A computational cognitive modeling approach to the development of second-order theory of mind
Arslan, Burcu

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

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

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Arslan, B. (2017). A computational cognitive modeling approach to the development of second-order theory of mind [Groningen]: University of Groningen

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

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

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 7: Putting the Pieces Together
Previous research on children's development of theory of mind has shown that children start to pass explicit first-order false belief tasks around the age of four\(^\text{26}\) (Wellman et al., 2001). However, it takes children a couple of years after that to pass second-order false belief tasks, in which children are expected to use theory of mind recursively (Perner & Wimmer, 1985; Sullivan et al., 1994). Because it has been shown that second-order false belief reasoning is important in many different aspects of human social cognition, it is essential to understand the underlying mechanisms of children's development as well as to find efficient ways to accelerate children's development. The main goal of this dissertation was to investigate the underlying mechanisms of this developmental lag between first-order and second-order false belief reasoning and to investigate the role of feedback in accelerating children's development of second-order false belief reasoning. These questions have not been studied extensively before. In addition to this goal, we also investigated how working memory and cognitive control contribute to children's development of first-order false belief reasoning. To achieve these goals, we combined a computational cognitive modeling approach with empirical research. In this last chapter, I summarize our main findings. Subsequently, I discuss what these findings mean in terms of the available theories, and I provide suggestions for future work.

### 7.1. Summary of the results

In Chapter 1, I introduced the methodology of this dissertation (see Figure 1.3) together with the research questions. Following that methodology, in Chapter 2, I first constructed two computational cognitive models in the light of the previous research and theories on children's development of second-order theory of mind. In addition to making precise predictions that can be tested empirically, the goal of the modeling approach was to provide a procedural explanation for the research questions, “Do 5-year-olds who fail in second-order false belief tasks predominantly use zero-order theory of mind reasoning or first-order theory of mind reasoning?” and “How do 5-year-olds revise their wrong theory of mind reasoning strategy to the correct second-order theory of mind reasoning strategy over time?” To this end, we constructed computational cognitive models by using ACT-R's two possible learning mechanisms in decision making, namely instance-based learning and reinforcement learning.

The main difference between these two learning mechanisms underlies where and how strategy selection occurs. In the instance-based learning model, the reasoning strategies (i.e., zero-order, first-order, second-order) are represented

\(^{26}\) But see Setoh, Scott, & Baillargeon (2016) for evidence that two-and-a-half-year-olds can pass explicit first-order false belief tasks when the processing demands are reduced.
as chunks in the declarative memory. Therefore, the strategy selection and revision are based on the activation of the strategy chunks, which are represented as declarative knowledge in the instance-based learning model. In contrast, in the reinforcement learning model, the strategy selection and revision are based on the utilities of the strategies, which are represented as procedural knowledge.

What does this difference between the two models mean? In the instance-based learning model, whenever a decision has to be made, the most active experience is retrieved from memory (i.e., the chunk with the highest activation) and used as the basis for the decision. The instance-based learning model revises its wrong reasoning strategy to a strategy\(^{27}\) one level higher when it gets the feedback “Wrong” and stabilize its current strategy when it gets the feedback “Correct”. Hence, the strategy selection and revision is explicit. On the other hand, in the reinforcement learning model, a reward/punishment is propagated back in time through the rules that have been used to make the decision, based on feedback. This reward/punishment mechanism updates the utility of those rules and finally, the model learns to apply a correct strategy. Therefore, the reinforcement learning model selects and revises its strategy implicitly.

The implicit vs. explicit strategy selection led to different predictions in the two models (Figure 2.2 and Figure 2.3). Unlike the reinforcement learning model, the instance-based learning model predicted that children who fail the second-order false belief tasks but have enough experience in first-order false belief reasoning would give answers to second-order false belief questions based on first-order reasoning as opposed to zero-order reasoning. This prediction was confirmed by an empirical study that we conducted with 72 five- to six-year-old children. The results showed that 17% of the answers were correct and 83% of the answers were wrong. In line with our prediction, 65% of the wrong answers were based on a first-order theory of mind strategy, while only 29% of them were based on a zero-order strategy, and the remaining 6% was “I don’t know”.

Both models predicted that it is possible to accelerate 5-year-olds development of second-order false belief reasoning with the feedback “Wrong” without any need to explain the reasons why children’s answers are wrong. This prediction is in contrast with the previous findings on children’s development of first-order theory of mind, which show that it is not possible to accelerate 4-year-olds’ development of first-order theory of mind by providing feedback without explanations when they are trained on first-order false belief tasks (Clements et al., 2000; Melot & Angeard, 2003).

In addition to this prediction, the instance-based learning model predicted that providing feedback with further explanations increases the odds of selecting the correct second-order reasoning strategy because the strategy revision is explicit. Therefore, if explanations are provided together with the feedback “Wrong”, children will have additional gain. On the other hand, the reinforcement learning model has nothing to say about further explanations because the strategy selection is implicit.

In order to test these predictions, in Chapter 3, we trained 106 5-year-old children with 12 different second-order false belief tasks in one of the following conditions: (i) Feedback with explanation; (ii) Feedback without explanations; (iii) No feedback (Figure 3.1). In the active control condition, children were trained with neutral stories that did not involve theory of mind reasoning. Confirming our instance-based learning and reinforcement learning models’ predictions, the results showed that there were significant improvements in children’s scores from pre-test to post-test in the ‘feedback without explanation’ condition (from 25% to 49%). Also, as the instance-based learning model predicted, children’s had additional gain when the feedback “Wrong” was provided with further explanations (from 31% to 68%). Moreover, surprisingly, children’s scores also improved in the ‘no feedback’ condition (from 33% to 55%). As expected, children did not show significant improvements in the ‘active control’ condition (from 29% to 35%). These improvements were not due to children’s age, verbal abilities and simple working memory span scores. Importantly, the children were able to generalize the training effect to another second-order false belief story type that they had not been trained on, and the training effect was stable at a follow-up session, which was 4 months after the pre-test.

In Chapter 2 and Chapter 3, we did not focus on the possible roles of executive functions and language in children’s development of second-order false belief reasoning. The assumption was that the children’s executive functioning and language abilities are sufficient to pass second-order false belief tasks. However, we know from previous studies that language and executive functions have an impact on children’s development of theory of mind (Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Carlson, Claxton, & Moses, 2014; Davis & Pratt; 1995, de Villiers & Pyers, 2002; de Villiers, 2005; de Villiers, 2007; Gordon & Olson, 1998; Keenan et al., 1998; Peterson & Siegel, 2000; Ruffman et al., 2002; Slade & Ruffman, 2005). Therefore, in Chapter 4, we conducted a cross-sectional study with 89 children in two age groups, one younger (4;6 – 6;5 years) and one older (6;7 – 8;10 years) in order to investigate the possible roles of syntactic recursion in the language domain and of working memory in the executive functions domain on children’s development of second-order theory of mind.

The reason why we focused on syntactic recursion was based on previous research showing that the syntactic component of language is related to children’s development of first-order theory of mind in terms of its hierarchical embedding.
structure (de Villiers, 2005; de Villiers, 2007; Hollebrandse & Roepen, 2014) and to children’s development of second-order theory of mind in terms of recursion (de Villiers et al., 2014; Hollebrandse et al., 2008). Different from those studies, for the first time in the literature, we used second-order relative clauses28 in order to investigate their relation with second-order false belief reasoning. Using second-order relative clauses instead of second-order complement clauses29 allowed us to specifically focus on the structural parallelism between second-order recursion in the language domain and in the thought domain by excluding the role of truth-value contrasts of complement clauses.

For the possible role of working memory in children’s development of second-order false belief reasoning, we invoked the serial processing bottleneck hypothesis (Verbrugge, 2009), which provides a procedural account for the role of complex working memory strategies in the development of second-order false belief reasoning. While simple working memory strategies only help people to build a representation of a list of information and to report it, complex working memory strategies allow people to process information in a more efficient way with the help of combined information processing steps to perform a task. The serial processing bottleneck hypothesis is based on the findings that working memory acts as a bottleneck, meaning that people can only hold one chunk of information in working memory at a time (Borst et al., 2010). This hypothesis assumes that children have a time threshold to give an answer in a given task and suggests that children need complex working memory strategies in order to process embedded30 beliefs in a way that chunks of information can pass through the working memory bottleneck within that time threshold.

In order to proceed in reasoning, due to the working memory bottleneck, at each step, the information in working memory needs to be sent to long-term memory to be retrieved later, if necessary. Retrieving information from long-term memory also takes time and increases the odds of forgetting and of retrieving wrong information (Anderson & Schooler, 2000). Therefore, having more inefficient rules instead of one efficient rule means that the process is more prone to errors and takes more time (Anderson et al., 2004; Taatgen & Anderson, 2002).

The evidence supporting the serial processing bottleneck hypothesis has been found in different cognitive domains, such as language and executive functions. Van Rij et al. (2010) focused on children’s poor performance on pronoun interpretation by constructing a computational model. Their model-based prediction was that once children have more time to interpret pronouns, their performance on pronoun comprehension should be increased. They validated this model-based prediction by presenting pronouns at a normal speech rate and at a slowed-down speech rate. Another empirical support to this view comes from Diamond et al.’s study (2002). They tested ninety-six 4-year-old children’s executive function abilities on the day-night Stroop-like task. In this task, children were supposed to say “day” when they see a picture of the moon, and supposed to say “night” when they see a picture of the sun. They inserted several seconds of delay between the stimulus and the response by introducing a little song saying “Think about the answer; don’t tell me”. This manipulation improved children’s performance on the task (from 56% correct to 86% correct). More recently, Ling et al. (2016) investigated whether this improvement was due to the delayed time that leads children to correct their mistakes or due to the task relevant information in the sentence “Think about the answer”. To investigate these competing hypotheses, Ling et al. tested seventy-two 4-year-old children in two different conditions. In one condition the little song was saying the task-relevant information “Think about the answer; don’t tell me”, and in the other condition the ditty was saying the task-irrelevant information “I hope you have a nice time; I like you”. The results showed that there was no difference between the task-relevant condition (83%) and the task-irrelevant condition (80%), and those conditions were significantly better than the standard version of the task (51%).

In order to test the predictions of the serial processing bottleneck hypothesis related to children’s development of second-order false belief reasoning, we tested 89 children in two age groups, one younger (4;6 – 6;5 years) and one older (6;7 – 8;10 years) with a simple working span and a complex working span task, in addition to second-order relative clauses. The analyses showed that although second-order syntactic recursion is significantly correlated with second-order false belief reasoning, the main predictor of children’s success in second-order false belief task is the complex working memory span task. Moreover, in line with the previous literature, both younger and older age groups’ justification scores (i.e., answers for the question “Why?” for their judgments) were far from perfect and lower than their judgment answers for the second-order false belief questions. Most of the correct justification answers involved implicit second-order answers (e.g., “Because she doesn’t know that Murat saw it”) for both age groups and none of the justification answers involved explicit second-order answers (e.g., “Because she believes that Murat doesn’t know that the chocolate is in the box”), except for one child in the older age group.

In addition to the highly significant effect of the complex working memory task, the simple working memory task explains significant variation in younger
children’s (4 – 6) justification answers. This significant correlation of the simple working memory task disappears for older children and only the complex working memory task is able explain the variance in children’s justification answers. Moreover, for the older age group, for the complex working memory task, we found that it only significantly predicts the second-order false belief justification score, not the second-order judgment score. The reason is that the judgment scores do not differ much among the older children, while their justifications still do. Thus, justifications seem to be a more sensitive variable for older children in the sense that they provide a finer distinction in their second-order reasoning abilities. While older children can give correct second-order false belief answers, their development still continues in terms of their justification abilities.

In Chapter 5 and Chapter 6, we stay with our interest in the role of executive functions but we shift our focus from children’s development of second-order false belief reasoning to children’s development of first-order false belief reasoning. In Chapter 5, the general research question that we wanted to investigate was “What kind of prior cognitive skills help children to pass explicit first-order false belief tasks?” More specifically, we investigated the role of simple and complex working memory strategies on children’s development of first-order false belief reasoning. By constructing computational cognitive models, we aimed to simulate how children’s prior skills in terms of simple and complex working memory strategies contribute to children’s success in first-order false belief tasks.

In order to simulate possible transfer of skills from working memory strategies to first-order false belief reasoning, we constructed three computational cognitive models by using the cognitive architecture PRIMs, which has been built specifically to explain transfer of skills (Taatgen, 2013). Instead of choosing working memory tasks that are constructed to test children in an experimental setting, we chose two real-life tasks which children might encounter before the age of four. We named the first task the ‘pencil task’, in which the goal is to count the total number of yellow and green pencils in a group of blue, red, yellow and green pencils (Figure 5.2). The second task was named the ‘marble task’, in which the goal is to find the two bags that contain the same number of marbles of the same color (Figure 5.3). While the pencil task calls for simple working memory strategies, the marble task calls for complex working memory strategies. Finally, as a third task, we modeled a first-order false belief task. We hypothesized that performing a number of versions of the marble task would help more than the pencil task in children’s transitions from failure to success in the first-order false belief task.

The simulation results showed that the first-order false belief task model learns to pass the task faster when it has prior experience of a task that needs simple or complex working memory strategies. Moreover, confirming our hypothesis, the first-order false belief task model mastered the first-order false-belief task much faster when it was first trained in the marble task, which required complex working memory strategies, than when it was first trained in the pencil task.

In Chapter 6, we investigated another important component of executive functions, which is cognitive control in relation to children’s development of first-order false belief reasoning. Previous research has shown that cognitive control and first-order false belief reasoning develop around the same age (Perner & Lang, 1999; Müller et al., 2005; Henning et al., 2011). In order to investigate whether children first develop cognitive control and then develop first-order false belief reasoning or vice versa, Kloo and Perner (2003) conducted a training study, in which 3-year-olds were trained on the Dimensional Change Card Sorting task of cognitive control in one experimental condition and in the first-order false belief tasks on the other experimental condition, both with explicit feedback with further explanations. Kloo and Perner found that there is a mutual transfer between the Dimensional Change Card Sorting task and the first-order false belief task, meaning that training children with one of the two tasks by providing feedback with further explanations significantly improved children’s performance on the other task. Based on these findings, Kloo and Perner proposed that the problem that 3-year-olds encounter is related to failure in redefining an object or situation and that training children with explicit feedback helps them to understand that an object or situation can be described differently from different perspectives. However, Kloo and Perner stated that the exact nature of the transfer effects remained to be determined.

In Chapter 6, our goal was to provide an explanation for Kloo and Perner’s results by constructing computational cognitive models. To achieve this goal, we constructed our models by using the PRIMs cognitive architecture. One of the two salient explanations of transfer of skills in PRIMs theory suggests that transfer of skills between different tasks occurs when performing one of the tasks trains a particular strategy (e.g., proactive strategy), which is already present in declarative memory, and the other tasks also require to use that particular strategy or parts of that strategy in order to succeed at the task instead of using a simpler strategy (e.g., reactive strategy). Our modeling approach showed that both the Dimensional Change Card Sorting task and the first-order false belief task have a similar structure in terms of having two competing strategies, only one of which led the model to give a correct answer (Figure 5.2). Our models’ results had a quite good fit to Kloo and Perner’s data (Figure 5.4). Based on our models, unlike Kloo and Perner’s theory, we proposed that the common element in the two tasks is two competing strategies, only one of which leads to a correct answer, namely the strategy of control. Providing children with explicit feedback trained them to use a strategy of control instead of using a simpler reactive strategy.
7.2. Discussion

In the previous section, I summarized the findings of this dissertation. In this section, I put the pieces together in the light of theories about children’s development of first-order and second-order theory of mind that I mentioned in Chapter 1, namely conceptual change and complexity. I also point out future research that will shed light on children’s development of false belief reasoning.

7.2.1. Conceptual change

The conceptual change explanation for children’s second-order false belief reasoning proposes that children need to realize that mental states such as beliefs can have other beliefs and not just events in the world as their content and can be used recursively (Miller 2009; 2012).

As we mentioned in Chapter 3, our findings that 5-year-old children do not need further explanations or any feedback in order to pass the second-order false belief tasks contrasts with the findings in the literature that 3-year-old children’s development of first-order false belief reasoning cannot be accelerated without providing both feedback and further explanations when they trained on false belief tasks. Moreover, in the no feedback condition of our training study, children’s second-order false belief scores also improved. This finding was surprising in terms of our models’ predictions. We surmise that exposing children to second-order false belief stories and asking them second-order false belief questions, together with the justification questions “Why?” helps children to reflect about their own judgments. Thus, asking justification questions helps children to revise their wrong first-order reasoning strategy to a correct second-order reasoning strategy.

Our findings related to children’s improvements in the ‘feedback without explanation’ and ‘no feedback’ conditions suggest that the conceptual change explanation alone cannot be the whole story because we did not use any second-order mental state embedding, neither in our stories nor at our training sessions. Of course, we cannot rule out the possibility that children go through a conceptual change by realizing that mental states can be used recursively between the ages three and five (see also Mahy, Moses, & Pfeifer, 2014 for a review stating that it is hard to rule out the possibility of a conceptual change at both the behavioral and neural levels).

Based on our computational modeling approach that we presented in Chapter 2, we propose that even if children go through another conceptual change after they pass the first-order false belief task, they still need experience in second-order false belief reasoning in order to revise their wrong first-order theory of mind strategy to a correct second-order strategy. Moreover, the confirmed two predictions of our instance-based learning model indicate that children select and revise the reasoning strategies in false belief tasks explicitly.

Where do those reasoning strategies come from? A possible answer is related to learning common sense knowledge for reasoning about false beliefs. Heyes and Frith (2014) propose that explicit theory of mind is culturally inherited, and that parental stories and “causal-explanatory” statements might be some of the possible sources of this common-sense knowledge.

In the following subsection, I turn to the complexity explanation in terms of language and executive functions.

7.2.2. Complexity

The complexity explanation suggests that the higher complexity of the second-order false belief tasks adds further demands on executive functions and involves more complex language in comparison to first-order false belief tasks (Miller, 2009; 2012).

We find that understanding that mental state words can be used recursively, in spite of what the conceptual change explanation suggests, is not enough to overcome the additional demands on executive functions and to parse or use embedded clauses, such as “She believes that he thinks that the key is in the car”. In order to do so, children need to process the embedded beliefs serially, which requires complex working memory strategies to pass the serial processing bottleneck (Verbrugge, 2009), as we argued in Chapter 4.

What might be the role of language in children’s development of second-order false belief reasoning? We argue that syntactic recursion helps chunking the hierarchical embedded beliefs by linearizing them to an efficient reasoning rule that passes through the serial processing bottleneck of working memory (see also Hollebrandse & Roeper, 2014 for a similar argument). Our finding showing that there is a high correlation between second-order relative clauses and the complex working memory task supports this view (Chapter 4).

The instance-based and reinforcement learning models that we presented in Chapter 2 involved complex and specialized reasoning rules in the form of IF-THEN rules. However, it is unlikely that children have these complex and specialized rules in their minds to give a specific answer to false belief questions. Our computational cognitive models that we introduced in Chapter 5 showed that complex working memory strategies that involve an element of cognitive control can contribute to children’s transitions from failure to success in first-order false belief tasks. Based on these findings, we propose that one of the important sources of combining those complex and specialized production rules might be children’s experience in working memory strategies that they apply in their daily...
lives. Future research is needed to test this explanation and the serial processing bottleneck hypothesis, possibly with a training study in which the children are trained with a complex working memory span task in one condition and with a simple working memory task in another condition and their performance in second-order false belief reasoning from pre-test to post-test is assessed.

In Chapter 6, our modeling approach showed us that giving explicit feedback in the Dimensional Change Card Sorting task and first-order false belief tasks train children to be more flexible in their behavior in terms of the current goals of the tasks. Because around the age four most children pass first-order false belief tasks, we can surmise that children start to have cognitive control, meaning that they learn to be flexible in their behavior depending on the current goal. Future research is needed to find out whether there is far transfer from complex working memory tasks to second-order false belief tasks. Based on our results in Chapter 3 and Chapter 4, we predict that training children with complex working memory tasks that involve not only remembering a list but also require cognitive control helps children to pass second-order false belief tasks.

However, as we have shown in Chapter 6, working memory by itself is not sufficient. Children also need to use cognitive control in order to give a correct answer. While for first-order false belief reasoning it is necessary for children to get feedback with explanations to understand that cognitive control is needed, for second-order false belief reasoning, children need to encounter many examples in which they can update their wrong strategies to the correct second-order false belief strategy.

### 7.3. Conclusions

In this dissertation, we combined computational cognitive modeling and empirical studies in order to investigate children’s development of second-order false belief reasoning. After reviewing the previous theories, we constructed computational cognitive models based on the available theories. Constructing models gave us the opportunity to implement cognitive processes and cognitive concepts with precision instead of using the concepts without teasing apart their components. The simulation results of the models revealed novel predictions that we tested empirically. Moreover, we used computational modeling approach in order to understand the underlying mechanisms of the available data in the literature. Based on our models and empirical findings together with the previous literature, I would like to propose a timeline for the contributing factors in children’s development of second-order theory of mind, which is depicted in Figure 7.1.

Development of first-order theory of mind is necessary for the development of second-order theory of mind. It seems more likely that a conceptual change is