Chapter 1: Introduction

In which we give an overview of this dissertation and discuss the underlying theories and applied methodologies.
What should I write into this Introduction in order to attract your attention to read the whole dissertation or at least until the end of the Introduction? To achieve this goal, first of all, I should take your perspective by thinking about your background, and then I should find a way to provide enough information to make you understand the rest of this dissertation without putting you to sleep. In other words, I should use theory of mind reasoning (Premack & Woodruff, 1978) by appreciating that you might have different knowledge, beliefs and desires than my own.

Maybe you think that I believe that I found a way of getting your attention but you think that I am wrong. When I reason that “you think that I am wrong”, I use first-order theory of mind reasoning by attributing a mental state to you. Furthermore, I use second-order theory of mind reasoning when I reason that “you think that I believe that I found a way” by attributing a mental state to you who attributes a mental state to myself. Second-order mental state attribution is important in many different social situations, such as idiom understanding (Caillies & Le Sourn-Bissaoui, 2013), maintaining a strategic lie (Hsu & Cheung, 2013), and irony understanding (Filippova & Astington, 2008). For example, although John says “You sure are a great researcher”, John doesn’t really want Stefan to believe that he is a great researcher.

This dissertation is part of a project called “Cognitive systems in interaction: Logical and computational models of higher-order social cognition”, awarded to my first supervisor Rineke Verbrugge. The general motivation of the project is to provide a better understanding of higher-order theory of mind for the benefits of cognitive scientists, logicians and computer scientists. In the near future, humans will work together with artificial agents in daily life. Investigating the underlying mechanisms of the limitations in humans higher-order theory of mind reasoning will allow us to build more effective systems for better communication, collaboration and negotiation between these artificial agents and humans. To this end, I focused on children’s development of second-order theory of mind in terms of learning in decision making, transfer of skills, cognitive control, working memory and language.

In addition to contributing to developmental psychology and cognitive science in general, studying children’s development of theory of mind can contribute to one of the emerging research domains, which is child-robot interaction. Child-robot interaction aims to enhance children’s healthcare and education with interactive robots. Considering that children do not perceive robots just as running programs and that they attribute characteristics of living things to the robots (Belpaeme et al., 2013), it is important that the robots that interact with children “know” the limits of children’s theory of mind depending on their ages together with the underlying mechanisms of these limitations. Moreover, to
enhance the robots’ social and cognitive abilities, it is important that they “know” the effective ways to improve children’s theory of mind abilities. This dissertation aims to contribute to all above-mentioned fields by providing new insights about children’s development of second-order theory of mind.

Before focusing on the specifics of children’s development, it might be useful to have a general view about theory of mind. For this purpose, in the light of the literature, I will first try to provide short answers to a couple of general questions about theory of mind.

1.1. What is the limit of recursion in adults’ theory of mind reasoning?

Adults use theory of mind recursively up to the fourth level (e.g., “You want me to believe that my mother thinks that my aunt believes that my cousin is getting married”) to answer questions related to vignettes (Kinderman, Dunbar, & Bentall, 1998).

Kinderman et al. tested adults with five different vignettes, which involved complex social situations. The participants were expected to choose one of two possible answers for each question. While the proportion of incorrect answers was around 20% up to the fourth-level of theory of mind question, proportion of incorrect answers increased dramatically to 60% at the fifth level. Importantly, participants’ lower performance at fifth-order theory of mind reasoning was not due to forgetting the story facts.

In Chapter 4, we invoke the serial processing bottleneck hypothesis (Verbrugge, 2009) in order to provide a possible explanation to the question why processing embedded mental states requires more complex working memory strategies than simply remembering facts.

1.2. Can adults apply higher-order theory of mind flawlessly in different contexts as well?

Previous studies have shown that adults do not apply second-order theory of mind flawlessly in strategic games (Flobbe, Verbrugge, Hendriks, & Krämer, 2008; Hedden & Zhang, 2002; Meijering, van Maanen, van Rijn, & Verbrugge, 2010; Meijering, Taatgen, van Rijn, & Verbrugge, 2014).

In order to investigate the effect of context on the performance of theory of mind, Flobbe et al. (2008) constructed a two-player sequential game in which a human player and the computer opponent were expected to jointly drive a car by taking turns. Figure 1.1 depicts a screenshot of a second-order theory of mind phase of the game. There were three decision points represented by road junctions. The end points of the game were represented by the dead ends. The number of blue (dark gray) and yellow (light gray) marbles at each dead end represented each player’s reward. The participants were instructed to maximize their own reward (i.e., attain the highest possible number of blue marbles) and were told that the computer opponent would try to do the same (i.e., attain the highest possible number of yellow marbles). At the first road junction, the human player had to make a decision (i.e., to go to the dead end and finish the game or to go straight to continue). If the human player chose to go straight, the computer opponent made a decision. If the computer decided to continue, the human player was expected to finish the game by choosing to go right or to go straight in order to attain the highest possible reward. Therefore, in order to get the highest possible reward, the human player was expected to apply second-order theory of mind (e.g., “the computer ‘thinks’ that I would go straight to get seven marbles, therefore, the computer would go to the dead end at the second road junction. So, I should go...
to the dead end at the first road junction to attain the highest possible four marbles. The results showed that adults’ performance was not perfect and 75% of the adults used the correct second-order theory of mind reasoning.

Moreover, it has been shown that adults have difficulties applying even first-order theory of mind in online communication games (Dumontheil, Apperly, & Blakemore, 2009; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003).

These findings together suggest that the context does matter and that use of theory of mind in a dynamic game is not an automatic and flawless process.

### 1.3. In which environments does use of higher-order theory of mind have an advantage?

Agent-based simulation research has provided elaborate answers for the question in which environments use of higher-order theory of mind might have an advantage. De Weerd, Verbrugge and Verheij (2013) have studied the function of higher-order theory of mind in competitive games. To this end, they compared computational agents’ behavior in different game settings when the agents, who have different levels of theory of mind, play the games against other agents. Their results showed that using first-order and second-order theory of mind in a competitive game setting has an advantage over a competitor agent that does not take the perspective of the opponents and acts just based on the opponents’ observed behavior.

Moreover, although agents were able to cooperate without using theory of mind, using first-order and second-order theory of mind in a cooperative setting allowed the agents come up with an agreement faster (de Weerd, Verbrugge, & Verheij, 2015). Finally, in all competitive, cooperative and mixed-motive situations, use of second-order theory of mind has an advantage over use of first-order theory of mind (de Weerd et al., 2013; 2015; 2017).

### 1.4. Development of theory of mind

More than three decades of research have shown that theory of mind reasoning develops with age (Perner & Wimmer, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994; Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). It has been argued that being able to attribute a false belief to someone else provides evidence that a person has a theory of mind (Dennett, 1978). Since then, the false belief task has become a litmus test for theory of mind reasoning (Wellman et al., 2001). In a standard explicit false belief task, children are required to 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.

In the rest of this section, I briefly review the development of theory of mind starting from infants’ implicit false belief reasoning to young children’s explicit first-order false belief reasoning, and finally, to the main focus of this dissertation, which is older children’s development of second-order false belief reasoning.

#### 1.4.1. Implicit false belief reasoning

A number of studies have found that infants can pass implicit first-order false belief tasks (Baillargeon, Scott, & He, 2010; Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005). Different from explicit false belief tasks, infants are tested with nonverbal false belief tasks in which infants’ eye movements and looking times are measured.

In the implicit false belief task of Onishi and Baillargeon (2005), infants are first habituated by watching an adult who is putting a toy into a green box (Figure 1.2a, Trial 1) and reaching for the object in the green box (Figure 1.2a, Trials 2 and 3). Subsequently, infants are presented with the belief-induction trial where the toy is either moved from the green box to a yellow box while the adult is absent (Figure 1.2b, False-belief-green condition) or moved to the yellow box in the adult’s presence but then moved back to the green box after the adult leaves (Figure 1.2b, False-belief-yellow condition). Finally, in the test trial, the adult returns and searches for the toy either in the yellow box (Figure 1.2c, Yellow-box event) or in the green box (Figure 1.2c, Green-box event). In the false-belief-green condition, the infants who saw the yellow-box event looked reliably longer than...
those who saw the green-box event and, in the false-belief-yellow condition, the infants who saw the green-box event looked reliably longer than those who saw the yellow-box event.

The researchers concluded that infants do expect an agent to search for an object based on the agent’s belief about the object’s location, instead of the real location of the object. However, a group of researchers have argued that passing implicit false belief tasks does not necessarily require to have an implicit understanding of false beliefs and that behavioral reading is enough to pass the implicit false belief tasks (Heyes, 2014; Perner & Ruffman, 2005).

### 1.4.2. Explicit first-order false belief reasoning

To test explicit first-order false belief reasoning, children are presented with a story and expected to predict or explain the protagonist’s action. Figure 1.3a depicts an example of an explicit first-order false belief task. Maxi eats some of his chips and puts the remainder into the cupboard. Once Maxi leaves the kitchen, Sally takes the chips from the cupboard and hides the bag of chips in the oven. After a while, Maxi comes back to the kitchen and says: “I want to eat my chips”.

At this point, children are expected to answer the first-order false belief question, “Where will Maxi (first) look for his chips?”, often together with the justification question “Why does he look there?”.

While children around the age of four can correctly predict Maxi’s false belief about the location of the chips by saying “in the cupboard”, most children around the age of three make systematic errors and give the answer “in the oven” without taking into account Maxi’s ignorance about the real location of the chips (Wellman et al., 2001; 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).

Why can most 3-year-olds not pass explicit first-order false belief tasks although infants show an implicit understanding of false belief reasoning? It has been proposed that human theory of mind reasoning involves two systems, which are independent from each other (Apperly and Butterfill, 2009; Low, Apperly, Butterfill, & Rakoczy, 2016). System 1 is fast, automatic and inflexible. On the other hand, System 2 is slow, effortful and flexible. While System 1 allows infants to pass implicit first-order false belief tasks, System 2 can only be tested explicitly and develops with age (see Carruthers, 2016a, 2016b for criticisms related to two-system accounts). Similar to this explanation, Heyes and Frith (2014) proposed that humans are born with a ‘start-up kit’. This ‘start-up kit’ involves genetic neurocognitive mechanisms that allow us to have accurate expectations
about behavior of other agents and allows infants to pass implicit theory of mind tasks. In contrast, they propose that explicit theory of mind is a culturally inherited skill which is learned by verbal instruction.

One of the proposed theories in order to explain 3- to 4-year-olds’ transition from failure to success in the explicit first-order false belief tasks is the conceptual change theory (Gopnik & Wellman, 1994; Wellman et al., 2001). An explanation of the conceptual change theory is that children first construct naive theories to predict others’ behavior based on observable instances. However, these naive theories do not always lead children to correct predictions, especially when others have false beliefs. Based on these incorrect predictions, just like scientists do, children accumulate evidence and revise their naive theories to more elaborate theories that can also predict others’ behavior based on unobservable instances (Goodman et al., 2006; Gopnik & Wellman, 1994). There are three major lines of research that support the view that 3-year-old children do need a conceptual change to pass first-order false belief tasks. The first one is related to the findings that children who have more siblings pass first-order false belief tasks earlier (Ruffman, Perner, Natio, Parkin, & Clements, 1998). The second one is related to the training studies showing that when children are trained on understanding the concept of belief, their performance on the first-order false belief tasks improves (Slaughter & Gopnik, 1996). Finally, it has been shown that it is not possible to accelerate children’s first-order false belief reasoning without explicit feedback with further explanations (Clements, Rustin, McCallum, 2000). These studies together have been used as evidence for children’s conceptual change because children do need to conceptually understand that people might have different mental states which can be different from their own mental states.

An important factor in children’s development of explicit first-order theory of mind is language. Because explicit false belief tasks are verbal, it is not surprising that language is related to children’s transitions from failure to success. However, the important question is in what ways language matters and which components of language are crucial in children’s development of explicit first-order theory of mind (see Astington & Baird, 2005 for further details).

A group of researchers have shown that the semantics component of language is crucial for explicit theory of mind because the mental state verbs such as ‘think’, ‘want’, and ‘believe’ refer to unobservable states of the mind (Peterson & Siegel, 2000; Ruffman, Slade, & Crowe, 2002). Alternatively, another group of researchers have proposed that syntax is the crucial component of language, because complement clauses that involve mental state verbs, such as in “Maxi thinks that the bag of chips is in the cupboard”, allow us to represent states that contrast with reality or with other people’s mental states in terms of truth-value (de Villiers & Pyers, 2002; de Villiers, 2005; de Villiers, 2007). For example, in Figure 1.3a, while the statement “The bag of chips is in the cupboard” is false, the whole sentence “Maxi thinks that the bag of chips is in the cupboard” is true. A longitudinal study on children’s false belief reasoning and language has shown that both semantics and syntax contribute to children’s first-order false belief reasoning (Slade & Ruffman, 2005).

In contrast to the conceptual change theory, it has been claimed that children’s failure in explicit first-order false belief tasks is mostly due to the complexity of the tasks in terms of executive functions (Leslie, 1994; Scholl & Leslie, 1999; Baillargeon et al., 2010). Executive function is an umbrella term that covers a set of cognitive processes, such as working memory, inhibition, and cognitive control. The complexity camp proposes that children’s failure in explicit false belief tasks is due to children’s inability to keep track of events and to infer and hold in mind the contrary beliefs of other agents. Their problems are related to working memory, as well as inhibiting one’s own perspective, and to cognitive control (Carlson & Moses, 2001; Carlson, Moses, & Breton, 2002; Carlson, Claxton, & Moses, 2014; Davis & Pratt; 1995; Gordon & Olson, 1998; Keenan, Olson, & Marini, 1998).

Moreover, although it is not possible to rule out that children go through a conceptual change, the findings that individual differences in children’s executive functioning predict improvements in first-order theory of mind tasks signal that the complexity of the explicit first-order false belief tasks contributes to children’s transitions from failure to success (Benson, Sabbagh, Carlson, & Zelazo, 2013; Sabbagh, Hopkins, Benson, & Flanagan, 2010).

1.4.3. Second-order false belief reasoning

In order to investigate children’s further development of false belief reasoning after the age of four, children are tested with second-order false belief tasks (Braüner, Blackburn, & Polanska, 2016; Hollebrandse, van Hout, & Hendriks, 2014; Fernald & Wimmer, 1985; Sullivan et al., 1994).

Figure 1.3b depicts an example of a second-order false belief task. Different from the first-order false belief task, while Sally takes the chips from the cupboard and hides it in the oven, Maxi passes by the kitchen door and sees Sally hiding the chips in the oven. However, Sally does not see Maxi. After Maxi comes back to the kitchen and says: “I want to eat my chips”, children are expected to answer the second-order false belief question, “Where does Sally think that Maxi

---

3 For simulation theory and hybrid theories of theory of mind, see Harris (2002), Goldman (2006), and Leslie, Friedman & German (2004).

4 See Westra & Carruthers (2017) for a pragmatic account.
will look for his chips?”, together with the justification question “Why does she think that?”. Most children around the age of five cannot attribute a false belief to Sally who is attributing a belief to Maxi, while most 6- to 7-year-olds can (Perner & Wimmer, 1985; Sullivan et al., 1994; Miller 2009; 2012).

Why does it take children another one or two years to pass second-order false belief tasks once they already know that mental states of others might be different from one’s own? In line with the first-order theory of mind literature, two types of explanations have been proposed for children’s developmental lag between first-order and second-order false belief reasoning, namely conceptual change and complexity explanations (Miller 2009; 2012).

The conceptual change explanation suggests 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 (e.g., “Sally thinks that [Maxi believes that [the chips are in the cupboard]]”). This view implies that children need to hear many examples of recursive mental state talk in order to realize that mental states can be used recursively (Miller 2009; 2012).

On the other hand, the complexity explanation suggests that the higher complexity of second-order false belief reasoning adds further demands on executive functions, as does the linguistic complexity of the stories and the questions, in comparison to first-order false belief tasks.

As you may have already noticed when reading the previous subsection on children’s development of first-order false belief reasoning, there is an extensive research on that topic. On the other hand, the number of studies that focus on children’s development of second-order false belief reasoning is much more scant (see Miller, 2009; 2012 for a detailed review). Because some of these studies used composite scores of first-order and second-order false belief reasoning, we do not have an extensive knowledge on children’s developmental transitions from first-order to second-order false belief reasoning.

The Chapters 2, 3 and 4 of this dissertation cover most of the studies on children’s development of second-order false belief reasoning. For this reason, I do not give detailed explanations about those studies here (see Chapter 2 for a literature review about the previous computational cognitive models of first-order and second-order false belief reasoning; see Chapter 3 for training studies on second-order theory of mind; and see Chapter 4 for studies related to language and executive functions in relation to second-order false belief reasoning). However, in the following section, I present the research questions of this dissertation, which have not been investigated before. Subsequently, I introduce the methodology that we used in order to find answers to these questions. Finally, I give an overview of the rest of the chapters of this dissertation.

### 1.5. Research questions

Why can children not pass second-order false belief tasks once they are able to pass first-order false belief tasks? The main goal of this dissertation is to provide a possible answer for this question by investigating the components of children’s development of second-order theory of mind. For this purpose, for the first time in the literature, we investigated the following research questions:

1. Once 5-year-old children already have zero-order and first-order theory of mind strategies in their repertoire, do they predominantly use a zero-order theory of mind strategy or a first-order theory of mind strategy when they fail in second-order false belief tasks?

After knowing the level of theory of mind strategies of 5-year-olds, the following research questions arise:

2. 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?

3. Which types of feedback help 5-year-olds to revise their wrong strategies to the correct second-order theory of mind strategy in second-order false belief tasks?

In the previous research questions, we disregard the possible roles of language and executive functions in children’s development of second-order false belief reasoning. Considering the previous literature showing that language and executive functions have a role in children’s development of theory of mind, we also investigated the following research question:

4. Does working memory or syntactic recursion predict children’s second-order false belief reasoning?

Since first-order false belief reasoning is necessary in second-order false belief reasoning, we also investigated contributing factors of children’s development of first-order false belief reasoning. Considering that children do not learn first-order false belief reasoning by practicing false belief tasks, we aimed to investigate what kind of prior cognitive skills help children to pass the first-order false belief tasks. More specifically, we focused on the following research questions for the role of children’s prior executive functioning skills:

5. What might be the role of working memory strategies in children’s development of first-order false belief reasoning?

6. What is the common mechanism of 3-year-old children’s performance on the cognitive control tasks and the first-order false belief tasks?
1.6. Methodologies

In order to investigate children’s development of second-order theory of mind, this dissertation, for the first time, combines empirical experiments with computational cognitive modeling. Along with studying children’s development empirically, the computational cognitive modeling approach is a powerful method to provide insight into the underlying processes of human cognition. Figure 1.4 depicts an overview of the methodology that is used in this dissertation.

Ideally, computational cognitive modelers review the previous theories and construct models based on the available theories. Constructing models gives modelers the opportunity to implement cognitive processes and cognitive concepts with precision instead of using the concepts without teasing apart their components (e.g., response inhibition, memory inhibition, goal inhibition). After implementing a computational cognitive model, modelers run simulations and the simulation result brings out new predictions. These predictions can be tested empirically. If the empirical results fit the model’s predictions, the theory is revised based on the model’s assumptions; otherwise, the model-prediction-experiment cycle is repeated again.

However, sometimes the literature has some empirical data that are explained by opposing theories. In these cases, computational cognitive models are constructed based on the available data and this helps to differentiate the theories more precisely and to bring out new predictions.

In particular, implementing computational cognitive models in cognitive architectures has further advantages. First, cognitive architectures are not just software for constructing cognitive models. Instead, they reflect unified theories of cognition (Newell, 1973; 1990) in which a wide variety of tasks can be implemented under the same architecture. Second, cognitive architectures have some parameters that are set to a default value based on previous psychological experiments to simulate average human performance. For example, it takes 200 milliseconds to press a button on the keyboard once a decision has been made and the finger is ready to press it. Therefore, there are fewer degrees of freedom in terms of the parameters of the models. Third, a single model can allow modelers to make predictions about different modalities, such as reaction times, accuracy, eye-movements and Blood Oxygenation Level Dependent signal (BOLD).

In this dissertation, I constructed computational cognitive models by using two different cognitive architectures, namely ACT-R and PRIMs. Below, I give a brief overview about the relevant mechanisms of these architectures.

1.6.1. ACT-R

Adaptive Control of Thought–Rational (ACT-R; Anderson, 2007) is a hybrid symbolic/sub-symbolic cognitive architecture. ACT-R consists of several modules, each associated with a specific region in the brain (see Figure 1.5).

Similar to the participants in behavioral experiments, ACT-R can “see” or “hear” the presented stimuli through its visual module, retrieve a chunk of information from its declarative module, hold the retrieved chunk in its “mind” temporarily through the problem state, change the current goal to another one through the goal state buffer, and finally, give a “verbal” or “motor” response through its manual module.

Knowledge is represented in two different memory systems in ACT-R. While the declarative memory represents factual knowledge in the form of chunks (i.e., “The capital of France is Paris”), procedural knowledge (i.e., how to ride a bicycle) is represented by the production rules in the form of IF-THEN rules. The procedural knowledge and the factual knowledge interact when production rules retrieve a chunk from the declarative memory.

At any time, the central pattern matcher checks which IF parts of the production rules match the current goal of the model, and if multiple production rules match the current goal, then the rule that has the highest utility value is executed. The utility value is calculated from estimates of the cost and probability of reaching the goal if that production rule is chosen. Noise is also added to the expected utility of a production rule, making production rule selection stochastic.
When a production rule is successfully executed, the central pattern matcher checks again for production rules that match the current goal. Thus, cognition unfolds as a succession of production rule executions.

Over the years, many ACT-R models have been constructed in different domains, such as perception and attention, language processing, problem solving and decision making, learning and memory, education and cognitive neuroscience.

### 1.6.2. PRIMs

Different from ACT-R, the primitive information processing elements theory (PRIMs; Taatgen, 2013) has been constructed specifically as a theory of skill acquisition and transfer of skills. It has been implemented in the cognitive architecture PRIMs (see Figure 1.6).

The cognitive architecture PRIMs adopts the mechanisms of the declarative memory of ACT-R. However, in addition to chunks of factual information, the PRIMs architecture has operators and goals in declarative memory. Operators, similar to production rules in ACT-R, are in the form of IF-THEN rules (condition-action). The PRIMs architecture breaks down the complex production rules of ACT-R, which represent procedural knowledge (i.e., how to drive a car), into a fixed number of smallest possible elements, named PRIMs. PRIMs only move, compare or copy information between modules (i.e., declarative, visual, and motor modules) independent from the content of the information. Operators combine these PRIMs together to perform a task. However, unlike production rules, operators, just like other chunks in declarative memory, have base-level activations and associative strengths.

In the PRIMs architecture, there is no hard connection between goals and operators. Current goals of the model activate operators to achieve those goals. If an operator is successfully used to complete a goal, the strength of association between the goal and the operator increases.

According to the PRIMs theory, transfer of skills occurs either based on the transfer of the task-general sequences of PRIMs or based on training a particular strategy, which is represented by operators. We provide a more detailed explanation of the transfer of skills based on the transfer of the task-general sequences of PRIMs in Chapter 5 and based on training a particular strategy in Chapter 6.

Taatgen (2013) showed the predictive power of the PRIMs architecture by modeling a variety of transfer experiments such as text editing (Singley & Anderson, 1985), arithmetic (Elio, 1986), and cognitive control (Chein & Morrison, 2010).
1.7. Overview of the Dissertation

As I mentioned in Section 1.5, the main goal of this dissertation is to provide a plausible explanation to the question why children cannot pass second-order false belief tasks once they are able to pass first-order false belief tasks. We will achieve this by breaking up the components of children's development of second-order theory of mind.

One of the components we investigate is children's strategy selection and revision in second-order false belief tasks. Chapter 2 is devoted to answer the research questions: Do 5-year-olds who fail in second-order false belief tasks predominantly use zero-order theory of mind or first-order theory of mind reasoning? And how do they revise their wrong theory of mind reasoning strategy to the correct second-order theory of mind reasoning strategy over time? To this end, in Chapter 2, we present two computational cognitive models in which two possible learning mechanisms of decision making are implemented in the ACT-R cognitive architecture, namely reinforcement learning and instance-based learning. Subsequently, we present our cross-sectional study with 5-year-old children, which is conducted to test the reinforcement learning and the instance based-learning models' different predictions about five-year-olds', wrong answers to second-order false belief questions.

The second prediction of the reinforcement learning and instance-based learning models is related to the role of different types of feedback in children's development of second-order false belief reasoning. Different from the first-order theory of mind literature, both of the models predict that it is possible to accelerate 5-year-olds' second-order false belief reasoning by providing feedback “Correct/Wrong” without further explanations. However, unlike the reinforcement learning model, the instance-based learning model predicts that if feedback with further explanations is provided, children's performance would increase more. In Chapter 3, we test these predictions by training children with many different second-order false belief stories in four different conditions, namely feedback with explanation, feedback without explanation, no feedback and an active control condition in which theory of mind reasoning played no role.

Until Chapter 4, the assumption was that children do not have major problems in their executive functioning abilities to attribute second-order false beliefs and with the linguistic structure of the second-order false belief questions. In Chapter 4, we investigate two important components of children's development of second-order theory of mind, namely working memory and language. More specifically, we focus on the relationship between simple and complex working memory strategies and children's second-order false belief reasoning as well as the relationship between syntactic recursion and second-order false belief reasoning. Furthermore, we investigate whether syntactic recursion or working memory is the best predictor of children's second-order false belief reasoning. Finally, we invoke the serial processing bottleneck hypothesis to propose a procedural account for the role of complex working memory strategies in second-order false belief reasoning.

Since first-order false belief reasoning is necessary in second-order false belief reasoning and since, in Chapter 4, we found that the main predictor of second-order false belief reasoning is working memory, in Chapter 5 and Chapter 6, we also investigate the role of executive functions in children's development of first-order false belief reasoning. In Chapter 5, we present a PRIMs model in order to investigate the question: How do simple and complex working memory strategies help children to pass first-order false belief tasks? In Chapter 6, we present another computational cognitive model of transfer of skills, which is implemented by using the PRIMs cognitive architecture. Our computational modeling approach provides a procedural account for the existing experimental data showing that there is a mutual transfer between children's cognitive control and first-order false belief reasoning, meaning that training 3-year-old children with a cognitive control task helps them to pass the first-order false belief task and vice versa.

Finally, in Chapter 7, I discuss our findings and give pointers to possible future research to have a better understanding of children's development of second-order theory of mind.