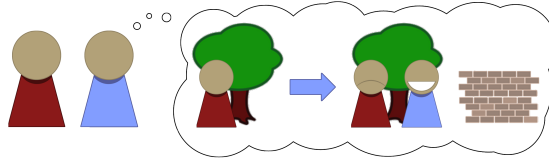
Fig. 2. If the blue agent j is a ToM_0 seeker, his beliefs concerning the hiding location of the hider i are based on her previous behavior. If the red agent i has typically hidden herself behind the tree, the blue agent j believes it is most likely that she will hide behind the tree again, and choose to search for her there.

4.2 Zero-order theory of mind

In addition to the reactive strategies described in the previous subsection, we also consider the possibility that participants apply a strategy based on the use of theory of mind. Note that the theory of mind strategies we describe here implement our theory of mind (ToM) agents that we studied in the game of rock-paper-scissors [9]. These agents are known to be able to take advantage of the use of higher-order theory of mind when playing rock-paper-scissors against a less sophisticated opponent. Moreover, our previous results show that these higher-order theory of mind agents maintain this advantage even if they do not accurately model the beliefs of their opponent [9]. These agents differ from the Bayesian theory of mind agents used in the study of Devaine et al. [14]. Importantly, this means that our random-effects Bayesian model selection does not include the exact strategies used by the computational theory of mind agents. By using a different agent model to estimate the use of theory of mind than was used to generate agent behavior in the hide-and-seek experiment, we aim to demonstrate that Bayesian RFX-BMS estimation [45] can estimate the level of theory of mind reasoning of a player, even if that player's strategy deviates from our specific implementation of theory of mind.

A zero-order theory of mind (ToM_0) agent is a goal-directed agent that is unable to represent or reason about mental content. Instead, the ToM_0 agent makes predictions about the behavior of the opponent based only on previously observed behavior. Figure 2 shows an example of this process in the hide-and-seek game. In this example, previous behavior of the red hider agent (agent i) leads the blue ToM_0 seeker agent (agent j) to believe that she has hidden herself behind the tree. As a result, the ToM_0 seeker j believes that he should look for his opponent at the same location.

A ToM_0 agent forms zero-order beliefs $b^{(0)}$ so that $b^{(0)}(a)$ represents the probability that the agent assigns to the event of its opponent playing some given

action a. For example, if player i is a ToM_0 agent, she believes that opponent j will play action l with probability $b^{(0)}(l) = 1 - b^{(0)}(r)$. Using these beliefs, the ToM_0 agent can calculate the expected value of playing a given action. For example, the expected value $EV^{(0)}(u)$ that ToM_0 agent i assigns to playing action u is

$$EV_i^{(0)}(u; b^{(0)}) = \sum_{a^j \in \{l,r\}} b^{(0)}(a^j) \cdot \pi^i(u, a^j) = b^{(0)}(r) = 1 - b^{(0)}(l). \quad (1)$$

The ToM_0 agent then chooses to play the action a^i that maximizes the expected value. That is, a ToM_0 agent with zero-order beliefs $b^{(0)}$ chooses to play action

$$t_i^{(0)}(b^{(0)}) = \arg\max_a EV^{(0)}(a; b^{(0)}). \quad (2)$$

Whenever a ToM_0 agent observes the behavior of its opponent, it updates its zero-order beliefs $b^{(0)}$. For example, when ToM_0 agent i observes player j playing action a^j, she updates her beliefs so that

$$b^{(0)}(a) := \begin{cases} (1-\lambda) \cdot b^{(0)}(a) + \lambda & \text{if } a = a^j, \\ (1-\lambda) \cdot b^{(0)}(a) & \text{otherwise.} \end{cases} \quad (3)$$

Note that the ToM_0 strategy is similar to the other-regarding strategy described in Section 4.1. Like the other-regarding strategy, the ToM_0 strategy is drawn to the action that would have won in the previous round. The distinction between the ToM_0 strategy and the reactive strategies described in the previous subsection is that a ToM_0 agent has an internal state that summarizes all previously observed behavior of the opponent, while the reactive other-regarding strategy only reacts to the actions in the most recently observed round of play. That is, when an other-regarding player j observes that player i has hidden behind the tree, he will respond by choosing to search near the tree with probability λ, irrespective of previous observations. In contrast, a ToM_0 player j is more likely to search near the tree when player i has hidden there three times in a row than he is to search near the tree when player i has only hidden there once.

For the purpose of Bayesian RFX-BMS estimation of theory of mind use, we follow Devaine et al. [14] and consider that choices may exhibit small deviations from the optimal decision rule defined by a strategy. We therefore employ the so-called 'softmax' probabilistic policy. That is, the probability that a ToM_0 player i will perform action u is given by
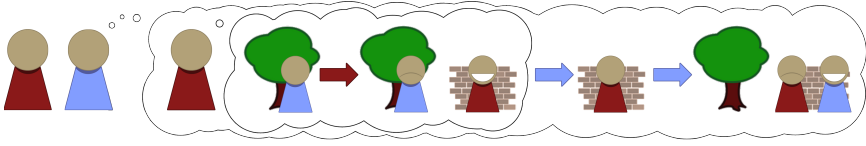
Fig. 3. If the blue agent j is a ToM_1 seeker, he puts himself in the position of the hider i to predict where she will hide. If the ToM_1 seeker j usually searches near the tree, the hider i may believe that he is going to search for her near the tree again. If this is what she believes, she would hide behind the wall to avoid the seeker. Therefore, the ToM_1 seeker j concludes, he should seek for the hider behind the wall.

$$
\begin{aligned}
P(A^i = u) \quad &= \quad s\left(\frac{EV_i^{(0)}(u; b^{(0)})}{\beta}\right) \\
&:= \quad \frac{\exp\left(EV_i^{(0)}(u; b^{(0)})/\beta\right)}{\exp\left(EV_i^{(0)}(u; b^{(0)})/\beta\right) + \exp\left(EV_i^{(0)}(d; b^{(0)})/\beta\right)},
\end{aligned}
\tag{4}
$$

where β is the exploration temperature, a free parameter that controls the magnitude of behavioral noise.

4.3 First-order theory of mind

In contrast to a ToM_0 agent, a first-order theory of mind (ToM_1) agent can reason about the mental content of its opponent and realize that the opponent has a goal of its own. A ToM_1 agent can place itself in the position of its opponent and calculate what the agent would have done itself in that position. Figure 3 shows an example of this process for a ToM_1 seeker agent j in the hide-and-seek game. Based on his own previous actions, the ToM_1 seeker j reasons that if he had been in the position of the hider, he would predict that the seeker is going to search behind the tree. To avoid this seeker, the ToM_1 agent would therefore have chosen to hide behind the wall. The ToM_1 seeker j attributes this reasoning process to his opponent and therefore concludes that she is most likely to hide behind the wall. As a result, the ToM_1 seeker decides to search behind the wall.

More in general, to model the beliefs of her opponent, a ToM_1 agent i forms first-order beliefs $b^{(1)}$ that represent what the agent's zero-order beliefs would have been if she had been in the position of her opponent. That is, according to the first-order beliefs $b^{(1)}$ of ToM_1 agent i, if agent i had been in the position of her opponent j, she would have believed that the probability that agent i will

perform action a^i is $b^{(1)}(a^i)$. If she had been in the position of her opponent j, agent i would therefore have chosen to play action $t_j^{(0)}(b^{(1)})$. Using first-order theory of mind, *ToM*$_1$ player i predicts that opponent j will do the same.

Based on the observed behavior of her opponent, a *ToM*$_1$ player i may come to believe that her first-order beliefs do not accurately predict the behavior of player j. In this case, she may decide to act as if she were a *ToM*$_0$ agent instead. This behavior is controlled by the *ToM*$_1$ agent's confidence $c_1 \in [0,1]$ in first-order theory of mind. The higher the confidence c_1, the more the behavior of a *ToM*$_1$ agent is determined by its first-order beliefs $b^{(1)}$. When the confidence c_1 reaches zero, the *ToM*$_1$ player will play as if it were a *ToM*$_0$ player. The expected value $EV_i^{(1)}(a^i)$ that *ToM*$_1$ agent i assigns to playing action a^i is

$$EV_i^{(1)}(a^i; b^{(0)}, b^{(1)}, c_1) = c_1 \cdot \pi^i\left(a^i, t_j^{(0)}(b^{(1)})\right) + (1 - c_1) \cdot EV_i^{(0)}(a^i; b^{(0)}). \quad (5)$$

The expected value for *ToM*$_1$ player j is constructed analogously. Similar to a *ToM*$_0$ agent, the *ToM*$_1$ agent chooses to play the action $t_i^{(1)}(b^{(0)}, b^{(1)}, c_1)$ that maximizes its expected value.

After observing the outcome of a game in which the *ToM*$_1$ agent i decided to play action a^i and the opponent j played action a^j, a *ToM*$_1$ agent i updates her confidence c_1 in first-order theory of mind so that

$$c_1 := \begin{cases} (1 - \lambda) \cdot c_1 + \lambda & \text{if } a^j = t_j^{(0)}(b^{(1)}) \\ (1 - \lambda) \cdot c_1 & \text{otherwise.} \end{cases} \quad (6)$$

That is, if the agent's first-order theory of mind accurately predicted that opponent j would play a^j, the *ToM*$_1$ agent increases her confidence c_1 in first-order theory of mind. Otherwise, the agent lowers her confidence.

Next, the *ToM*$_1$ agent i updates her zero-order beliefs as described in Equation (3). In the same way, the *ToM*$_1$ agent i updates her first-order beliefs $b^{(1)}$ to increase the belief that she will play action a^i again. Note that the *ToM*$_1$ agent i updates her first-order beliefs $b^{(1)}$ using her own learning speed λ. Unlike Devaine's Bayesian agents [14], our *ToM* agents do not attempt to estimate the learning speed λ of their opponent. However, simulation results show that our theory of mind agent can take advantage of the use of higher-order theory of mind even when the implicit assumption of equal learning speeds is violated [9].

Similar to the procedure of the *ToM*$_0$ agent, we use a softmax policy with exploration temperature β, so that the probability that a *ToM*$_1$ player i will perform action u is given by

$$P(A^i = u) = s\left(EV_i^{(1)}(u; b^{(0)}, b^{(1)}, c_1)/\beta\right), \quad (7)$$

where β is the exploration temperature.

4.4 Higher orders of theory of mind

A kth-order theory of mind (ToM_k) agent considers the possibility that its opponent is a ToM_{k-1} agent in addition to the possibility that the opponent is reasoning at even lower orders of theory of mind. To predict the behavior of its opponent, a ToM_k agent takes the perspective of the opponent and determines what it would have done itself, based on its own kth-order beliefs $b^{(k)}$ and confidence c_k in kth-order theory of mind.

Analogous to the way a ToM_1 agent integrates the predictions of different orders of theory of mind, a ToM_k player i calculates the expected value $EV^{(k)}(a^i)$ of playing actions a^i as

$$EV_i^{(k)}(a^i; b^{(0)}, \ldots, b^{(k)}, c_1, \ldots, c_k) =$$
$$c_k \cdot \pi^i \left(a^i, t_j^{(k-1)}(b^{(1)}, \ldots, b^{(k)}, 1, 0, \ldots, 0) \right) +$$
$$(1 - c_k) \cdot EV_i^{(k-1)}(a^i; b^{(0)}, \ldots, b^{(k-1)}, c_1, \ldots, c_{k-1}), \tag{8}$$

and decides to play the action $t_i^{(k)}(b^{(0)}, \ldots, b^{(k)}, c_1, \ldots, c_k)$ that maximizes this expected value.

Following our previous work [9], our ToM agents do not attempt to model their opponent's confidence in theory of mind. Rather, a ToM_k agent i maintains $k+1$ models of opponent behavior (corresponding to her $0, 1, \ldots k$th-order theory of mind) and performs the actions prescribed by the model that she believes to most accurately predict the behavior of her opponent.

After observing the outcome of a game in which player i played action a^i and player j played action a^j, the ToM_k agent updates its confidences c_n in nth-order theory of mind according to Equation (6) and corresponding nth-order beliefs $b^{(n)}$. For each even numbered order of theory of mind n, player i updates her beliefs $b^{(n)}$ according to Equation (3) to reflect that she believes it to be more likely that her opponent will play action a^j again. For each odd numbered order of theory of mind m, player i updates her beliefs $b^{(m)}$ to reflect that she predicts her opponent to believe that she is more likely to play action a^i again.

To obtain the likelihood that a ToM_k player i will play action u, we use the softmax policy so that

$$P(A^i = u) = s \left(EV_i^{(k)}(u; b^{(0)}, \ldots, b^{(k)}, c_1, \ldots, c_k) / \beta \right), \tag{9}$$

where β is the exploration temperature.

5 Results

To determine the extent to which theory of mind is used when playing matching pennies, the strategies described in Section 4 are used as the basis for random-effects Bayesian model selection [45]. These strategies include two randomizing strategies (Nash, biased), three behavior-based strategies (other-regarding, self-regarding, ToM_0), as well as four theory of mind strategies (ToM_k, $k \in \{1, 2, 3, 4\}$). Note that only the Nash strategy is parameter-free. The biased, other-regarding, and self-regarding strategies each have a single parameter λ, while the ToM_k strategies ($0 \leq k \leq 4$) have two free parameters: λ and β. The values of these parameters were allowed to vary between subjects, but were assumed to be fixed within subjects.

The experimental data from Devaine et al. [14] and Sher et al. [43] contain the behavioral responses of human participants following an unknown strategy, but also the actions performed by their opponents, who strictly follow a fixed strategy. In Section 5.1, we start by estimating strategy use of these opponents to show that Bayesian RFX-BMS estimation can successfully recover known strategies from behavioral data. In Section 5.2, we estimate the strategies used by human participants in these matching pennies games.

5.1 Validation

To show that Bayesian RFX-BMS estimation can successfully recover player strategies, we apply this method to the behavior of players that follow a known strategy. Since the ToM_k strategies have an additional free parameter compared to the simpler, reactive strategies, a particular concern is that these theory of mind strategies can more accurately model a broader range of behavior. In this section, we show that this does not result in a tendency to erroneously classify strategies that are known to be simple, reactive behavior-based strategies as theory of mind strategies instead.

In the experimental study of Sher et al. [43], children played the sender-receiver game against a confederate who was instructed to follow a fixed other-regarding strategy. After the initial choice, the confederate would always select the action that would have won in the previous round. That is, confederates followed an other-regarding strategy with learning speed $\lambda = 1.0$. Figure 4 shows the estimated proportions of each of the nine strategies we consider based on the behavioral data of confederates in the sender-receiver game. As the figure shows, Bayesian RFX-BMS estimation successfully recovers the confederate
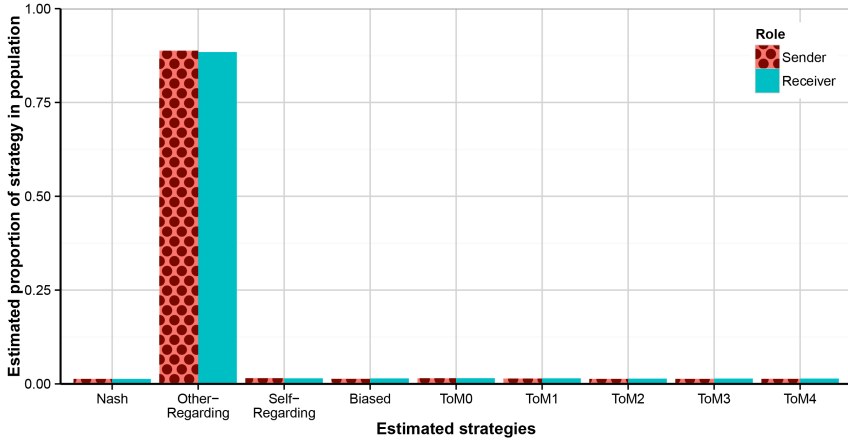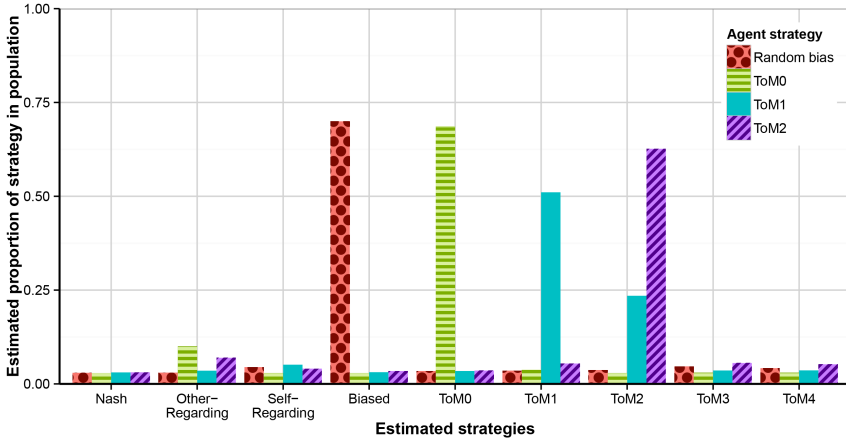
Fig. 4. Estimated proportions of strategies used by confederates in the sender-receiver game [43].

other-regarding strategy. In addition, the average estimated value of the learning speed parameter λ was 0.98. This shows that Bayesian RFX-BMS estimation can successfully recover a player strategy if it is included in the estimation as a population strategy.

The experimental study of Devaine et al. [14] also included players following a known strategy. Both in the hide-and-seek task and in the casino task, participants played against four different computer agents. These agents included a random biased agent that always chose one of the options with a 65% probability and three Bayesian theory of mind agents. Note that although Devaine's Bayesian agents make use of theory of mind, the exact specifications of these agents differ from our *ToM* agent descriptions in Section 4. That is, the strategies used by Devaine's Bayesian agents are not included as population strategies and can therefore not be recovered from the empirical data. This allows us to test whether Bayesian RFX-BMS estimation can accurately classify a strategy as a theory of mind strategy, even if the specific implementation of the theory of mind strategy used by the player differs from our *ToM* agent definition. In addition, by using different agents to generate behavior and to classify behavior, we can determine to what extent different orders of theory of mind reasoning are consistent among our *ToM* agents and Devaine's Bayesian agents.

Figure 5 shows the results of Bayesian RFX-BMS estimation on the behavioral data of computer agents in the Devaine et al. study, aggregated across the two task settings. The figure shows that the Bayesian RFX-BMS estimation
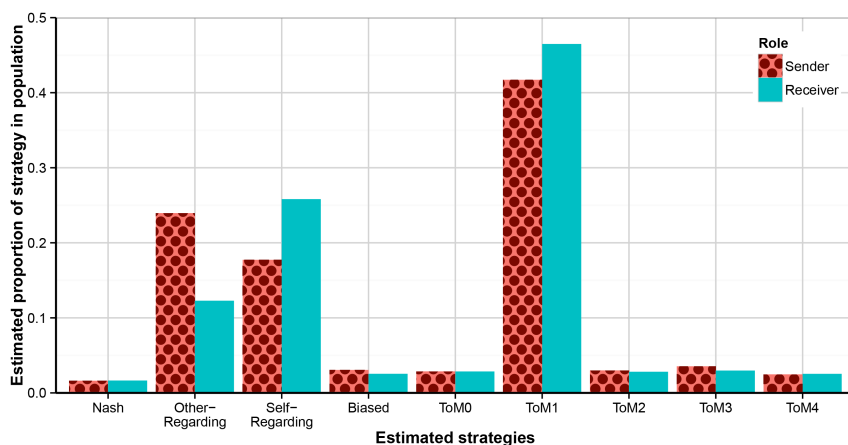
Fig. 5. Estimated proportions of strategies used by Devaine's Bayesian agents in the hide-and-seek game and the casino game [14].

method accurately distinguishes between theory of mind strategies and simpler, reactive strategies. In addition, the theory of mind abilities of Devaine's Bayesian agents are estimated remarkably well, despite the differences in agent specification[1]. In our Bayesian RFX-BMS estimation results, only Bayesian first-order theory of mind agents are regularly misclassified and identified as ToM_2 agents. This suggests that Devaine's Bayesian first-order theory of mind agents may be capable of more complex opponent modeling than the ToM_1 agent described in Section 4. Interestingly, Bayesian zero-order and second-order theory of mind agents are rarely misclassified, which suggests a good fit between the two models of second-order theory of mind.

The results in this section show that Bayesian RFX-BMS estimation can accurately recover known strategies used when playing matching pennies as well as distinguish between theory of mind strategies and reactive strategies. Moreover, Bayesian RFX-BMS estimation accurately distinguishes between levels of theory of mind reasoning, even when the level of theory of mind reasoning is estimated using a different model than the model that generated the behavior. This suggests that Bayesian RFX-BMS estimation may be useful in determining the

---

1 Note that Devaine's Bayesian theory of mind agents are not classified as using Devaine's Bayesian theory of mind strategy because these strategies are not included in the RFX-BMS estimation.
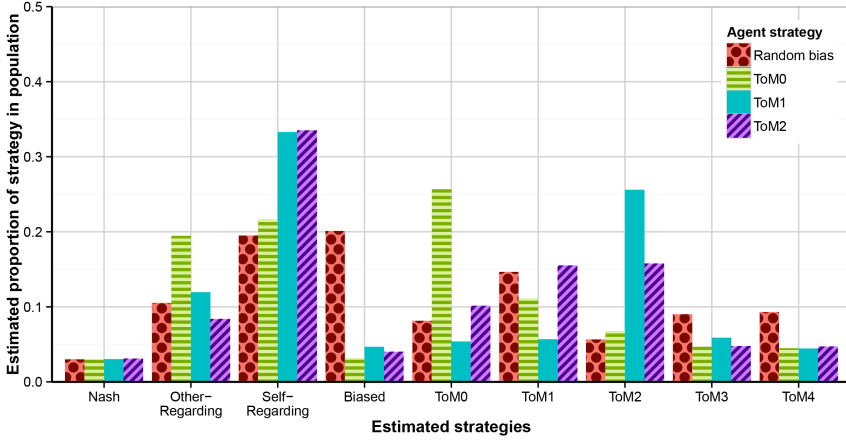
Fig. 6. Estimated proportions of strategies used by children playing the sender-receiver game [43].

extent to which human participants make use of theory of mind when playing this game. In the following subsection, we will consider this issue in detail.

5.2 Participant strategy use

The results of Bayesian RFX-BMS estimation on behavioral data of known strategies suggests that this method can accurately recover player strategies across several variants of the matching pennies game. In this section, we apply Bayesian RFX-BMS estimation to human participant data to determine the extent to which they make use of theory of mind.

Sher et al. [43] let children play a sender-receiver game against a confederate following a strict other-regarding strategy. Note that the best response against this strategy is to follow a self-regarding strategy that alternates between the two possible actions. Figure 6 shows the results of Bayesian RFX-BMS estimation of the strategies used by these children. The figure shows that Bayesian RFX-BMS estimation classifies over 40% of these children as ToM_1 agents, both in the sender role and in the receiver role. However, Figure 6 shows that in the sender role, 24% of the children are classified as using an other-regarding strategy and 17% as using a self-regarding strategy. In the receiver role, 12% of the children are classified as using an other-regarding strategy and 26% as using a self-regarding strategy. That is, the behavior of many of these children is best described as the use of a reactive strategy.

Fig. 7. Estimated proportions of strategies used by participants playing the hide-and-seek game and the casino game [14].

Devaine et al. [14] let participants play against Bayesian theory of mind agents in the social setting of a hide-and-seek task and the non-social setting of a casino task. As in the previous subsection, we aggregate data across the two tasks. Unlike Devaine et al., however, we do not aggregate across opponent strategy. That is, we explicitly take into account that participants may adjust their strategy based on the observed behavior of an opponent, while we ignore the effects of task context. Figure 7 shows the results of Bayesian RFX-BMS estimation of the strategies used by the participants. These results show strong variation in the strategy use of participants. Interestingly, participants are poorly described by randomizing strategies (Nash or biased), with one exception. When participants play against a biased opponent, 20% of them are best described as using a biased strategy as well. Note that in this case, the best response is indeed for participants to follow a biased strategy.

Compared to the estimated strategy use of children in Figure 6, adult participants are less often classified as making use of a theory of mind strategy. One notable exception is when participants play against a Bayesian first-order theory of mind agent, where the proportion of participants using a second-order theory of mind strategy is estimated at 25%.

The results in Figure 7 suggest that participants engage in some opponent modeling. Participants that play against a biased agent are more likely to be classified as using a biased strategy themselves, while participants that play against a first-order Bayesian theory of mind agent are more likely to be classified as using

second-order theory of mind. However, across the different opponent strategies, many participants are well-described as using a self-regarding strategy. This suggests that a sizable portion of participants may rely on simple reactive strategies when playing matching pennies in the context of the hide-and-seek or the casino game, irrespective of the behavior of the opponent [2].

To summarize, the results of our Bayesian RFX-BMS estimation suggest that there are both child and adult participants that make use of theory of mind strategies while playing matching pennies. However, there also appear to be many participants whose behavior is better described as following a simpler reactive strategy.

6 Discussion and conclusion

Both empirical and simulation studies suggest that players can benefit from reasoning about unobservable mental content of opponents in simple competitive games such as matching pennies or rock-paper-scissors [9, 14, 15, 17, 18, 26]. In this paper, we combined computational agents with Bayesian RFX-BMS estimation [45] to determine to what extent human participants actually make use of this so-called 'theory of mind' when playing repeated versions of matching pennies in two different empirical studies.

Sher et al. [43] let children play the sender-receiver game against a human confederate following a fixed strategy, while Devaine et al. [14] let participants play against computational Bayesian theory of mind agents in both a social hide-and-seek task and a non-social casino task. Our results show that Bayesian RFX-BMS estimation can accurately recover the known strategies used by both confederates and computational agents. In particular, players that were known to use a behavior-based strategy were not classified as theory of mind reasoners, while players that were known to follow a theory of mind strategy were classified as such. Our results also show that Bayesian RFX-BMS estimation can accurately identify a theory of mind strategy when the model used to estimate theory of mind use differs from the model that generated theory of mind behavior. This suggests that Bayesian RFX-BMS estimation can be used effectively to

---

2 When individual differences are ignored and all participants are assumed to best described by the same model, participants are classified as using fourth-order theory of mind. This result is partially due to the fact that the higher-order theory of mind models can typically generate behavior that resembles lower-order theory of mind more easily than the other way around.

determine theory of mind use in human participants, even when the the details of their theory of mind strategy are unknown.

Our results about human participants suggest that both children and adults engage in theory of mind when playing simple repeated games such as matching pennies. Many children make use of first-order theory of mind in the sender-receiver game, both in the role of sender and in the role of receiver. In addition, adult participants appear to make use of second-order theory of mind when facing a Bayesian first-order theory of mind agent. However, both the child data and the adult data show that many participants are better described as making use of a simpler reactive strategy. In addition, Devaine et al. [14] report that adult participants won, on average, 51% of the games they played. This may indicate that human players do not use their theory of mind abilities primarily to compete with others in one-shot situations such as matching pennies, or may even be unable to profitably apply theory of mind in these game situations.

One reason for the apparent lack of participants' use of theory of mind is that there are many viable simple strategies available in the matching pennies setting. The availability of these simple strategies may make it less appealing for participants to select a more complex and cognitively demanding theory of mind strategy. In addition, the wide range of strategies may make it more difficult for participants to use theory of mind to accurately predict the behavior of an opponent [cf. 21]. Future research could disentangle the effects of the appeal of using simple strategies from effects caused by the unpredictability of other players. Settings such as the centipede game and Marble Drop [34–37], for example, have fewer viable simple strategies than matching pennies. In such settings, less sophisticated players would therefore exhibit more predictable behavior, which may encourage human players to make more use of theory of mind as well. Experimental evidence shows that participants in these settings vary in their strategy use and level of theory of mind reasoning [see, for example, 21, 27, 36, 40], which suggests that Bayesian RFX-BMS estimation may yield additional insights.

Alternatively, human participants may be more likely to use their theory of mind in more cooperative settings. Experiments in the communicative-cooperative setting of the Tacit Communication Game, for example, demonstrate that participants readily reason about unobservable mental content of others [8]. In this setting, participants do not only reason about the way unfamiliar partners will interpret a novel signal, but also adjust their behavior depending on whether they believe to be partnered with a child or with an adult [4]. Additionally, simulation experiments with artificial agents have shown that the use of higher-order theory of mind can partially explain the high human performance in the Tacit Communication Game [12]. These results seem to suggest that par-

ticipants more readily make use of higher-order theory of mind in cooperative settings. Additionally, participants also seem to be more likely to make use of higher-order theory of mind in mixed-motive settings, where both cooperative and competitive aspects play a role[3].

In our analysis, we have only made use of participant choice data in the matching pennies game. Our results show that Bayesian RFX-BMS estimation can be effective in extracting information concerning the level of theory of mind reasoning from a sequence of binary choices. Other research has shown that information such as eye movements [31, 36] and reaction times [3, 42] can give additional insights concerning the strategies used by participants. Future research could extend the Bayesian RFX-BMS estimation to models that include predictions of reaction times and eye movements. In the setting of the Marble Drop game, for example, this could be done with the ACT-R and PRIMs models described by Meijering et al. [37], Ghosh et al. [20], and Ghosh and Verbrugge [19].

References

[1]   Ayala Arad and Ariel Rubinstein. The 11–20 money request game: A level-k reasoning study. *The American Economic Review*, 102(7):3561–3573, 2012.

[2]   M. Bacharach and D. O. Stahl. Variable-frame level-n theory. *Games and Economic Behavior*, 32(2):220–246, 2000.

[3]   Gerben Bergwerff, Ben Meijering, Jakub Szymanik, Rineke Verbrugge, and Stefan M Wierda. Computational and algorithmic models of strategies in turn-based games. In Paul Bello, Marcello Guarini, Marjorie McShane, and Brian Scassellati, editors, *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, pages 1778–1783, 2014.

[4]   Mark Blokpoel, Marlieke van Kesteren, Arjen Stolk, Pim Haselager, Ivan Toni, and Iris van Rooij. Recipient design in human communication: Simple heuristics or perspective taking? *Frontiers in Human Neuroscience*, 6(253), 2012. 10.3389/fnhum.2012.00253.

[5]   C.F. Camerer, T.H. Ho, and J.K. Chong. A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898, 2004.

[6]   Colin Camerer and Teck Hua Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874, 1999.

[7]   Miguel Costa-Gomes, Vincent P Crawford, and Bruno Broseta. Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235, 2001.

[8]   Jan Peter De Ruiter, M Noordzij, Sarah Newman-Norlund, Peter Hagoort, and Ivan Toni. On the origin of intentions. *Attention & Performance XXII*, pages 593–610, 2007.

---

3 For first results in this direction, see [11, 13].

[9] Harmen de Weerd, Rineke Verbrugge, and Bart Verheij. How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence*, 199-200:67–92, 2013. 10.1016/j.artint.2013.05.004.

[10] Harmen de Weerd, Rineke Verbrugge, and Bart Verheij. Theory of mind in the Mod game: An agent-based model of strategic reasoning. In A. Herzig and E. Lorini, editors, *Proceedings of the European Conference on Social Intelligence (ECSI-2014)*, volume 1283 of *CEUR Workshop Proceedings*, pages 128–136, 2014.

[11] Harmen de Weerd, Eveline Broers, and Rineke Verbrugge. Savvy software agents can encourage the use of second-order theory of mind by negotiators. In D. Noelle, R. Dale, A. Warlaumont, J. Yoshimi, T. Matlock, C. Jennings, and P. Maglio, editors, *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, pages 542–547, 2015.

[12] Harmen de Weerd, Rineke Verbrugge, and Bart Verheij. Higher-order theory of mind in the Tacit Communication Game. *Biologically Inspired Cognitive Architectures*, 11: 10–21, 2015. 10.1016/j.bica.2014.11.010.

[13] Harmen de Weerd, Rineke Verbrugge, and Bart Verheij. Negotiating with other minds: The role of recursive theory of mind in negotiation with incomplete information. *Autonomous Agents and Multi-Agent Systems,*, 31:250–287, 2017.

[14] Marie Devaine, Guillaume Hollard, and Jean Daunizeau. The social Bayesian brain: Does mentalizing make a difference when we learn? *PLoS Computational Biology*, 10 (12):e1003992, 2014. 10.1371/journal.pcbi.1003992.

[15] Marie Devaine, Guillaume Hollard, and Jean Daunizeau. Theory of mind: Did evolution fool us? *PloS ONE*, 9(2):e87619, 2014. 10.1371/journal.pone.0087619.

[16] Prashant Doshi, Xia Qu, Adam Goodie, and Diana Young. Modeling recursive reasoning by humans using empirically informed interactive POMDPs. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 1223–1230. IFAAMAS, 2010.

[17] Michael Franke and Paolo Galeazzi. On the evolution of choice principles. In Jakub Szymanik and Rineke Verbrugge, editors, *Proceedings of the Second Workshop Reasoning About Other Minds: Logical and Cognitive Perspectives, co-located with Advances in Modal Logic, Groningen, The Netherlands*, volume 1208 of *CEUR Workshop Proceedings*, pages 11–15, 2014.

[18] Seth Frey and Robert L. Goldstone. Cyclic game dynamics driven by iterated reasoning. *PloS ONE*, 8(2):e56416, 2013. 10.1371/journal.pone.0056416.

[19] Sujata Ghosh and Rineke Verbrugge. Studying strategies and types of players: Experiments, logics and cognitive models. *Synthese*, 2017. 10.1007/s11229-017-1338-7. in press.

[20] Sujata Ghosh, Ben Meijering, and Rineke Verbrugge. Strategic reasoning: Building cognitive models from logical formulas. *Journal of Logic, Language and Information*, 23(1):1–29, 2014. 10.1007/s10849-014-9196-x.

[21] Sujata Ghosh, Aviad Heifetz, and Rineke Verbrugge. Do players reason by forward induction in dynamic perfect information games? In R. Ramanujam, editor, *Proceedings of the 15th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2015)*, pages 121–130, 2015.

[22] P. Gmytrasiewicz and E. Durfee. A rigorous, operational formalization of recursive modeling. In *Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems*, pages 125–132, 1995.

[23] Piotr J Gmytrasiewicz and Prashant Doshi. A framework for sequential planning in multi-agent settings. *Journal of Artificial Intelligence Research*, 24:49–79, 2005.

[24] Piotr J Gmytrasiewicz, Sanguk Noh, and Tad Kellogg. Bayesian update of recursive agent models. *User Modeling and User-Adapted Interaction*, 8(1-2):49–69, 1998.

[25] Jacob K Goeree and Charles A Holt. A model of noisy introspection. *Games and Economic Behavior*, 46(2):365–382, 2004.

[26] Adam S. Goodie, Prashant Doshi, and Diana L. Young. Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making*, 25(1):95–108, 2012. 10.1002/bdm.717.

[27] Tamoghna Halder, Khyati Sharma, Sujata Ghosh, and Rineke Verbrugge. How do adults reason about their opponent? Typologies of players in turn-taking games. In D. Noelle, R. Dale, A. Warlaumont, J. Yoshimi, T. Matlock, C. Jennings, and P. Maglio, editors, *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, pages 854–859, 2015.

[28] Trey Hedden and Jun Zhang. What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, 85(1):1–36, 2002. ISSN 0010-0277. 10.1016/S0010-0277(02)00054-9.

[29] Aviad Heifetz. *Game Theory: Interactive Strategies in Economics and Management*. Cambridge University Press, 2012.

[30] Junling Hu and Michael P. Wellman. Online learning about other agents in a dynamic multiagent system. In Katia P. Sycara and Michael Wooldridge, editors, *Proceedings of the Second International Conference on Autonomous Agents*, pages 239–246, 1998. 10.1145/280765.280839.

[31] E. J. Johnson, C. F. Camerer, S. Sen, and T. Rymon. Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory*, 104:16—47, 2002. 10.1006/jeth.2001.2850.

[32] Peter Kinderman, Robin I.M. Dunbar, and Richard P. Bentall. Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, 89(2):191–204, 1998. 10.1111/j.2044-8295.1998.tb02680.x.

[33] R.D. McKelvey and T.R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, 1995.

[34] Ben Meijering, Leendert van Maanen, Hedderik van Rijn, and Rineke Verbrugge. The facilitative effect of context on second-order social reasoning. In Stellan Ohlsson and Richard Catrambone, editors, *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, pages 1423–1428, 2010.

[35] Ben Meijering, Hedderik van Rijn, Niels A. Taatgen, and Rineke Verbrugge. I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. In Laura Carlson, Christoph Hoelscher, and Thomas F. Shipley, editors, *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 2486–2491, 2011.

[36] Ben Meijering, Hedderik van Rijn, Niels A Taatgen, and Rineke Verbrugge. What eye movements can tell about theory of mind in a strategic game. *PLOS ONE*, 7:e45961, 2012. 10.1371/journal.pone.0045961.

[37] Ben Meijering, Niels A. Taatgen, Hedderik van Rijn, and Rineke Verbrugge. Modeling inference of mental states: As simple as possible, as complex as necessary. *Interaction Studies*, 15:455–477, 2014. 10.1075/is.15.3.05mei.

[38] Rosemarie Nagel. Unraveling in guessing games: An experimental study. *The American Economic Review*, pages 1313–1326, 1995.

[39] David Premack and Guy Woodruff. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(04):515–526, 1978. ISSN 0140-525X. 10.1017/S0140525X00076512.

[40] Maartje E. J Raijmakers, Dorothy J. Mandell, Sara E. van Es, and Marian Counihan. Children's strategy use when playing strategic games. *Synthese*, 191(3):355–370, 2014. 10.1007/s11229-012-0212-x.

[41] Amnon Rapoport and David V. Budescu. Randomization in individual choice behavior. *Psychological Review*, 104(3):603–617, 1997. 10.1037/0033-295X.104.3.603.

[42] Ariel Rubinstein. A typology of players: Between instinctive and contemplative. *The Quarterly Journal of Economics*, 131(2):859–890, 2016.

[43] Itai Sher, Melissa Koenig, and Aldo Rustichini. Children's strategic theory of mind. *Proceedings of the National Academy of Sciences*, 111(37):13307–13312, 2014.

[44] D.O. Stahl and P.W. Wilson. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254, 1995.

[45] Klaas E. Stephan, Will D. Penny, Jean Daunizeau, Rosalyn J. Moran, and Karl J. Friston. Bayesian model selection for group studies. *Neuroimage*, 46(4):1004–1017, 2009.

[46] James Stiller and Robin I.M. Dunbar. Perspective-taking and memory capacity predict social network size. *Social Networks*, 29(1):93–104, 2007. 10.1016/j.socnet.2006.04.001.

[47] Iris van de Pol, Iris van Rooij, and Jakub Szymanik. Parameterized complexity results for a model of theory of mind based on dynamic epistemic logic. In Ramanujam R., editor, *Proceedings of the 15th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 239–248, 2016.

[48] Willem A. Wagenaar. Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin*, 77(1):65–72, 1972. 10.1037/h0032060.

[49] James R. Wright and Kevin Leyton-Brown. Beyond equilibrium: Predicting human behavior in normal-form games. In *Proceedings of the 24th Conference on Artificial Intelligence*, pages 901–907, 2010.

[50] Wako Yoshida, Ray J Dolan, and Karl J Friston. Game theory of mind. *PLoS Computational Biology*, 4(12):e1000254, 2008.