The Role of Simple and Complex Working Memory Strategies in the Development of First-order False Belief Reasoning: A Computational Model of Transfer of Skills

Burcu Arslan (barslan.cogs@gmail.com), **Stefan Wierda** (wierda.stefan@gmail.com),
Niels Taatgen (N.A.Taatgen@rug.nl), **Rineke Verbrugge** (L.C.Verbrugge@rug.nl)

Institute of Artificial Intelligence, University of Groningen, P.O. Box 407,
9700 AK Groningen, The Netherlands

Abstract

In their fourth year, most children start to understand that someone else might have a false belief, which is different from the reality that the children know. The most studied experimental task to test this development is called the first-order false belief task. What kind of prior cognitive skills help children to pass the false belief task? There are hundreds of correlational studies that have shown that language and executive functions (such as inhibition and working memory) play a role. Moreover, several training studies have shown the importance of language and inhibition in the development of false belief reasoning. However, to the best of our knowledge there has been no training study (with normally developing children) to investigate the role of working memory strategies in the development of false belief reasoning.

We present here a computational cognitive model to investigate transfer from working memory strategies to false belief reasoning. For this reason, in addition to the false belief task, we constructed two tasks that children encounter in their daily life: a pencil task (simple working memory) and a marble task (complex working memory). Our simulation results confirm our hypothesis that there is more transfer from the marble task to the first-order false belief task than from the pencil task to the first-order false belief task, because of the more complex working memory strategies that appear to be necessary in the false belief task. The results of our simulations suggest conceptual predictions to be tested experimentally.

Keywords: Theory of Mind; False Belief Reasoning; Working Memory; Transfer; Cognitive Modeling; Actransfer.

Introduction

Children's development of reasoning about other people's representational mental states such as beliefs, desires and knowledge has been one of the most studied areas in developmental psychology. In order to conclude that an agent has such a theory of mind (ToM, Premack & Woodruff, 1978), Dennett (1978) argued that it is necessary to test whether the agent can correctly attribute a false belief to another agent. Since then, the explicit false belief task (Wimmer & Perner, 1983) has become one of the most commonly used tasks that verbally tests children's ToM. In the explicit first-order false belief task, children are required to make and report a decision about another person's mental state while they know the real situation, which happens to be different from the other person's false belief. Various studies have shown that children cannot pass the explicit first-order false belief tasks until the age of four (Wimmer & Perner, 1983; Wellman, Cross & Watson, 2001).

One of the most commonly studied explicit first-order false belief tasks is called the unexpected location change task. In this task the story goes more or less as follows: 'Sally and Anne are in the room. Sally puts her chocolate into the basket. After that, she leaves the room. Anne takes the chocolate from the basket and puts it into the box and she also leaves the room. Later, Sally comes back to the room.' The first-order false belief question is "Where will Sally look for the chocolate?" If a child correctly reasons about Sally's mental state, s/he reasons that because Sally did not see Anne taking the chocolate from the basket and putting into the box, Sally will look for the chocolate in the place where she last saw it—thus, the child would answer that Sally will look in the basket.

Interestingly, until the age of 4, children make systematic errors by reporting the real location of the chocolate, which is the box in the above story. This phenomenon is called 'reality bias' (Mitchell et al., 1996). Previous studies of the explicit false belief task showed that 3-year-old children's accuracy is around 30%, 4-year-olds' accuracy is around 50%, 6-year-olds' accuracy is around 80%, and finally around the age of 8, children's performance is at ceiling, similar to adults' performance (Wellman, Cross & Watson, 2001). According to the 'reality bias' view, in order to give correct answers, children should inhibit their own response and take into account others' perspectives.

What kind of cognitive skills are required for children to overcome their 'reality bias' and pass the explicit first-order false belief task? It is a matter of debate whether the development of first-order ToM is purely a matter of conceptual change. In fact, it has been shown that other cognitive factors contribute to the development of first-order false belief reasoning. Several studies have examined the so-called 'far transfer' of skills by training children with different cognitive tasks and investigating whether children's performance on the first-order false belief task has improved or not after the training. Those studies revealed that there is indeed a far transfer of skills from language (Hale & Tager-Flusberg, 2003) and inhibition (Kloo & Perner, 2003) to first-order false belief reasoning. We believe that the working memory strategies that children use also contribute to the development of false belief reasoning. The important role of working memory for first-order false belief reasoning has already been shown by correlational studies (Gordon & Olson, 1998; Hughes, 1998; Keenan, Olson, & Marini, 1998). Moreover, we have evidence for a significant effect of the complex working memory task but not the simple working memory task in second-order false belief reasoning (Arslan, Hohenberger, &Verbrugge, *submitted*). However, there has so far been no experimental training study focused on the role of working memory strategies in the development of first-order false belief reasoning.

Training studies need more time and effort than correlational studies. For this reason, constructing computational cognitive models to predict what kind of skills might be transferred to another domain (far transfer) is an effective way of designing an appropriate training study. There have been a few computational models of the development of explicit false belief reasoning (Wahl & Spada, 2000; Triona, Masnick & Morris, 2002; Bello & Cassimatis, 2006; Hiatt & Trafton, 2010; Arslan, Taatgen & Verbrugge, 2013). However, none of those models are aimed to predict and explain far transfer from daily life tasks to explicit false belief reasoning.

In the current study, we aim to investigate the possible transfer of cognitive skills from working memory strategies that children use in their daily-life tasks to first-order false belief reasoning by constructing a computational cognitive model that helps us to make more precise predictions. To investigate the role of working memory strategies, we modeled one simple working memory task (the pencil task) and one complex working memory task (the marble task) together with the first-order false belief task. The pencil and marble tasks were inspired by Brain Quest game cards for children of ages 5 to 6 (http://www.brainquest.com/) and they differ from each other in terms of the complexity of the working memory strategies required to solve them (see the sections "A cognitive model of the pencil task" and "A cognitive model of the marble task" for details). We hypothesized that there would be more transfer from the marble task to the first-order false belief task than from the pencil task, because of the more complex working memory strategies required by the marble task, which are also necessary in the false belief task.

In order to model transfer from the pencil and the marble tasks to first-order false belief reasoning, we modeled the tasks using Actransfer (Taatgen, 2013). Actransfer implements the primitive elements theory (Taatgen, 2013) of the nature and transfer of cognitive skills. Actransfer builds on the symbolic computational cognitive architecture Adaptive Control of Thought–Rational (ACT-R; Anderson, 2007). The Actransfer architecture uses ACT-R modules, buffers and mechanisms such as production compilation (Taatgen, 2002).

The primitive elements theory (Taatgen, 2013) breaks down the complex production rules typically used in ACT-R models into the smallest possible elements (PRIMs) that move, compare or copy information between modules. There is a fixed number of PRIMs in the Actransfer architecture. When PRIMs are used often over time, production compilation combines them to form more complex production rules. While those PRIMs may have some task-specific elements, PRIMs also have task-general elements that can be used by other tasks. Transfer occurs if two tasks have common task-general elements: One task can benefit from another trained task because of the already compiled production rules that are learned through production compilation. Taatgen (2013) showed the predictive power of Actransfer by modeling a variety of transfer experiments such as text editing (Singley & Anderson, 1985), arithmetic (Elio, 1986), and cognitive control (Chein and Morrison, 2010).

In the following sections, we will explain our Actransfer models in detail, present the results of the simulations and discuss our findings.

A Cognitive Model of the First-order False Belief Task

Our Actransfer model for the first-order false belief task was inspired by Arslan, Taatgen and Verbrugge's (2013) ACT-R model and Wierda and Arslan's (2014) Actransfer model of first- and second-order false belief reasoning. A simulated storyteller presents the first-order false belief story to our model. The way we implemented this is by updating the perceptual buffer every 4 seconds with new story facts. For each picture in the story, the storyteller tells what happens in that particular picture. The model "listens" to the story and stores what happened in each picture in its declarative memory. The pictures that have actions related to changing the location of the object of interest are chained together in chronological order. Adding a pointer that refers to the previous picture fact realizes the chaining of the picture facts. Also, all related action facts are linked in a similar manner with the corresponding picture fact.

At the end of the story, the storyteller presents the model with a first-order false belief question ('Where will Sally look for the chocolate?'). First, the model creates a first-order chunk in declarative memory that represents the first-order false belief question ("Where will Sally look for the chocolate?"). Next, the model creates a zero-order chunk that represents the corresponding zero-order question ("Where is the chocolate?") by breaking up the first-order false belief question. The model keeps a reference to the zero-order chunk in working memory, which in turn has a pointer towards the first-order chunk. After the question is presented, the model uses two strategies to reason about the question. The first strategy is a memory strategy in which the model always tries to retrieve a picture fact that has an action related to the object's location change. It then looks at that picture when remembering facts about it, such as "Anne put the chocolate into the box". The second strategy is a perception strategy, which is used whenever the model has forgotten the story facts. The model looks at each picture in detail and extracts the story facts from the picture. Below, we present the details of these two strategies (memory and perception) in detail.

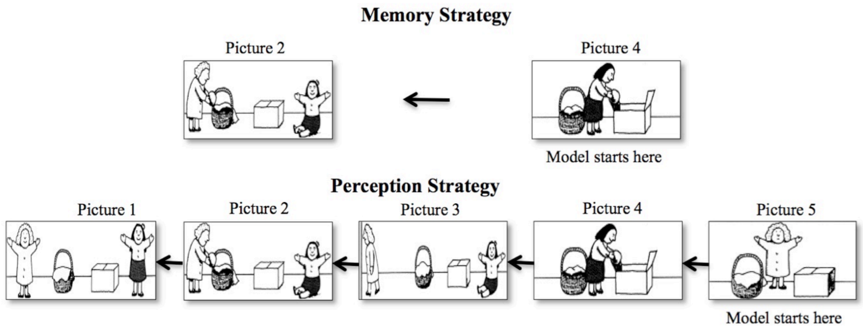

Figure 1. The order of the pictures that the false belief model attends in memory and perception strategies

The memory strategy

The *memory strategy* is the first strategy that the model uses. The model tries to retrieve what was the last picture in which an action happened that was related to the location of the object. If it retrieves that picture fact, it then tries to remember what exactly happened in that picture (for example, a location change of the chocolate). If the model successfully remembers that Anne put the chocolate into the box, it puts the location of the chocolate ("the box") in its working memory and then tries to recall the question. First, the zero-order question is retrieved by the reference that is kept in working memory. If the zero-order chunk does not point to a first-order chunk, the model gives an answer by reporting the location from its working memory ("the box"). However, in this particular task, the actual question put to the model is the first-order false belief question.

Thus, the model then tries to recall the first-order question ("Where will Sally look for the chocolate?"). If it retrieves the first-order question, it checks whether the person in the question performed the action in that picture. Because it was not Sally but Anne who put the chocolate into the box and Sally is absent in the picture, the model tries to retrieve another picture fact at which another action towards the object happened and again it tries to recall what exactly happened in that picture (Figure 1). This process continues until the person who moved[1] the chocolate is the same person who is mentioned in the question.

If the model's run-time passes a preset threshold, the model stops reasoning and answers whatever it currently has in working memory. In this way, we simulate that the model gives up for whatever reason (for example, it takes too long or it gets distracted). As a result, the model will at first give either no answer at all or a zero-order answer. Note that this is because the model first stores the most recent location of the chocolate in its working memory, which corresponds to the zero-order answer ("the box"). When the model reaches the part of the story where the first-order answer ("the basket") can be found, this location will be stored in working memory and the model starts giving the correct first-order answer.

[1] We used the action of moving the chocolate, but the model could also easily be adapted for seeing.

The perception strategy

In our behavioral study (Arslan, Verbrugge, Taatgen, & Hollebrandse, 2014), we have successfully trained 5-6 year old children to pass the second-order false belief tasks. We experienced that on most occasions, children look back in the pictures. Similarly, our model uses the *perception strategy* by looking at the pictures in more detail if it fails to apply the memory strategy because it has forgotten some of the facts of the story as told by the storyteller. In the perception strategy, the model first focuses its attention at the most recently seen picture and inspects whether there is an action related to the salient object in the picture. If there is a person present in the picture, it checks whether this person performed an action or not. Subsequently, it creates a new action fact about the picture in memory and starts to reason with those newly created chunks in the same way as in the memory strategy.

Note that both the perception strategy and the memory strategy use almost the same mechanism for reasoning about the question. The difference is that the irrelevant pictures for finding the answer are skipped in the memory strategy, whereas every picture has to be inspected in the perception strategy (see Figure 1). This is because the memory strategy broke down, and the model cannot immediately recall Picture 4 and subsequently Picture 2 of the false belief story (see Figure 1) at which there are actions related to the location of the object.

A Cognitive Model of the Pencil Task

As we mentioned above, we modeled one simple working memory task, the pencil task, and one complex working memory task, the marble task. In the former task, the goal is to count the total number of yellow and green pencils in a group of blue, red, yellow and green pencils (Figure 2). We modeled this task as follows. The model first looks at a pencil that is in its perceptual buffer. If the color of the pencil is blue or red, it focuses its attention to another pencil. This procedure is repeated until the model finds a yellow or green pencil. It then initializes counting by retrieving a counting fact from its declarative memory and copying the retrieved number to the working memory. It keeps on searching pencils until it finds another yellow or green pencil. When it finds one of those, the counter in working memory is updated by retrieving the next counting fact. After attending all pencils, the model reports the total number of yellow and green pencils. As becomes clear from this explanation, this task does not need any complex working memory strategies. It simply uses one slot in the working memory buffer and it updates that slot whenever it is necessary.

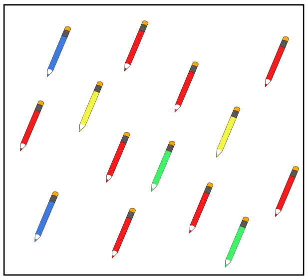

Figure 2. The pencil task (simple working memory)

A Cognitive Model of the Marble Task

The goal of this task is to find, out of a small number of bags of marbles, the two bags that contain the same number of marbles of the same color (Figure 3). Our model uses a strategy that focuses on one color in a bag and counts that color of marbles in each bag until finding a bag that shares the same number of that color. We assume that this is one of the strategies[2] that children use in general. Because we are interested in comparing a complex working memory strategy with a simple one, the strategy that we used for modeling will suffice for our purposes.

The model starts by looking at the first bag and retrieving a color fact from its declarative memory to count the marbles of that color. For example, if the model retrieves the color red, it copies red to one of four working memory slots. At the same time it copies the identity of the bag (Bag-1) in another working-memory slot to report it back when necessary. Then, it looks at the first marble that is in its perceptual buffer, which is blue in our example. Since blue is not the same as the color that is in working memory (red), the model focuses its attention to another marble and repeats that procedure until it finds a marble that matches the color in working memory. After it finds a red marble, it initializes counting by requesting the retrieval of a counting fact from its declarative memory and copying the retrieved number to a third working memory slot. The model then updates that counting slot if it attends another red marble.

Once all marbles of the current color in the current bag (Bag-1) are counted, the model tries to remember if it has already seen another bag that has the same number of marbles of the same color. In the example, because it is the first bag, the model cannot remember a bag that has the same number of red marbles and focuses its attention on another bag to continue to count the red marbles. It carries out the same procedures for the second and the third bags.

After counting all the red marbles in all bags and not remembering any bags that have the same number of red marbles, the model creates a new working memory chunk by emptying all its slots except the slot that has the current color (red). This process also consolidates all information present in working memory and thus creates a new chunk in declarative memory that can be retrieved later on—effectively it remembers which bags it has seen with how many marbles of a given specific color.

Later, it repeats the procedures above by retrieving another color from its declarative memory. Let's say the color blue is retrieved this time. The model counts the blue marbles in the first and second bags, and checks if they have the same number of blue marbles. Because this is not the case, it moves its attention to the third bag and counts the blue marbles. At this point the model can successfully retrieve the first bag with the same number of blue marbles, which is 1, from its declarative memory. Finally, it gives an answer by reporting the first and third bag.

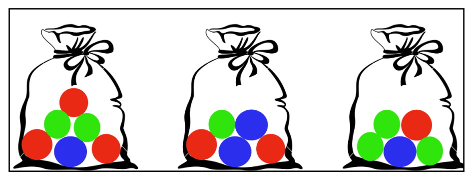

Figure 3. The marble task (complex working memory)

Results

In order to investigate transfer from the simple working memory task (pencil) and the complex working memory task (marble) to the first-order false belief task, we ran simulations in three conditions. In the first condition (FB-only), we ran 100 simulations of a child doing the first-order false belief task 100 times (thus, a total of 100 × 100 = 10,000 trials were simulated).

In the second condition (Marble-FB), we first ran the marble task for 10,080 minutes (24 hours in 7 days) in ACT-R's time. The model would perform as many trials as it could possibly do within that time. Subsequently, the model performed 100 trials of the first-order false belief task. This condition was also simulated 100 times, simulating 100 children.

In the third condition (Pencil-FB), we followed the same protocol as in the second condition but first we ran the pencil task instead of the marble task. Table 1 shows the mean and the standard deviations of the number of simulations for each task. As can be understood from Table 1, the model could squeeze more trials of the pencil task than trials of the marble tasks into the 10,080 minutes. After all, each trial of the marble task, in which several numbers of objects need to be compared, takes much more time than the pencil task, which just involves counting an easily recognizable subset of objects. Therefore, the model has much more previous experience as expressed in number of

[2] Another possible strategy would be focusing on a bag and counting the number of marbles of all its colors, and repeating this procedure until another bag has the same number of a color with the bag in focus. Because both strategies use similarly complex working memory strategies, this would probably not change the simulation results of transfer.

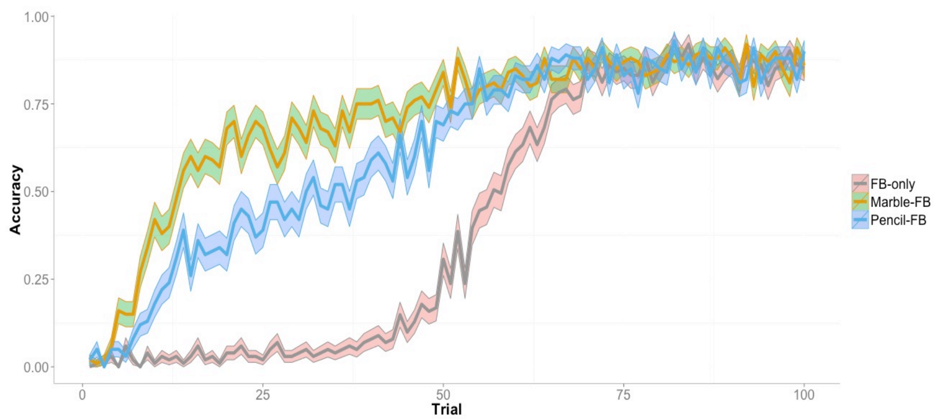

Figure 4. The results of the simulations of 100 trials averaged over 100 runs representing 100 children. The FB-only condition represents 100 trials of false belief task simulation only. The Marble-FB condition represents 100 trials of false belief task simulation after 10,080 minutes (ACT-R time) of training on the marble task. The Pencil-FB condition represents 100 trials of false belief task simulation after 10,080 minutes (ACT-R time) of training on the pencil task.

trials in the pencil task before we run the false belief task model compared to if it is first trained with the marble task. However, as mentioned above, the amount of exposure as expressed in seconds is equal for both tasks.

Table 1. The mean and the standard deviations of the number of simulations for each task

Task	Mean	Standard Deviation
FB-Only	100	0
Marble-FB	217	8.6
Pencil-FB	850	21.0

Figure 4 shows the results of the simulations. In the FB-only condition, in which the model starts without any prior knowledge other than the PRIMs as described in the Introduction, the first-order false belief task model gives the zero-order answer ("reality bias") by reporting the real location of the chocolate (i.e., "the box") until around the 60th trial. After that, it gives the correct answer for the first-order false belief question (i.e., "the basket").

In the Marble-FB condition, in which the first-order false belief task model experienced the prior practice of the marble task, the model starts to give the correct answer much earlier, around the 15th trial. Finally, in the Pencil-FB condition, the model starts to give the correct answer for the first-order false belief question around the 35th trial, which is earlier than in the FB-only condition, but later than in the Marble-FB condition.

Discussion

Our goal was to investigate the role of working memory (WM) strategies in the development of first-order false belief reasoning. In order to achieve this goal, we modeled two real life examples, the pencil task and the marble task, corresponding to a simple and a complex working memory strategy, respectively, by using the cognitive architecture Actransfer.

In agreement with the previous behavioral studies that have shown the correlation between working memory and the development of first-order false belief reasoning (Gordon & Olson, 1998; Hughes, 1998; Keenan, Olson, & Marini, 1998; see Arslan, Hohenberger & Verbrugge, forthcoming for second-order false belief reasoning), our results show that having an experience with tasks that need working memory strategies contribute to this development. Because more complex working memory strategies are needed in our first-order false belief task model than a simple strategy that needs to just update the WM, we predicted that there would be more transfer from the marble task (complex working memory) to the first-order false belief task than from the pencil task (simple WM) to the first-order false belief task. The results confirm our hypothesis.

The first-order false belief task model learns to pass the task faster when it has a prior experience of a task that needs simple or complex WM strategies. This result is straightforward, as we compare the simulations with prior knowledge to a model that has no prior experience at all. More interestingly, the model that was first trained in the marble task, which required complex working memory strategies, mastered the first-order false-belief task much faster—even though the model was able to do fewer trials of the marble task in a given time period ($M_{\text{no of simulations}}$ =217, SD= 8.6) than the model that was first trained in the pencil task, which required simple WM strategies ($M_{\text{no of simulations}}$ =850, SD= 21.0). Note that the amount of exposure to both models was similar in terms of time, as stated above. Together with the experimental training studies that we mentioned in the Introduction, our work implies that passing false belief tasks is not a skill acquired through maturation, but by experience.

Future directions

Although the amount of exposure-time in the Marble-FB and the Pencil-FB conditions was the same, one could argue that it is the general complexity of the marble task (complex working memory), which causes the transfer to the false belief task. In addition to comparing the marble task to the pencil task (simple working memory), including a third task that has the same complexity as the marble task but that does not require complex working memory strategies might be a better control condition. Also, finding a task to model that has the same complexity as the first-order false-belief task but without the need of working memory might be worthwhile.

The results of our simulations suggest conceptual predictions that should be tested experimentally in experiments with 3-4 year old children.

Acknowledgments: We are grateful to the Netherlands Organization for Scientific Research for Vici grant NWO-277-80-01, awarded to Rineke Verbrugge.

References

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.

Arslan, B., Verbrugge, R., Taatgen, N. & Hollebrandse, B. (2014). Teaching children to attribute second-order false belief: A training study. In Szymanik, J. & Verbrugge, R. (eds.), *Proceedings of the Second Workshop Reasoning About Other Minds: Logical and Cognitive Perspectives,* CEUR Workshops Proceedings, 1208, 1-5.

Arslan, B., Taatgen, N., Verbrugge, R. (2013). Modeling developmental transitions in reasoning about false beliefs of others. In R. West & T. Stewart (eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*, Ottawa: Carleton University.

Arslan, B., Hohenberger, A. & Verbrugge, R. (submitted). The role of language and memory in the development of second-order theory of mind.

Bello, P. & Cassimatis, N. (2006). Developmental accounts of theory-of-mind acquisition: Achieving clarity via computational cognitive modeling. *Proceeding of the 28th Annual Meeting of the Cognitive Science Society*, 1014-1019. Vancouver: Cognitive Science Society.

Chein, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, 17, 193–199.

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1, 568-570.

Elio, R. (1986). Representation of similar well-learned cognitive procedures. *Cognitive Science*, 10, 41–73.

Gordon, A. C. L., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology*, 68, 70–83.

Hale, C., M., & Tager-Flusberg, H. (2003). The influence of language in theory of mind: A training study. *Developmental Science*, 6 (3), 346-359.

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

Hughes C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, 16, 233–253.

Keenan, T., Olson, D. R., & Marini, Z. (1998). Working memory and children's developing understanding of the mind. *Australian Journal of Psychology,* 50, 76-82.

Kloo, D. & Perner, J. (2003). Training transfer between card sorting and false belief understanding: Helping children apply conflicting descriptions. *Child Development,* 74, 1823-1839.

Mitchell, P., Robinson, E. J., Isaac, J. E. & Nye, R. M. (1996). Contamination in reasoning about false belief: An instance of realist bias in adults but not children. *Cognition*, 59, 1-21.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development,* 61, 973-983.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515-526.

Singley, M. K., & Anderson, J. R. (1985). The transfer of text-editing skill. *International Journal of Man-Machine Studies*, 22, 403–423.

Tager-Flusberg, H., & Sullivan, K. (1994). A second look at second-order belief attribution in autism. *Journal of Autism and Developmental Disorders,* 24, 577-586.

Taatgen, N.A. (2013). The nature and transfer of cognitive skills. *Psychological Review,* 120(3), 439-471.

Taatgen, N.A. & Anderson, J.R. (2002). Why do children learn to say "broke"? A model of learning the past tense without feedback. *Cognition,* 86(2), 123-155.

Triona, L. M., Masnick, A. M., & Morris, B. J. (2002). What does it take to pass the false belief task? An ACT-R model. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (p. 1045). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wahl, S. & Spada, H. (2000). Children's reasoning about intentions, beliefs and behaviour. *Cognitive Science Quarterly*, 1, 5-34.

Wellman, H., Cross, D. & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false-belief. *Child Development,* 72 (3), 655-684.

Wierda, S. & Arslan, B. (2014). Modeling theory of mind in Actransfer. In Szymanik, J. & Verbrugge, R. (eds.), *Proceedings of the Second Workshop Reasoning About Other Minds: Logical and Cognitive Perspectives* , CEUR Workshops Proceedings, 1208, 40-44.

Wimmer, H. M. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.