Reasoning about mental states in sequential games: As simple as possible, as complex as necessary

Ben Meijering[1] (b.meijering@rug.nl), Niels A. Taatgen[1], Hedderik van Rijn[2], & Rineke Verbrugge[1]

[1]Institute of Artificial Intelligence, PO Box 407, 9700 AK Groningen, The Netherlands
[2]Department of Experimental Psychology, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

Abstract

Because of limited cognitive resources, humans use heuristics in problem solving and decision-making. We argue that reasoning about mental states also comprises the use of heuristics. This has been tested in sequential games, in which one player's outcomes depend on another player's decisions. Empirical findings show that participants apply simple strategies, or heuristics, for as long as these yield expected outcomes. However, participants were able to revise their strategies when presented with superficially similar but more difficult games. We have built a flexible and task-general computational cognitive model that can simulate these findings. The model uses a heuristic for as long as the heuristic yields expected outcomes. If the model's decisions yield suboptimal outcomes, the model updates its strategy level. This updating can be considered a deliberate process, as it is based on an interaction between factual knowledge and problem solving skills.

Keywords: Theory of mind; sequential games; cognitive model; decision making; heuristics.

Introduction

In social interactions we try to predict others' behavior by reasoning about their goals, intentions, beliefs, and other mental states. Reasoning about mental states requires a *theory* of how minds work. This theory has often been referred to as *theory of mind*, abbreviated ToM (Wellman, Cross, & Watson, 2001). ToM has been implemented in computational cognitive models 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). In contrast, we present a model that simulates task-dependent application of various ToM levels, ranging from simple heuristics to recursive ToM.

Many studies show suboptimal reasoning about mental states, particularly in two-player sequential games (e.g., Flobbe, Verbrugge, Hendriks, & Krämer, 2008; Hedden & Zhang, 2002; Raijmakers, Mandell, Van Es, & Counihan, 2013). Sequential games require reasoning about complex mental states, because Player 1 has to reason about Player 2's subsequent decision, for which Player 2 in turn has to reason about Player 1's subsequent decision. A possible explanation for suboptimal performance is that we can never be sure whether our ideas about someone else's mental states are truly correct. By means of hypothesis testing, we try and figure out which theory works best in predicting behavior (Gopnik & Wellman, 1992). However, a particular action or behavior can have many possible mental state interpretations (Baker, Saxe, & Tenenbaum, 2009), and testing these strains our cognitive resources.

To alleviate strain on cognitive resources, humans oftentimes start testing simple theories, or experience-based techniques that have been proven successful before (Todd & Gigerenzer, 2000). Experience-based techniques or so-called heuristics persist for as long as they yield expectations that come true. We argue that reasoning about mental states also comprises the use of heuristics, as cognitive resources are strained when applying ToM (e.g., Qureshi, Apperly, & Samson, 2010). In fact, Raijmakers et al.'s study about mental state reasoning implies that participants might have used heuristics: Children consistently used simple strategies that were incongruent with the logical structure of the games presented to them.

In this study, we present a computational cognitive model that shows how suboptimal reasoning about mental states is due to the use of heuristics. Here, we regard heuristics as simple strategies that prove themselves successful even though they do not take into account all task aspects. The model starts out using a simple heuristic, and gradually revises the heuristic when its decisions yield unexpected suboptimal outcomes. This revising can be considered a deliberate process, based on an interaction between factual knowledge and problem solving skills, similar to Van Rijn, Van Someren, and Van der Maas's (2003) model of children's developmental transitions on the balance scale task. The model can be generalized to other two-player games, because it reuses a small set of production rules when reasoning about increasingly more complex mental states. The idea of reusing a small set of production rules is inspired by Taatgen's *primitive elements theory*, which he presented in his paper on the nature and transfer of cognitive skills (Taatgen, 2013). Before we explain the model, we will first explain what task was used to measure reasoning about mental states.

Sequential games

The two-player sequential games in this study can be represented by the graph in Figure 1. 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. As a player's outcome depends on the other player's decision, both players have to reason about each other's mental states. Participants are always assigned to the role of Player 1, and decide at the first decision point whether to stop the game at A or to continue to the next decision point, which is Player 2's decision between his payoff in B and his payoff in either C or D, which in turn depends on Player 1's decision between Player 1's payoffs in C and D. Thus, before making a decision at the first decision point, participants have to reason about Player 2, who in turn has to reason about Player 1's subsequent decision. In other words, participants have to apply second-order ToM when making a decision.

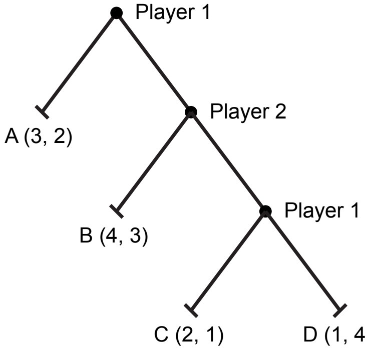

Figure 1: An extensive form representation of a two-player sequential game. Player 1 decides first, Player 2 second, and Player 1, again, third. Each end node has a pair of payoffs, of which the left-side is Player 1's payoff and the right-side Player 2's payoff. Each player's goal is to obtain their 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.

Empirical findings

In a previous study about second-order ToM reasoning in sequential games, participants' performance was significantly influenced by the type of training presented to them (Meijering, Van Rijn, Taatgen, & Verbrugge, 2011). In *stepwise training*, participants were familiarized with sequential games by successively presenting each additional decision point, and thus each ToM level, in subsequent blocks of games (Figure 2). This procedure facilitates embedding the application of second-order ToM in the decision making process. In *undifferentiated training*, in contrast, participants were immediately presented with games that had three decision points. However, these games could be considered 'easier' to play than the superficially similar but more difficult game in Figure 1. The 'easy' or so-called trivial games in *undifferentiated training* (see Figure 2, rightmost panel) required first-order ToM at most: As Player 2's payoff in B is either lower or higher than both his payoffs in C and D, Player 1 would only have to reason about Player 2 considering his own payoffs, irrespective of Player 1's decision at the third decision point. First-order ToM would suffice during *undifferentiated training* but not anymore during the experimental phase, which consisted of truly second-order games. Participants who had received *stepwise training* performed better during the experiment than participants who had received *undifferentiated training* (Figure 5).

Based on our cognitive model, we argue that participants in the *undifferentiated training* condition fell prey to using heuristics. As application of first-order ToM was sufficient during *undifferentiated training*, participants strengthened the corresponding "ToM1" strategy level, whereas higher strategy levels would have yielded correct decisions as well. However, from the start of the experimental phase, first-order ToM did not suffice anymore. Consequently, participants in the *undifferentiated training* condition performed worse in the experimental phase than participants in the *stepwise training* condition, whose reasoning had been scaffolded by subsequent blocks of increasingly higher-order ToM games during *stepwise training*.

Computational cognitive model

The model is implemented in the ACT-R cognitive architecture (Anderson, 2007), and it can be downloaded from http://www.ai.rug.nl/~meijering/iccm2013. Our model is based on an interaction between factual knowledge and problem solving skills. Arslan, Taatgen, and Verbrugge (2013) successfully modeled the development of second-order ToM using a similar approach. 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 plays the same payoff structures (i.e., items) as were presented to the participants. The 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's initial strategy, or heuristic, 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.

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 0 in the case of the heuristic 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 level 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 heuristic 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 future decisions.

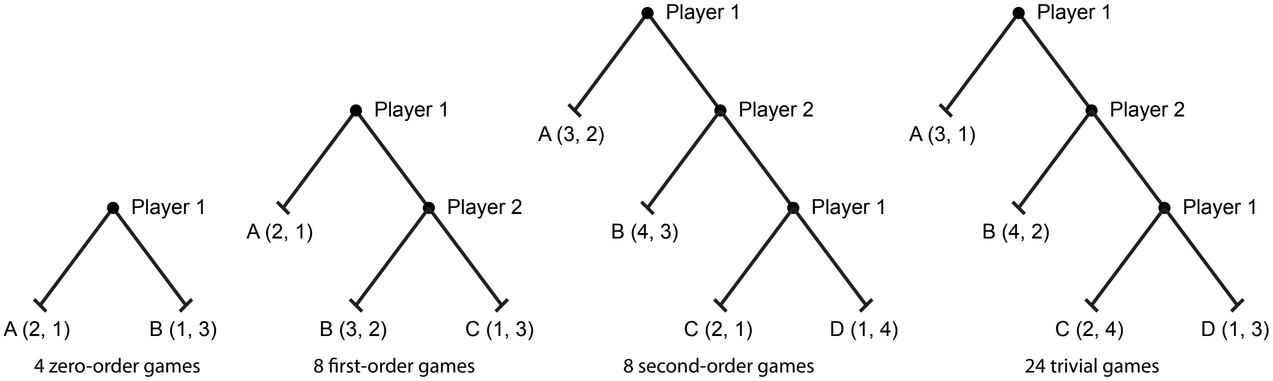

Figure 2: Extensive forms of example games. *Stepwise training* consisted of 4 zero-order, 8 first-order, and 8 second-order games. *Undifferentiated training* consisted of 24 trivial games.

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 (e.g., level 2). 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 two 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. This simple strategy can be considered a heuristic, as participants who are using it ignore the consequences of possible future decisions.

Our second assumption is that participants attribute their own strategy to the other player whenever the other player, at the second decision point, makes a decision that yields a payoff incongruent with the participant's expected outcome. If participants obtain expected outcomes, they do not have to revise their strategy. However, if participants obtain unexpected outcomes, they have to acknowledge that the unexpected turn of events was caused by the other player deciding at the next decision point. Reasoning about the other player, participants can only attribute a strategy they are familiar with. This idea is based on variable frame theory (Bacharach & Stahl, 2000). For example, if two persons have 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. The colorblind person can only predict or guess what object the other would select based on which shape is the least abundant. The same applies to reasoning about others: We can only attribute to others goals, intentions, beliefs, strategies and/or heuristics that we are familiar with ourselves.

Mechanisms

The simple strategy or heuristic 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 3). 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.

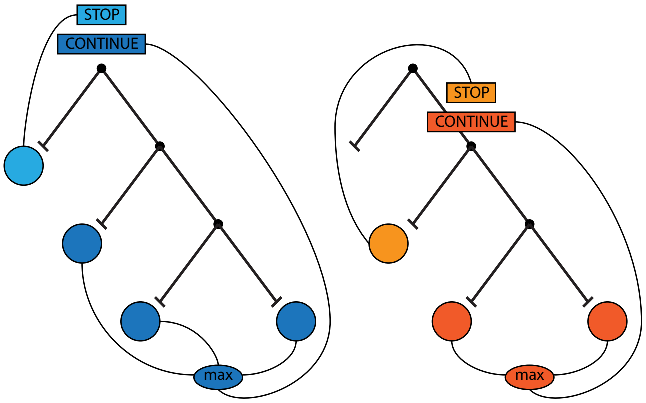

Figure 3: Simplified representation of heuristic. In the left panel, the model compares its payoff if it would stop (light blue) against its maximum possible payoff if it would continue (dark blue). In the right panel, the model compares Player 2's payoff if Player 2 would stop (light orange), against Player 2's maximum possible future payoff (dark orange). The left panel schematically represents the application of zero-order ToM, and the right panel the application of first-order ToM.

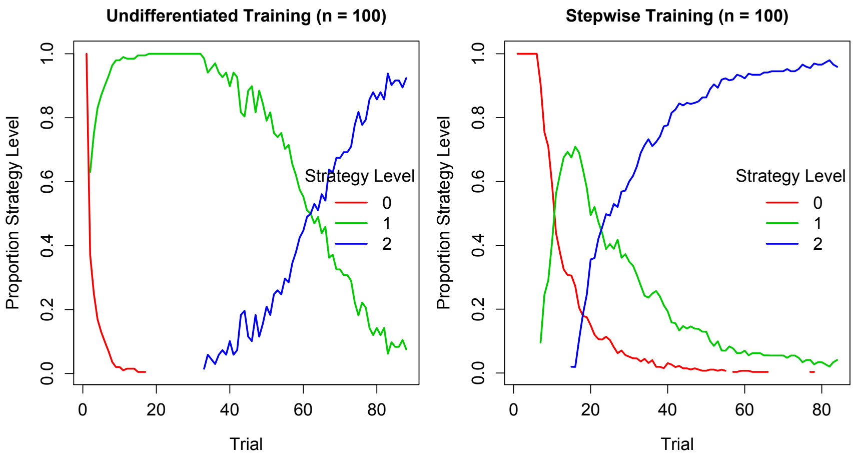

Figure 4: Proportion of models that apply strategy level 0, 1, and 2; plotted as a function of trial. The left panel depicts these proportions for the model that received undifferentiated training; the right panel depicts the proportions for the model that received stepwise training.

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 3, 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. However, before determining what payoff is associated with a *continue* decision, the model considers the next decision point (i.e., decision point 2) and attributes strategy level 0 to Player 2, who is deciding there.

The model will apply strategy level 0, but from the perspective of Player 2 (Figure 3, right panel). When reasoning about Player 2's decision, the model constructs a new game state, which references the previous one, to which it needs to jump back. 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 *stop* and *continue* decisions, but from the perspective of Player 2.

If the model would apply zero-order ToM from its own perspective, it would make a decision if it had determined the payoffs associated with *stop* and *continue* decisions. The model would make a decision because it would not have a previous game state to jump back to. However, the model's current game state representation references a previous one, and therefore the model will backtrack to that previous game state representation. Note that the previous game state did not have a payoff associated with a *continue* decision. However, the payoff associated with that *continue* decision 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 working memory, or the *problem state* or *imaginal* buffer (Anderson, 2007, Chapter 1), and the previous game state is stored in the *retrieval* buffer. 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 in working memory.

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. 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 when necessary. 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 4 shows the proportions of models that apply strategy level 0, 1, and 2, calculated per trial. The left panel of Figure 4 shows the output of 100 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 5. Therefore, models start applying strategy level 1, corresponding 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. Models do not start applying strategy level 2, because strategy level 1 yields correct decisions in each of the undifferentiated training games, which can be seen in Figure 5. However, in the experimental games, which are truly second-order games, strategy level 1 yields too many errors, and models start applying strategy level 2, which corresponds with second-order ToM. Initially, the accuracy drops, but it increases again as a greater proportion of models start applying second-order ToM, as can be seen in Figure 5.

The right panel of Figure 4 shows the output of 100 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 the correct answer in the first four games during stepwise training, because those are zero-order games. As can be seen in Figure 5 (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

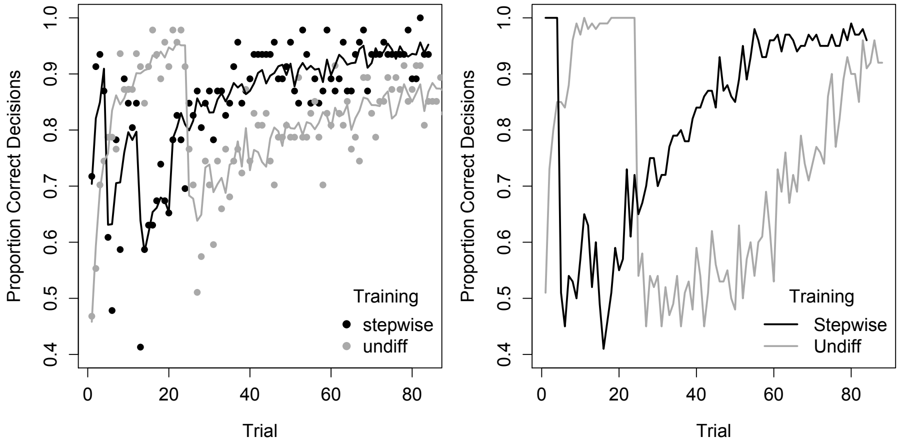

Figure 5: Proportion correct decisions, or accuracy, across subjects (left panel) and models (right panel). The solid lines in the left panel represent the fit of statistical model, which is added to visualize the proportion trends.

proportion of models applying strategy level 1 increases. In the next eight second-order training games (Trials 13 – 20), the proportions of models that apply strategy levels 0 and 1 decrease, as both strategy levels yield too many errors. Simultaneously, the proportion of models applying 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 5 (right panel).

The accuracy trends in the models' output qualitatively fit the accuracy of the participants in Meijering et al.'s study (Meijering et al., 2011). The quantitative differences are probably due to the fact that not all participants start out with the simple heuristic, whereas all models do. Some participants probably start with intermediate-level strategies, and due to large proportions of optimal outcomes, do not proceed to the highest level of reasoning. The model trends, changing as a function of type of game, correspond with our prediction that humans use heuristics, or simple strategies, for as long as these yield expected outcomes.

Conclusions

Based on previous empirical findings (Meijering et al., 2011) and our computational cognitive model, we argue that humans use heuristics when reasoning about others. We show that interplay between factual knowledge and problem solving skills, in contrast to a more implicit process of utility learning (Taatgen & Anderson, 2002; Van Rijn et al., 2003), allows the model to exploit the possibility of using simple strategies, not considering all task aspects. Although the update rule to assign a particular strategy to the other player might seem simplistic at first sight, the model does gradually master second-order ToM. As the model does not need to have task-specific productions rules, the model is flexible and can accommodate many two-player sequential games.

The methodological implication of this study is that experimenters should be careful in selecting 'practice' items, as participants exploit the possibility of using heuristics when possible. The theoretical implication is that participants 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. Over time, participants' reasoning becomes as simple as possible, as complex as necessary.

References

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

Arslan, B., Taatgen, N., Verbrugge, R. (2013). Modeling Developmental Transitions in Reasoning about False Beliefs of Others. *Proc. of the 12th International Conference on Cognitive Modelling (this volume)*.

Bacharach, M., & Stahl, D. O. (2000). Variable-frame level-n theory. *Games and Economic Behavior*, *32*(2), 220–246. doi:10.1006/game.2000.0796

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

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.

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: Drexel University.

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). doi:10.1371/journal.pone.0045961

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.

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*.

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "Broke?" A model of learning the past tense without feedback. *Cognition*, *82*(2), 123–155.

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.