RUNNING HEAD:
Modeling inference of mental states

TITLE:
Modeling inference of mental states: As simple as possible, as complex as necessary

AUTHORS:
Ben Meijering[1] (b.meijering@rug.nl; +31 50 363 7603)
Niels A. Taatgen[1] (niels@ai.rug.nl; +31 50 363 6435)
Hedderik van Rijn[2] (hedderik@van-rijn.org; +31 50 363 6290)
Rineke Verbrugge[1*] (l.c.verbrugge@rug.nl; +31 50 363 6334)

1. Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen, PO Box 407, 9700 AK, Groningen, The Netherlands
2. Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands

* Corresponding author

Abstract

Behavior oftentimes allows for many possible interpretations in terms of mental states, such as goals, beliefs, desires, and intentions. Reasoning about the relation between behavior and mental states is therefore considered to be an effortful process. We argue that people use simple strategies to deal with high cognitive demands of mental state inference. To test this hypothesis, we developed a computational cognitive model, which was able to simulate previous empirical findings: In two-player games, people apply simple strategies at first. They only start revising their strategies when these do not pay off. The model could simulate these findings by recursively attributing its own problem solving skills to the other player, thus increasing the complexity of its own inferences. The model was validated by means of a comparison with findings from a developmental study in which the children demonstrated similar strategic developments.

Keywords

Theory of mind; developmental study; computational cognitive model; two-player games; strategies.

Introduction

In social interactions, we try to understand others' behavior by reasoning about their goals, intentions, beliefs, and other mental states. Such reasoning about mental states is often called *theory of mind*, abbreviated ToM (Baron-Cohen, Leslie, & Frith, 1985; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). There has been an ongoing debate among philosophers and cognitive scientists whether our everyday reasoning about mental states constitutes a theory, as claimed by the 'theory-theorists' or rather that people directly simulate others' mental states, as claimed by the 'simulation-theorists'. In this work, even though we do use the term 'theory of mind', we do not adhere to either of the two claims.

When we ascribe a simple mental state to someone, we apply first-order theory of mind. For example, if Bob thinks: "Ann *knows* that my birthday is tomorrow", he is applying first-order theory of mind, which covers a great deal of daily social interactions. However, first-order theory of mind does not suffice for reasoning about more complex situations. If Carol thinks: "Bob *thinks* that Anne *knows* that his birthday is tomorrow", she is making a second-order mental state attribution, as she does if she thinks: "Bob *believes* that Anne *intends* to throw him a surprise party".

In this work, we present a computational cognitive model that simulates application of various ToM strategies, ranging from simple strategies to full-blown recursive ToM. ToM has been modelled before (Hiatt & Trafton, 2010; Van Maanen & Verbrugge, 2010). However, these models either simulated one specific instance of ToM (Hiatt & Trafton, 2010) or attributed too much rationality to human reasoning (Van Maanen & Verbrugge, 2010).[1] Our model is based on previous empirical results (Meijering, Van Maanen, Van Rijn, & Verbrugge, 2010; Meijering, Van Rijn, Taatgen, & Verbrugge, 2011) and is validated by means of a re-analysis of a previous developmental study by Flobbe et al. (2008). The model can explain why people use strategies that are relatively simple, while still being successful at inferring mental states of others.

Many studies have shown that people cannot always account for another's mental states in order to predict their behavior, particularly in the context of two-player sequential games (e.g., Flobbe et al., 2008; Hedden & Zhang, 2002; Raijmakers, Mandell, Van Es, & Counihan, 2013; Zhang, Hedden, & Chia, 2012). Sequential games require reasoning about complex mental states, because Player 1 has to reason about Player 2's subsequent decision, which in turn is based on Player 1's subsequent decision (Figure 1). Typically, performance is suboptimal and that is probably because players do not have a correct model of the other player's mental states (Johnson-Laird, 1983). By means of hypothesis testing, they may try to figure out which model works best in predicting the other player's behavior (Gopnik & Wellman, 1992; Wellman et al., 2001). However, a particular

[1] Note that in the remainder of this article we use 'simulate' in its usual meaning in the field of computational cognitive modeling as 'fitting well to the experimental data', so not in the sense of ToM as understood by simulation-theorists.

action or behavior can have many possible mental state interpretations (Baker, Saxe, & Tenenbaum, 2009), and testing all these interpretations strains our cognitive resources.

To alleviate cognitive demands, people generally start testing simple models or strategies that have been proven successful before (Todd & Gigerenzer, 2000). Because application of ToM and especially recursive ToM is an effortful process (Keysar, Lin, & Barr, 2003; Lin, Keysar, & Epley, 2010; Qureshi, Apperly, & Samson, 2010), reasoning about mental states probably also comprises the use of simple strategies. So where do these strategies come from? We hypothesize that they are a legacy of our childhood years, possibly developed in other domains, where they have proven themselves successful. Raijmakers et al.'s (2013) findings corroborate this claim, as the children in their study consistently used strategies that were not fit to deal with the logical structure of the games presented to them. The strategies sometimes did yield the best possible outcome, however, which may be an explanation for why they still exist in adult reasoning: Simple strategies do not exhaust cognitive resources and are appropriate in a wide range of circumstances. Indeed, our computational cognitive model will show that the presence of simple strategies depends on the proportion of games in which they yield an optimal outcome.

In this study, we present a computational cognitive model that simulates inference of mental states in sequential games. The model initially uses a simple strategy that ignores many task aspects. However, if the model's strategy does not work, it learns to acknowledge that the other player has a role in its outcome. The model will therefore start attributing its own strategy to the other player. We will show that this process can account for the differential learning effects in Meijering et al.'s study (2011), in which participants adopted distinct strategies based on the training regimen that was administered to them. To validate the model, the developmental study of Flobbe et al. (2008) was re-analyzed, searching for patterns that are indicative of the use of simple strategies in children.

Before we explain the model, we will first explain the empirical findings on which it is based.

Empirical findings

Meijering et al. (2011) studied second-order ToM reasoning in two-player sequential games. Take the game in Figure 1 as an example game: Each end node contains a pair of payoffs, left-side payoffs belonging to Player 1 and right-side payoffs belonging to Player 2. The end node in which a game is stopped determines the payoff each player obtains in that particular game. Each player's goal is to obtain his or her greatest attainable payoff. Players take turns in making a decision, and Player I begins. Because each player's outcome depends on the other player's decision, both players have to reason about one another's mental states.

These two-player games might appear at first sight to be similar to the Prisoner's Dilemma (PD), because in PD there are also two players whose outcomes depend on the other's decisions. However, PD is a simultaneous-move game, whereas in sequential games, players take turns. The turn-taking aspect of sequential games naturally delineates the required depth of ToM reasoning. In PD,

in contrast, there is not necessarily a limit on ToM depth (see Osborne & Rubinstein, 1994), which complicates the analysis of human ToM reasoning in PD.

In our turn-taking game, participants were always assigned to the role of Player 1, and had to decide at the first decision point whether to stop the game in A or to continue it to the next decision point. To make that decision, they had to reason about what Player 2 would do at II, and because Player 2, in turn, would have to reason about what Player I would do at III, participants had to apply second-order ToM.

It might at first sight appear that participants could be using alternative strategies that do not concern the opponent's mental states at all; for example, they might be applying backward induction, which essentially consists of three pairwise comparisons of pay-offs, from the leaves of the game tree upwards. Our previous studies Meijering et al. (2012, 2013) and Bergwerff et al. (2014), however, support the assumption that participants are indeed imputing mental states such as beliefs and plans to their opponent, and that they also reason about the opponent's beliefs about their own plan of action. Reaction-time and accuracy measures for a similar turn-taking game in Meijering et al. (under submission) also show that a second-order prediction requires significantly more cognitive resources than a first-order decision, even if they correspond to exactly the same payoff comparisons.

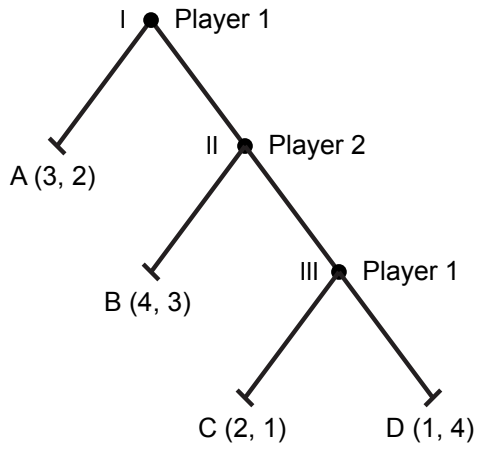

Figure 1: An extensive form representation of a two-player sequential game. Player 1 decides first, Player 2 second, and Player 1, again, third. The decision points are indicated in Roman numerals (I – III). Each end-node has a pair of payoffs, of which the left-side is Player 1's payoff and the right-side Player 2's payoff. Each player's goal is to obtain their highest possible payoff. In this particular game, the highest possible payoff for Player 1 is a 4, which is obtainable because Player 2's highest possible payoff is located at the same end node (i.e., B). Player 2's payoff of 4 is not obtainable because Player 1 would decide *left* instead of *right* at the third decision point (III).

Experimental setup

Meijering et al. (2011) investigated the effect of two distinct training regimens. The training regimens were followed by two test blocks of 32 games each, which were administered to test for longer-term effects of training regimen.

One training regimen was based on the training phase in Hedden and Zhang's (2002; 2012) study, which consisted of 24 so-called trivial games (Figure 2; top panel). In these games, Player 2 did not necessarily have to reason about Player 1's last possible decision, because Player 2's payoff in B was either lower or higher than both his payoffs in C and D. Consequently, Player 2 did not have to apply ToM, and Player 1 had to apply first-order ToM at most.

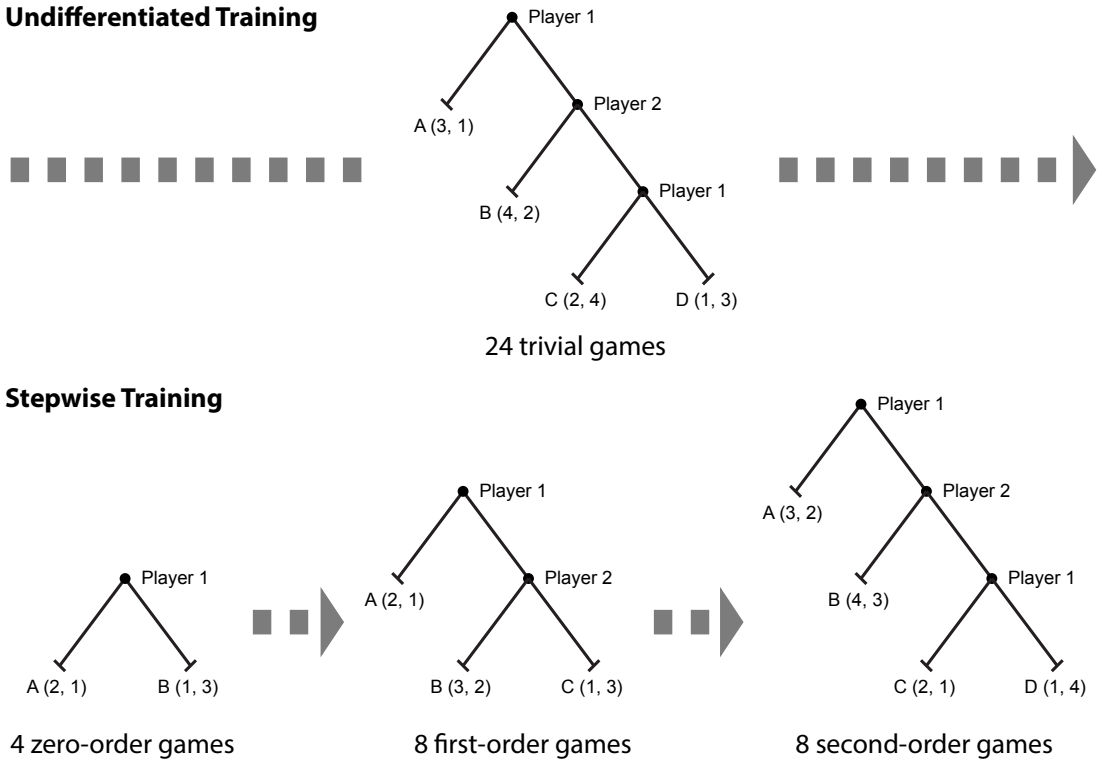

Figure 2: Schematic overview of the two training regimens (see Figure 1 for a detailed explanation of a single game). *Undifferentiated training* consisted of 24 trivial games. *Stepwise training* consisted of 4 zero-order, 8 first-order, and 8 second-order games. Each game had a unique distribution of payoffs.

Although it is not uncommon to include 'simple' practice items, Meijering et al. (2011) believed that the trivial games might have had the opposite effect of familiarization (see Flobbe et al., 2008 for similar concerns). Participants might have adopted first-order reasoning during training. Such reasoning would yield expected outcomes during the training phase, but not anymore during the experimental phase. The games in the latter phase required second-order ToM, and first-order ToM would yield suboptimal outcomes. To test the hypothesis that the practice regimen might have entrained participant to use first-order ToM,

Meijering et al. contrasted Hedden and Zhang's training phase with another training regimen.

In the other training regimen, participants were presented with subsequent blocks of games of increasing complexity (see Figure 2; bottom panel). Each block (three in total) introduced a new decision point, thus increasing the required ToM reasoning. This regimen allowed for adoption of simple strategies at first, and gradually more complex strategies in subsequent blocks. Crucially, the games in the last training block required second-order ToM, and simpler strategies would not suffice anymore. This training regimen is henceforth referred to as Stepwise training; Hedden and Zhang's training regimen is henceforth referred to as Undifferentiated training.

Meijering et al. (2011) hypothesized that these two training regimens would have distinct effects on strategy formation and performance. They predicted that Stepwise training would facilitate participants to incorporate mental states of increasing complexity into their decision making process, yielding high accuracy. Undifferentiated training, in contrast, would not motivate participants to develop recursive ToM, as they could suffice with application of first-order ToM.

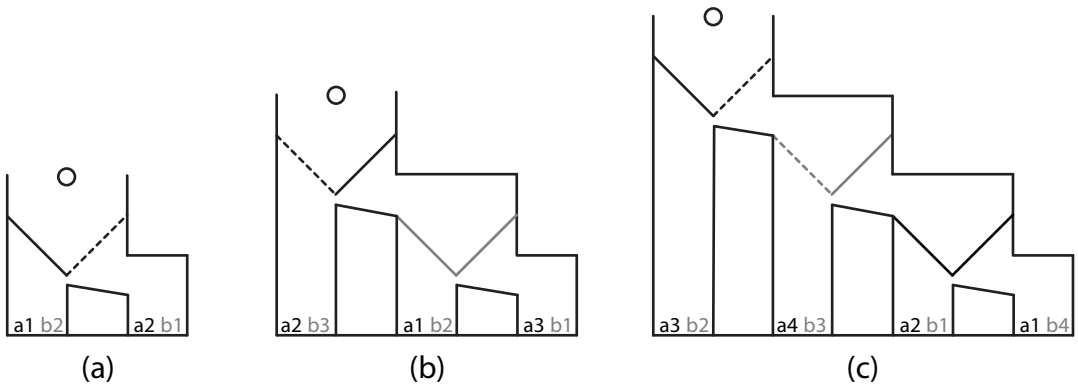

Figure 3: A zeroth-order (a), first-order (b) and second-order (c) Marble Drop game. The participant's payoffs are represented by a1 – a4, the computer's by b1 – b4, both in increasing order of value. The goal is to let the white marble end up in the bin with the highest attainable payoff. The diagonal lines represent trapdoors. At the first set of trapdoors, the participant decides which of both trapdoors to remove, at the second set the computer decides, and at the third set the participant again decides. The dashed lines represent the trapdoors that both players should remove to attain the highest payoff they can get.

Meijering et al. (2010; 2011) designed a new visual task representation to facilitate reasoning in sequential games. Importantly, this representation did not change the logical structure of the sequential games, which were still equivalent to Hedden and Zhang's. Figure 3 shows examples of the new task representation, which is called Marble Drop. Note that in this game, both players are aware of the pay-off structure for the whole game. In this respect, Marble Drop differs from well-known large turn-taking games such as chess and checkers.

Results

As expected, the participants that were assigned to Stepwise training performed better than the participants assigned to Undifferentiated training (see Figure 6). One specific behavioral pattern is of particular interest to validate the model: The performance of participants assigned to Undifferentiated training rose to ceiling during the training phase and dropped again when the experimental phase started (Figure 6). We hypothesize that the participants applied simple, child-like strategies during the training phase, because these strategies worked and did not consume much cognitive resources. At the start of the experimental phase, however, these strategies did not work anymore and accuracy dropped, because the games, while superficially similar, required more complex reasoning. Nevertheless, accuracy increased again over the course of the experimental phase, as the participants were able to revise their strategies. We will show that our computational cognitive model can simulate this process: The model's most important characteristic is that the complexity of its reasoning gradually increases by repeatedly attributing its own (evolving) strategy to the other player.

Computational cognitive model

The model[2] is implemented in the ACT-R cognitive architecture (Anderson, 2007; Anderson et al., 2004). ACT-R comprises a production system, which executes if-else rules, and contains declarative knowledge, which is presented as memory representations, or so-called chunks. In addition, ACT-R also includes modules that simulate specific cognitive functions, such as vision and attention, declarative memory, motor processing, et cetera. The results of these simulations appear as chunks in the modules' associated buffers, which the model continually checks (and manipulates) by means of its production system. ACT-R imposes natural cognitive constraints, as buffers can hold just one chunk at a time, and production rules can only fire successively, whenever their pre-specified conditions are matched. ACT-R does allow for parallel processing whenever a task induces cognitive processing in distinct modules. The model that we present here runs atop of ACT-R.

The model's behavior partially depends on memory dynamics. It needs to retrieve factual knowledge from declarative memory, and both the speed and success of retrieval depend on the so-called base-level activation of a fact (or chunk). The higher the base-level activation is, the greater the probability and speed of retrieval. The base-level activation in turn is positively correlated with the number of times a fact is retrieved from memory and the recency of the last retrieval.

The model simulates inference of mental states in sequential games. It uses a simple strategy at first (explained in the section Assumptions) and gradually revises that strategy until it can process recursive mental states. We consider the application of a particular strategy, and revising that strategy, to be deliberate

[2] The model can be downloaded from http://www.ai.rug.nl/~meijering/iccm2013

processes. Therefore, application and revision are implemented by means of an interaction between factual knowledge and problem solving skills. Arslan, Taatgen, and Verbrugge (2013) successfully used a similar approach in modeling the development of second-order ToM in another ToM paradigm (i.e., the false-belief task). Van Rijn, Van Someren, and Van der Maas (2003) have successfully modeled children's developmental transitions on the balance scale task in a similar vein. Factual knowledge is represented by chunks in declarative memory, which store what strategy the model should be using. The problem solving skills, or strategy levels, are executed by (recursively) applying a small set of production rules. The model's goal is to make decisions that yield the greatest possible payoff. Decisions are either 'stop the game' or 'continue it to the next decision'. The model was presented with the same distributions of payoffs (i.e., items) as were presented to the participants.

The model's initial simple strategy is to consider only its own decision at the first decision point and to disregard any future decisions. The model's decision is based on a comparison between its (i.e., Player 1's) payoff in A and the maximum of its payoffs in B, C, and D. If the model's payoff in A is greater, the model will decide to stop. Otherwise, the model will decide to continue. By using this simple strategy the model seeks to maximize its own payoff, which can be considered a direct translation of the instructions given to the participants.

This strategy will work in some games but not in all. Whenever the strategy works, the model receives positive feedback and stores in declarative memory what strategy it is currently using. In fact, the model stores a strategy level, which is level-0 in the case of the simple strategy described above. Whenever the strategy does not work, the model receives negative feedback and stores in declarative memory that it should be using a higher strategy level (e.g., level-1).
The higher strategy level means that the model should attribute whatever strategy it was using previously to the other player at the next decision point. In the case of strategy level-1, the model attributes the model's initial simple strategy (i.e., level-0) to Player 2. Accordingly, the model is applying first-order ToM, as it reasons about the mental state of Player 2, who considers only his own payoffs and disregards any future decisions.

Again, this strategy will work in some games but not in all. Whenever it does not work, the model receives negative feedback and stores in declarative memory that it should be using a higher strategy level. At a higher strategy level, the model will attribute whatever strategy level it was using previously to Player 2. At strategy level-2, the model attributes strategy level-1 to Player 2, who in turn will attribute strategy level-0 to the player deciding at third decision point: Player 1. Now the model is applying second-order ToM.

Assumptions

The model is based on three assumptions. The first assumption is that participants, unfamiliar with sequential games, start playing according to a simple strategy that consists of one comparison only: Participants compare their current payoff, when stopping the game, against the maximum of all their future payoffs, when continuing the game. They stop if the current payoff is highest; they continue otherwise. Participants who are using this strategy ignore the consequences of any

possible future decision, whether their own or the other player's, hence the label 'simple strategy'.

The simple strategy can also be phrased in terms of risk attitudes that bias one's expectations of a certain outcome. Expectations might either be too liberal or too conservative, causing risk-seeking or risk-averse strategies. Importantly, risk attitudes are prevalent in both children and adults. For example, Harbaugh, Krause, and Vesterlund (2002) have investigated the role of risk attitudes in choices between gambles and certain outcomes, and they have shown that children are risk seeking when faced with high-probability prospects of gains. Harbaugh et al. have also found that adults' choices, too, were affected by similar risk attitudes. Importantly, these risk attitudes did not have to be task specific, and could generalize across a multitude of domains.

If participants obtain their expected outcomes, they do not have to revise their strategy. However, participants might obtain unexpected outcomes if the other player's decision is incongruent with their goals. As the games are fully animated, and played from start to end, participants would see that the unexpected turn of events is due to the other player's unexpected decision. Our second assumption is that in future games, participants will acknowledge the role of the other player.

Reasoning about the other player, participants can only attribute a strategy they are familiar with themselves. This is our third assumption, which is based on variable frame theory (Bacharach & Stahl, 2000). Imagine a scenario in which two persons are asked to select the same object from a set of objects with differing shapes and colors but one person is completely colorblind. The colorblind person cannot distinguish the objects based on color, nor can he predict how the other would do that. Therefore, the colorblind person can only predict or guess what object the other would select based on which shape is the least abundant. The seeing person should account for the colorblind person's reasoning and also choose the object with the least abundant shape. This variable frame principle also applies to reasoning about others: We can only attribute to others goals, intentions, beliefs, and strategies that we are familiar with ourselves.

Mechanisms

The simple strategy is implemented in two production rules. The first production rule determines what the payoff will be when stopping the game; the other production rule determines what the highest future payoff could possibly be when continuing the game. Both productions are executed from the perspective of whichever player is currently deciding (Figure 4). The model will attribute this simple strategy from the current decision point to the next, each time the model updates its strategy level (i.e., incrementing strategy level by one). The model will thus heighten its level, or order, of ToM reasoning.

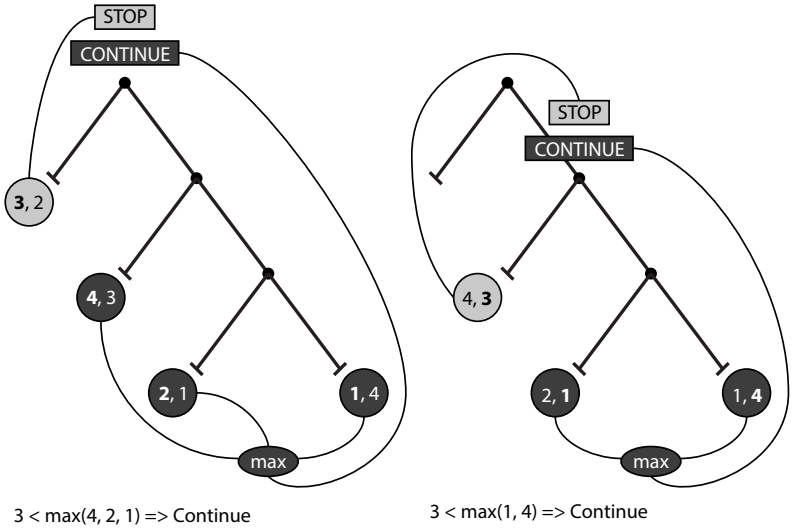

Figure 4: Depiction of the simple strategy. In the left panel, the model compares its payoff if it would stop (grey) against its maximum possible payoff if it would continue (black). In the right panel, the model compares Player 2's payoff if Player 2 would stop (grey), against Player 2's maximum possible future payoff (black). The left panel schematically represents the application of zero-order ToM, and the right panel the attribution of zero-order ToM to the other player.

Zero-order ToM

Before the model starts applying its strategy, it needs to construct a game state representation to store the payoffs that are associated with a *stop* and *continue* decision, respectively. To construct a game state, the model first retrieves from declarative memory what strategy level it is currently using. At the beginning of the experiment, strategy level has a value of 0, which represents the simple strategy. After retrieving strategy level, the model constructs its current game state.

Starting with the simple strategy, the model will determine its own *stop* and *continue* payoffs (see Figure 4, left panel), which will be stored in the game state representation. The model will then compare these payoffs and make a decision. After the model has made a decision, it will update declarative memory by storing what strategy level the model should be playing in the next game: If the model's decision was correct, the model should continue playing its current strategy level; otherwise the model should be playing a higher strategy level.

After playing a couple of games in which the simple strategy (i.e., level-0) does not work, the higher strategy level (i.e., level-1) will have a greater probability of being retrieved, as its base-level activation increases more than the simple strategy's base-level activation. At the start of the next few games, before the model constructs its game state, it will begin retrieving strategy level-1 from declarative memory.

First-order ToM

Playing strategy level-1, the model will first determine what payoff is associated with a *stop* decision at the first decision point (I). However, before determining what payoff is associated with a *continue* decision, the model needs to reason about the future and therefore consider the next decision point (II). It attributes strategy level-0 to Player 2, who is deciding at II. Later, the model will return to the first decision point and determine what payoff is associated with a *continue* decision.

At II, the model will apply strategy level-0, but from the perspective of Player 2 (Figure 4, right panel). When reasoning about Player 2's decision, the model constructs a new game state, which references the previous one. The previous game state is referenced, because the model needs to jump back to that game state and determine what payoff is associated with a *continue* decision in that game state. At II, the model will execute the same production rules that it executed before when it was playing according to strategy level-0: It will determine what payoffs are associated with a *stop* and a *continue* decision, but from the perspective of Player 2.

The model will not produce a response whenever it determines the *stop* and *continue* payoffs at II, because the problem state at II references a previous one (i.e., I). The model will therefore backtrack to the previous game state representation, which did not yet have a payoff associated with a *continue* decision. That payoff can now be determined based on the current game state (i.e., Player 2's decision). The model will retrieve the previous game state from declarative memory.

After retrieving the previous game state representation, the model has two game states stored in two separate locations, or buffers: The current game state is stored in a so-called *problem state* buffer, which stores intermediate results (Anderson, 2007; Borst, Taatgen, & Van Rijn, 2010); the previous game state is stored in the *retrieval* buffer, which belongs to the declarative memory module. The model will determine what payoff is associated with a *continue* decision in the previous game state (stored in the *retrieval* buffer) given the decision based on the current game state (in the *problem state* buffer). It will update the previous game state and store it as an intermediate result in the *problem state* buffer.

Playing strategy level-1 and being back in the previous game state, there is no reference to any previous game state and the model will make a decision based on a comparison between the payoffs associated with the *stop* and *continue* decisions. As explained previously, the model will stop if the payoff associated with stopping is greater; otherwise the model will continue.

Again, after the model has made a decision, it will update declarative memory by storing what strategy level the model should be playing in the next game(s). If the model's decision is correct, it will apply the current strategy level. Otherwise, the model will revise its strategy level by storing in declarative memory that it should be using strategy level-2 in the next game(s).

Second-order ToM

The model will first determine what payoff is associated with stopping the game and then consider the next decision point. There, the model proceeds as if it were

playing strategy level-1, but from the perspective of Player 2. In other words, the model is applying second-order ToM.

The strategy described above closely fits the strategy of *forward reasoning plus backtracking* (Meijering, Van Rijn, Taatgen, & Verbrugge, 2012). Meijering et al. (2012) conducted an eye-tracking study, and participants' eye movements reflected a forward progression of comparisons between payoffs, followed by backtracking to previous decision points and payoffs. Such forward and backward successions are present in strategy level-2 as well: Payoffs of *stop* decisions are determined one decision point after another, and this forward succession of payoff valuations is followed by backtracking, as payoffs of previous *continue* decisions are determined in backward succession.

Results

The model was presented with the same trials as in Meijering et al.'s (2011) study, with stepwise training versus undifferentiated training as a between-subjects factor. The model was run 100 times for each training condition. Each model run consisted of 20 (stepwise) or 24 (undifferentiated) training games, followed by 64 truly second-order games. The results are presented in Figures 4 and 5.

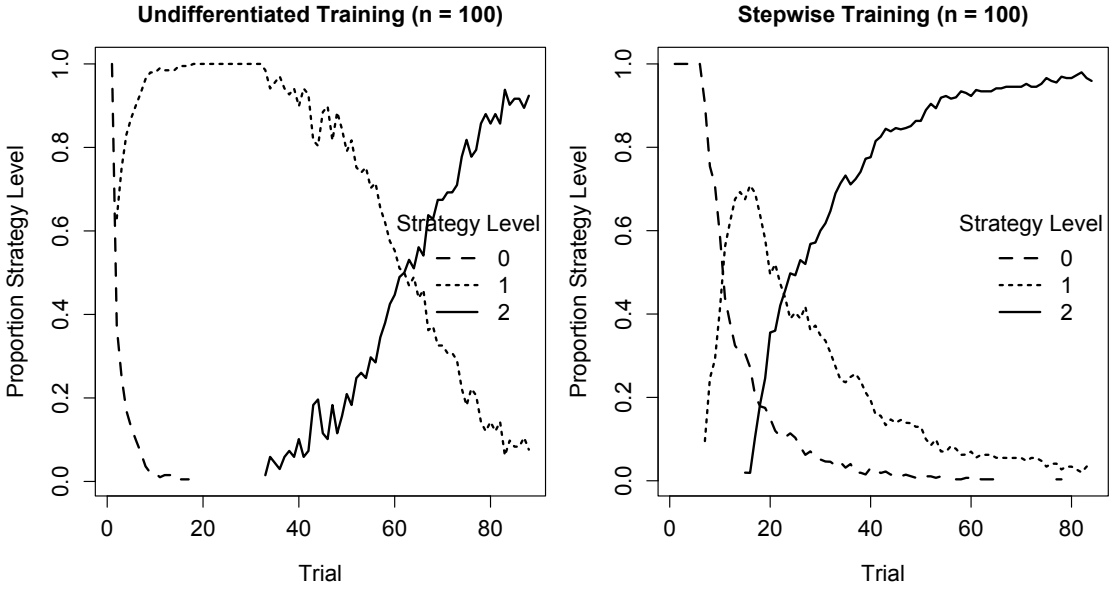

Figure 5: Proportion of models that apply strategy levels 0, 1, and 2; plotted as a function of trial. The left panel depicts these proportions for the model that received undifferentiated training; the right panel depicts the proportions for the model that received stepwise training.

Figure 5 shows the proportions of models that apply strategy levels 0, 1, and 2, calculated per trial. The left panel of Figure 5 shows the output of the models that received 24 undifferentiated training games before playing 64 second-order games. As can be seen, initially all models apply strategy level-0, corresponding with zero-order ToM, but that proportion decreases quickly in the first couple of games. The proportion of models applying zero-order ToM decreases because that

strategy yields too many errors, which can be seen in Figure 6. The models store in declarative memory that they should be using strategy level-1, but it takes a few games before the base-level activation of the level-0 chunk drops below the retrieval threshold. After it does, the models start retrieving level-1 chunks and will apply strategy level-1, which corresponds with first-order ToM. The proportion of models that use strategy level-1 increases up to 100% towards the end of the 24 undifferentiated training games. The models do not start applying strategy level-2 during the training phase, because strategy level-1 yields correct decisions in all undifferentiated training games, which can be seen in Figure 6. However, in the experimental games, which are truly second-order games, strategy level-1 yields too many errors, and accuracy drops. It takes approximately 40 games before the base-level activation of the level-1 chunk has dropped below the threshold in at least half of the models. The models gradually start using strategy level-2, and accuracy starts to increase again, as can be seen in Figure 6.

The right panel of Figure 5 shows the output of the models that were presented with 20 stepwise training games (4 zero-order, 8 first-order, and 8 second-order games) before playing 64 second-order games during the experimental phase. As can be seen, all models start applying strategy level-0, and they use it longer than the models that received undifferentiated training. The reason is that strategy level-0 yields a correct answer in the first four games during stepwise training, because those are zero-order games. As can be seen in Figure 6 (right panel), accuracy is 100% in the first few games. In the next eight first-order training games (Trials 5 – 12), the proportion of models that apply strategy level-0 decreases, as strategy level-0 yields too many errors. Simultaneously, the proportion of models applying strategy level-1 increases, as the base-level activation of the level-0 chunk decreases and the models start retrieving the level-1 chunk. In the next eight second-order training games (Trials 13 – 20), the proportions of models that apply strategy level-0 and level-1 decrease, as both strategy levels yield too many errors. Simultaneously, the proportion of models that apply strategy level-2 increases. As strategy level-2 yields a correct decision in the remainder of the games, accuracy increases up to ceiling, which can be seen in Figure 6 (right panel).

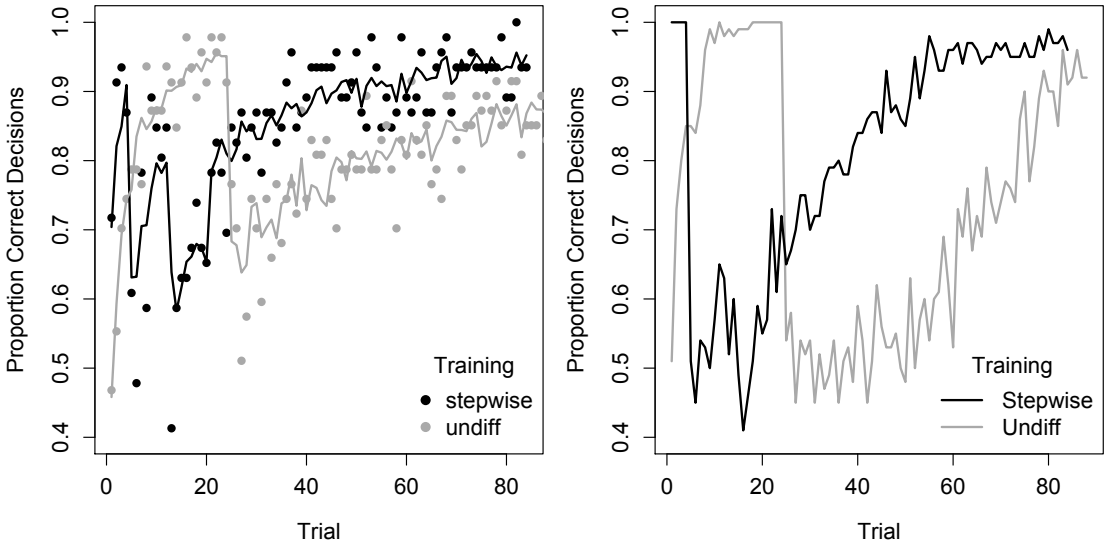

Figure 6: Proportion of correct decisions, or accuracy, across participants (left panel) and models (right panel). The solid lines in the left panel represent the fit of the statistical model, which is added to visualize the proportion trends.

The accuracy trends in the models' output qualitatively fit those of Meijering et al.'s study (2011). The quantitative differences are probably due to the fact that not all participants started out using the simple strategy, whereas all models did. One possible explanation is that some participants started with intermediate-level strategies and, due to large proportions of optimal outcomes, did not proceed to the highest level of reasoning. We could account for this by storing level-0, level-1, and level-2 chunks in declarative memory, and having the base-level activation of these chunks follow the distribution of zero-order, first-order, and second-order ToM in the adult population. A meta-review of (higher-order) ToM in adults and children may be a good starting point to find the appropriate distributions. Nevertheless, the qualitative trends in the model data, changing as a function of game complexity, correspond with the response patterns in the behavioral data. The trends suggest that people use simple strategies for as long as these yield expected outcomes.

In the introduction we hypothesized that simple strategies are a legacy of our childhood years, and that adults keep using those strategies that have proven themselves successful during development. To test this hypothesis, we have re-analyzed the data from Flobbe et al.'s (2008) developmental study. We expected that few children would have sufficient cognitive resources to apply second-order ToM, and that performance levels would therefore align well with lower and intermediate strategy levels. The most obvious prediction is that prevalence of level-0, level-1, and level-2 strategies can be ranked, where level-0 is the most dominant strategy and level-2 is least frequent.

Developmental study

Flobbe et al. (2008) studied the application of second-order ToM in children that were in between 8 and 10 years (*M* = 9;2). They presented the children with sequential games that were game-theoretically equivalent to the game we use in the current study (see Fig. 1), but they used a different cover story. The child was told that she and the computer opponent would jointly control a car. The current position in the game was represented by the location of the car. Decision points were represented by road junctions. Each junction was marked with a color (blue for the child, yellow for the computer) to show who could decide there. Leaves were represented by dead ends. Each dead end contained a reward for the child (a number of blue marbles) as well as one for the computer opponent (a number of yellow marbles). The rewards at a dead end could be different for each player, and the rewards to be amassed at each dead end differed. Crucially, all rewards were visible throughout the entire round of the game (car ride). The reward for a player consisted of 1, 2, 4, or 7 marbles. These numbers were chosen to make the payoffs easy to distinguish visually and to eliminate the need for counting. At the junctions, the child and the computer opponent would alternately decide either to turn to a dead end, where both 'drivers' would receive their rewards, or to continue on the main road, so that other rewards at subsequent dead ends could be reached. The child was told to maximize her own reward (i.e., the number of blue marbles), and was told that the opponent would try to maximize the number of yellow marbles.

Performance in this 'road game' was just above chance-level (57% correct). As children of age 9 can correctly apply second-order theory of mind in false belief tasks and are at the brink of mastering application of second-order ToM in other contexts (Flobbe et al., 2008; Miller, 2009; Perner & Wimmer, 1985), we expect the lower and intermediate strategies to be most prevalent in Flobbe et al.'s study, which is thus perfect to validate our model.

We hypothesize that children apply the same simple strategies that are implemented in our computational cognitive model. We predict that the children start out with the simplest (i.e., zero-order) strategy, and that some will learn to attribute that strategy to the other player. Probably few children will learn that the other player, in turn, attributes the simple strategy to the player who decides next (i.e., to them). As each child was first asked to predict the other player's decision, before they were asked to make a decision themselves, we have a direct measure of the child's perspective of the other player's strategy. We will analyze both the predictions and the decisions.

Predictions

We applied a binomial criterion to reliably categorize a participant's predictions as belonging to either level-1 or level-2: The predictions in at least 8 out of 10 consecutive games had to be congruent with one particular strategy level to label the predictions accordingly. This might seem strict, but 8 out of 10 is the minimum quantile that is still significant with a significance level of 0.05. As the experiment consisted of 40 second-order games, we categorized each child's responses in 4 sets of 10 games. Figure 7 depicts the proportion of children that applied either first-order or second-order ToM. These ToM-orders correspond with level-1 and level-2 in the computational model.

Note that sets of predictions that could not be categorized level-1 or level-2 do not necessarily imply the use of level-0, because the predictions in those sets could have been completely random, or a mixture of the various strategy levels. The decisions are therefore analyzed to determine the prevalence of strategy level-0.

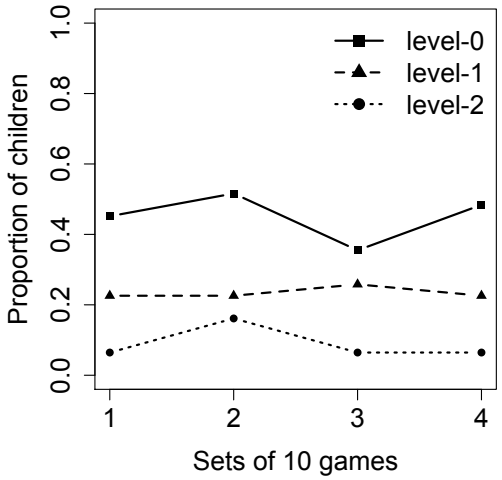

Figure 7: The proportion of children that applied zero-order ToM (level-0), first-order ToM (level-1), or second-order ToM (level-2) to the other player; depicted in 4 consecutive sets of 10 games.

As can be seen in Figure 7, the proportion of children that applied first-order ToM by attributing strategy level-0 to the other player is greater than the proportion of children that applied strategy second-order ToM. Furthermore, many children's predictions could not be labeled according to one of the strategies at all (13 out of 40). These children probably switched frequently between multiple possible perspectives, and such switching is difficult to reliably capture by means of a statistical model. However, Bayesian model selection could be used to determine which strategies are most likely (see for example Lee & Sarnecka, 2010). Nevertheless, most of the children whose responses could be categorized, were applying first-order ToM by attributing the simple (i.e., level-0) strategy to the other player. Almost none of the children was able to consistently attribute strategy level-1 to the other player, thereby applying second-order ToM.

Decisions

As explained above, the predictions required application of first-order ToM at minimum and could therefore not be indicative of zero-order ToM. Therefore, the decisions were analyzed to determine how many children applied zero-order ToM, ignoring the other player entirely. Again, we categorized the decisions based on the binomial criterion that at least 8 out of 10 consecutive responses should be consistent with application of zero-order ToM (i.e., level-0 in the model). As can be seen in Figure 7, most of the children that consistently responded according to one of the strategies applied zero-order ToM when making a decision. This is remarkable, because each child that participated in the experimental phase successfully passed a training block in which they were required to apply first-order ToM. This finding suggests that the children could not see how first-order

ToM would fit in the more complex games in the experimental blocks. They may have recognized that it did not work, but still could not revise their strategy to incorporate an additional ToM level.

To conclude, a re-analysis of Flobbe et al.'s (2008) study shows that few children were able to apply second-order ToM (level-2), and that most children used simple strategies. The most dominant strategy was the simplest one that did not account for any future decision points. Most children seemed to apply zero-order ToM (level-0) while making a decision. Some children, though, were able to attribute that simple strategy to the other player, thereby applying first-order ToM (level-1). These strategies are the same as those implemented in our computational cognitive model. The model is thus supported in two ways: (1) Its most simple strategies are found in children, and (2) it learns to revise its strategies as adults do.

Conclusions

In this study we presented a computational cognitive model that simulates inference of mental states in sequential games. More specifically, the model was required to apply ToM recursively, a skill that appears to be unique to human intelligence. Many studies have shown that people oftentimes fail to apply ToM to interpret the behavior of others (e.g., Apperly et al., 2010; Keysar et al., 2003; Lin et al., 2010). In this study, in contrast, we show that people do not necessarily fail to apply ToM, but rather first apply simple strategies that are computationally less costly. Only when necessary do people revise their strategies to account for complex mental states.

The model is based on previous empirical findings (Meijering et al., 2011) that seemed to imply that people exploit the possibility of using simple strategies for as long as these pay off. We implemented one such simple strategy that ignores any future decisions and simply compares the immediate payoff, when stopping a game, against the maximum of all future possible payoffs. By means of simple memory dynamics the model either retrieves a chunk that specifies that the model should continue using this strategy, or chunks that specify that the model should attribute the simple strategy to the player who decides next. Although this updating process may seem simplistic at first sight, the model does gradually master second-order ToM, but only because that is required in the games in this study. In other words, the model's most important dynamics are not task-specific, and because of that, the model is flexible and can accommodate many other two-player sequential games.

We found support for the model in the data from Flobbe et al.'s (2008) developmental study in which 9-year-old children were presented with similar sequential games. Most children used the simple, level-0, strategy when making a decision. The second-most prevalent strategy was the level-1 strategy. Using that strategy, the children attributed the simplest possible strategy (i.e., level-0) to the other player. Few children were able to apply second-order ToM mind. They did not recognize that the other player, in turn, attributed the simplest strategy (i.e., level-0) to them. These findings show that the children used the same simple strategies as the adults initially used in Meijering et al.'s study. However, the adults

were able to revise their strategies to achieve the highest required level of ToM reasoning, whereas the children may not have had sufficient cognitive resources to achieve that same level of reasoning. This interpretation is supported by the study of Omaki et al. (2013), indicating that children of around 5-6 years of age have trouble revising their interpretations when incrementally interpreting complex questions such as "Where did Emily tell someone that she hurt herself?"

Our notion of zero-order ToM (i.e., strategy level-0) closely maps with the instruction given to the participants: to maximize their payoff. This strategy corresponds with a risk-seeking perspective, because it does not account for the fact whether higher future payoffs are actually attainable. There are other notions of a level-0 strategy, however. A risk-seeking strategy can be contrasted with a risk-aversive strategy according to which one would stop if there were any lower future payoffs. There is still another notion of a level-0 strategy: Hedden and Zhang (2002; 2012) defined a so-called myopic level-0 strategy that only considers the current payoff and the closest future payoff. Player 1, for example, would only compare his payoffs in A and B, ignoring his payoffs in C and D. These strategies, however, are almost non-existent in Flobbe et al.'s dataset.

This study has at least two methodological implications: One, experimenters should be careful in selecting 'practice' items, as participants exploit the possibility of using simple strategies when possible. Two, average proportions of correct answers, a popular statistic in most ToM studies, may not be as informative as a categorization of responses (also see Raijmakers et al., 2013). Flobbe et al., for example, reported that performance was just above chance-level (i.e., 57% correct), and the most common interpretation would be "on average children were able to apply second-order ToM in 57% of the games." However, the current study shows that this score can be obtained if 1 or 2 children are applying second-order ToM and most of them below-optimal strategies such as zero-order and first-order ToM.

The theoretical implication of this study is that people do not necessarily perceive sequential games in terms of interactions between mental states. They know that there is another player making decisions, but they have to learn over time, by playing many games, that the other player's depth of reasoning could be greater than initially thought. Learning takes place when people obtain unexpected outcomes and start recognizing that the other player has a role in their outcomes. They will have to attribute their own, simple, strategies to the other player, thereby developing increasingly more complex strategies themselves. Over time, reasoning will become as complex as necessary, as simple as possible.

Acknowledgments

This study was supported by the Vici grant "Cognitive systems in interaction: Logical and computational models of higher-order social cognition" from the Netherlands Organization of Scientific Research (NWO), which was awarded to Rineke Verbrugge (grant 227-80-001). We would like to thank the three anonymous reviewers for their useful suggestions for improvements. Finally, we are grateful for several rounds of feedback from special issue editor Marjorie McShane, who constantly kept the possible future readers' mental states in mind, without losing track of the mental states of the authors.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffit, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217.

Arslan, B., Taatgen, N. A., & Verbrugge, R. (2013). Modeling developmental transitions in reasoning about false beliefs of others. In *Proceedings of the 12th International Conference on Cognitive Modelling* (pp. 77–82). Ottawa, CA: Carleton University.

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

Baker, C., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind?" *Cognition*, *21*(1), 37–46.

Bergwerff, G., Meijering, B. , Szymanik, J. Verbrugge, R. & Wierda, S. M. (2014). Computational and algorithmic models of strategies in turn-based games. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, memory, and cognition*, *36*(2), 363–382.

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.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind and Language*, *7*(1-2), 145–171.

Harbaugh, W. T.,Krause, K., & Vesterlund, L. (2002). Risk attitudes of children and adults: Choices over small and large probability gains and losses. *Experimental Economics*, *5*(1), 53-84.

Hedden, T., & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, *85*(1), 1–36.

Hiatt, L. M., & Trafton, J. G. (2010). A cognitive model of theory of mind. In D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the 10th International Conference on Cognitive Modeling* (pp. 91–96). Philadelphia, PA: Drexel University.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behavior in number concept development. *Cognitive Science*, *34*(1), 51-67.

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556.
Meijering, B., Van Maanen, L., Van Rijn, H., & Verbrugge, R. (2010). The facilitative effect of context on second-order social reasoning. In R. Catrambone & S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 1423–1428). Austin, TX: Cognitive Science Society.
Meijering, B., Van Rijn, H., Taatgen, N. A., & Verbrugge, R. (2012). What eye movements can tell about theory of mind in a strategic game. *PloS one*, *7*(9).
Meijering, B., Van Rijn, H., Taatgen, N., & Verbrugge, R. (2011). I do know what you think I think: Second-order theory of mind in strategic games is not that difficult. In L. Carslon, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2486–2491). Austin, TX: Cognitive Science Society.
Meijering, B., van Rijn, H., Taatgen, N. A., & Verbrugge, R. (2013). Reasoning about diamonds, gravity and mental states: The cognitive costs of theory of mind. In: M. Knauff et al. (eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3026-3031). Austin, TX: Cognitive Science Society.
Meijering, B., Van Rijn, H., Taatgen, N. A., & Verbrugge, R. (under submission). Reasoning about self versus others: Changing perspective is hard.
Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*(5), 749–773.
Omaki, A., Davidson White, I., Goro, T., Lidz, J., & Phillips, C. (2013). No fear of commitment: Children's incremental interpretation in English and Japanese Wh-Questions. *Language Learning and Development*, (ahead-of-print), 1-28.
Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. MIT press.
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.
Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236.
Raijmakers, M. E. J., Mandell, D. J., Van Es, S. E., & Counihan, M. (2013). Children's strategy use when playing strategic games. *Synthese*.
Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, *23*(1), 727–780.
Van Maanen, L., & Verbrugge, R. (2010). A computational model of second-order social reasoning. In D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the 10th International Conference on Cognitive Modeling* (pp. 259–264). Philadelphia, PA: Drexel University.
Van Rijn, H., Van Someren, M., & Van der Maas, H. L. J. (2003). Modeling developmental transitions on the balance scale task. *Cognitive Science*, *27*(2), 227–257.
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.
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.
Zhang, J., Hedden, T., & Chia, A. (2012). Perspective-taking and depth of theory-of-mind reasoning in sequential-move games. *Cognitive Science*, *36*(3), 560–573.

Ben Meijering recently acquired his PhD at the Institute of Artificial Intelligence at the University of Groningen. Previously, he obtained an MSc in psychology (Brains and Behavior) and an MSc in Human-Machine Communication. In Rineke Verbrugge's Vici-project, Ben Meijering has investigated higher-order social cognition in adults. His research methods include both behavioral experiments (including eye-tracking) and computational cognitive models.

Niels Taatgen is professor of Cognitive Modeling in the institute of Artificial Intelligence at the University of Groningen. He received his PhD on the topic of "Learning without Limits" in 1999, also at the University of Groningen. From 2003 to 2009 he worked at Carnegie Mellon University as part of the core ACT-R modeling team. He currently specializes in multitasking and skill acquisition, and in the transfer of cognitive skills and studies these using cognitive models, human experimentation and neuroimaging.

Hedderik van Rijn is associate professor of Experimental Psychology at the Department of Psychology at the University of Groningen. He received his PhD from the University of Amsterdam in 2003 and was a postdoc at Carnegie Mellon University from 2002 to 2004. From 2004 onwards, Hedderik van Rijn has worked at the University of Groningen, first at the Institute of Artificial Intelligence and Cognitive Engineering, and since 2009 at the Department of Psychology. He specializes in computational and mathematical models of cognition and experimental psychology and is editor-in-chief of the journal Timing and Time Perception.

Rineke Verbrugge received an MSc (cum laude) and, in 1993, a PhD in Mathematics from the University of Amsterdam. She was postdoc at the Universities of Prague and Gothenburg, and Assistant Professor at MIT and Vrije Universiteit Amsterdam. From 1997, she has worked at the University of Groningen's Institute of Artificial Intelligence and Cognitive Engineering, since 2009 as full professor of Logic and Cognition. Rineke Verbrugge has published a large number of research papers and some books. She is associate editor of the Journal of Logic, Language and Information and has been program chair of several conferences and a large summer school.