The Relationship Between Theory of Mind, Inhibitory Control and Children's Behavioral Problems – A Multi-Informant Approach

Dissertation zur Erlangung des akademischen Grades
eines Doktors der Philosophie (Dr. phil.)
der
Erziehungswissenschaftlichen Fakultät
der Universität Erfurt

vorgelegt von Karen Kühn

Erfurt 2016

Erstes Gutachten: PD Dr. David Buttelmann (Universität Bern)
Zweites Gutachten: Prof. Dr. Claudia Steinbrink (Universität Erfurt)
Drittes Gutachten: Prof. Dr. Gisa Aschersleben (Universität des Saarlandes)
Tag der Disputation: 07.04.2017
Datum der Promotion: 07.04.2017
urn:nbn:de:gbv:547-201700061

Acknowledgements

This thesis would not have been possible without the help and dedication of many people. Firstly, I want to thank my supervisor David Buttelmann. I'm very grateful for his support, his valuable ideas, and his continuous motivation, which impressed me again and again during the time of my research. He was a reliable research partner and always available for intensive discussions, sharing his extensive knowledge and experience to help me in the best possible way with my work. I would also like to thank Tilmann Betsch for his helpful suggestions and constructive criticism, in particular, his expert advice on the evaluation of the data. Furthermore, I want to express my gratitude to the members of the Junior Research Group. I was actively supported with well-organized scheduling, technical support, and regular professional advice by Kata Gellén, Frances Buttelmann, Andy Schieler and Anja Flödl. I would like to particularly thank Kata Gellén for the linguistic assistance she provided during the writing process. Likewise, I would like to thank the members of the Center for Empirical Research in Economics and Behavioral Sciences to whom I was allowed to present my work regularly and who provided valuable feedback. My gratitude also extends to numerous people who were involved in the data collection process: Angelika Radeck, Friederike Degwitz, Carolin Mehrmann, Carolin Seifert, Daniel Knackstedt, Katja Richter, Maria Rintelmann, Franziska Köppen, Sophie Lachmann, Franziska Schmidt, Daniel Neumeister, Christina Predatsch, Sandra Patting, Laura Andres-Klein, Antje Pötzsch, Silvana Karau, Daniela Mückenheim, Isabell Weiher, Jan Kotzerke and Laura Lindner.

I want to express my heartfelt thanks to my mother and father who encouraged me not only during this thesis, but my whole life with good advice and unconditional warm-heartedness. Finally, I am especially grateful to the man who has always been by my side at the end of the day, who encouraged me when I doubted and motivated me by believing in me; I deeply thank Frank Sonnabend.

Abstract (German)

Die Beziehung zwischen Theory of Mind (ToM), inhibitorischer Kontrolle (IK) und Verhaltensproblemen von Kindern im Vorschulalter wurde bisher empirisch wenig untersucht und zeigt zudem uneinheitliche Ergebnisse. Eine Erklärung für diese Ergebnisse könnte darin begründet sein, dass der Fokus meist auf einer einzelnen Dimension der ToM lag, dem Verstehen falscher Überzeugungen. Die alleinige Untersuchung dieser Dimension reicht jedoch nicht aus, um ein umfassendes Profil der ToM im Vorschulalter abzubilden. Zudem blieb IK, welche häufig einen positiven Zusammenhang zur ToM zeigt, oftmals unbeachtet. Die Mehrheit der Studien verwendete verbale Aufgaben und untersuchte daher Kinder ab 4 Jahren. Die Frage nach einem früheren Zeitpunkt eines Zusammenhangs bleibt damit ebenfalls offen. Um umfassende Informationen über die Zusammenhänge zwischen ToM, IK und Verhaltensproblemen in der frühen Kindheit zu liefern, wurden Kinder im Alter von 2, 3 und 4 Jahren (*N*= 252) mit einer umfangreichen Batterie von ToM- und IK-Aufgaben untersucht. Zudem wurde ihr Sprachverständnis, Verhalten und Temperament mit einem Multi-Informanten Ansatz erfasst. Das Sprachverständnis steht ab dem Alter von 2 Jahren in positivem Zusammenhang zur IK, sowie ab 4 Jahren in positivem Zusammenhang zur ToM. Zudem zeigte sich, dass 4-jährige Kinder mit hohen IK-Werten wenige Verhaltensprobleme zeigten. Gegensätzlich dazu zeigten 4-jährige Kinder mit hohen ToM-Werten mehr Verhaltensprobleme. Das Temperament zeigte keinen einzigartigen Zusammenhang zu IK und ToM. Die Ergebnisse weisen drauf hin, dass sich ein direkter Zusammenhang zwischen ToM, IK und Verhaltensproblemen erst ab einem Alter von 4 Jahren entwickelt. Des Weiteren werden mögliche Erklärungen für die Beziehung von ToM, IK und Verhaltensproblemen unterschiedlicher Erscheinungsform diskutiert.

Schlagwörter: Theory of Mind, Inhibitorische Kontrolle, Temperament, Verhaltensprobleme

Abstract (English)

The relationship between Theory of Mind (ToM), inhibitory control (IC) and behavioral problems has attracted little empirical investigation to date and delivered mixed findings. An explanation for these mixed findings might lie in focusing on a single mental state (i.e., false-belief understanding), which might not account for a comprehensive profile of children's ToM. Furthermore, IC, often positively correlated to ToM, remained unattended by the majority of studies. Due to the use of verbal tests, investigations mostly focused on children from 4 years of age onwards. Thus, the time of emergence of a relationship also remains an open question. To provide comprehensive information on possible correlations between ToM, IC and behavioral problems in children's early years, 2- , 3- and 4-year-old children (*N*= 252) were presented with a broad battery of ToM and IC tasks, and tested for receptive language abilities, complemented by comparable multi-informant assessment of their behavior and temperament. Language was positively correlated to IC from 2 years onwards, and to ToM only at 4 years of age. With regard to caregiver ratings, I found that for 4-year-old children higher scores in IC were associated with fewer behavioral problems. In contrast, higher scores in ToM were associated with more behavioral problems. No such associations were found for 2- and 3-year-old children. Considering temperament dimensions, only the measure of activity was negatively correlated to IC at the age of 2 years. However, taking language abilities into account the unique contribution disappeared. The results suggest that robust relationships between ToM, IC and behavioral problems start to develop at the age of 4 years. Different explanations for the patterns of association will be discussed, especially for the contribution of ToM and IC to the development of different manifestations of behavioral problems.

Keywords: Theory of Mind, Executive Functions, Inhibitory Control, Temperament, Behavioral Problems

Table of Contents

Acknowledgements III

Abstract (German) IV

Abstract (English) V

Table of Contents VI

1 Introduction 1

1.1 Theoretical Background 3

Theory of Mind – Different Mental States 3

Executive Functions – Inhibitory Control 14

Temperament 17

Behavioral Problems in Childhood – Classification, Prevalence and Possible Correlates 21

1.2 Relations Between the Constructs 24

The Relationship Between Theory of Mind and Inhibitory Control 24

Language Abilities – an Important Correlate of Theory of Mind and Inhibitory Control 27

The Relationship Between Theory of Mind and Children's Behavior 30

The Relationship Between Inhibitory Control and Children's Behavior 32

The Relationship Between Theory of Mind, Inhibitory Control and Children's Temperament 34

2 Dissertation Project 36

2.1 Method 41

Participants 41

Procedure 42

Measures 42

2.2 Results 69

Descriptive Statistics 69

Gender Differences 70

Session Order Effects 70

Correlation and Comparison of Parents' and Caregivers' Ratings 71

Correlations Among Age, Language, Theory of Mind, Inhibitory Control, Behavioral Problems and Temperament of 2-Year-Old Children 73

Multiple Regression Analyses for the Sample of 2-Year-Old Children 75

Correlations Among Age, Language, Theory of Mind, Inhibitory Control, Behavioral Problems and Temperament of 3-Year-Old Children 76

Correlations Among Age, Language, Theory of Mind, Inhibitory Control, Behavioral Problems and Temperament of 4-Year-Old Children 78

Multiple Regression Analyses for the Sample of 4-Year-Old Children 81

2.3 Discussion 85

The Relationship Between ToM, IC, Behavioral Problems and Language 85

The Relationship Between IC and Behavioral Problems 92

The Relationship Between ToM and Behavioral Problems 98

Caregivers' and Parents' Judgments 105

Temperament as a Possible Correlate 107

Graphical Representation of the Current Model of Correlations 111

Remarks About Test-Retest Reliability and Task Construction 112

Gender Differences and Internal Consistency of Tasks 120

Possible Limitations of the Current Study 123

2.4 Conclusion 126

References 128

Appendix 157

Appendix A - Material of the Intention-Understanding Task for 4-Year-Old Children 157

Appendix B – Material of the Preference-Understanding Task for 4-Year-Old Children 162

Appendix C - Analyses of ToM and IC Batteries 166

1 Introduction

The research field of psychology has always tried ever since to explain mental processes and behavior and to set both areas in relation to each other in order to predict human behavior based on psychological characteristics. An important aspect which distinguishes humans from other species is the development of social-cognitive abilities, such as social engagement and mental representations (Tomasello, Carpenter, Call, Behne, & Moll, 2005). Humans as social beings surround themselves and communicate with others. To enable sophisticated social cooperation, it is essential to interpret other people's thinking and behavior. Only in this way is it possible to successfully interact with one another. During the course of development, human beings acquire the unique ability to understand others' mental states. Cognitive growth is caused by interactive and co-operational processes with others (Vygotskiĭ & Cole, 1978), and cognitive processes, in turn, are reflected in behavior. Thus, it is logical to assume that the social-cognitive ability to understand the others' mind may influence one's own behavior and emotional state. Everything a person perceives will be considered in relation to herself, integrated into existing schemes, and will be further used to create a view of the world and the people in the person's environment. If the perceived content of the social environment thus provides the basis for how a person interprets the world, then this perceived content would also affect the way this person reacts in response to her environment. In addition to social-cognitive abilities, other cognitive processes might also play a role in behavioral manifestations; more precisely, the ability to manage one's own thoughts and actions is of significance. A society, conceived as a space of cohabitation, is subject to normative rules, which at best should be tracked by all members (Durkheim, 1950). Thus, social coexistence requires a certain degree of normative behavior. Deviating behaviors might be perceived as unusual, disturbing or inappropriate, because they hold the potential to jeopardize a successful participation in social interactions. The occurrence of behavioral problems for centuries in works of literature, e.g., „The Story of Fidgety Philip“ (Hoffmann, 1847), can be interpreted as signifying that certain forms of behavior have been considered problematic since time immemorial. The research field of psychopathology continuously captures the differences in behavior, classifies behavioral problems and tries to explain their origins. However, to predict the occurrence of behavioral problems or to alter affect existing behavior, crucial determinants must be identified. The investigation of possible risk factors, would allow for early prevention of behavioral problems or to initiate interventions to support the child's development in a positive way. Studies have shown that behavioral problems are not always a temporary phenomenon and that behavioral problems, which roots can be found early in childhood, can manifest later (Campbell, Pierce, March, Ewing, & Szumowski, 1994; Hinshaw, 1992). Therefore, one responsibility of research is the identification of early predictors. Besides other factors, it is conceivable that social-cognitive abilities and the ability to inhibit one's own thoughts and actions are related to the formation of behavioral problems, since both concepts underlie human actions.

Therefore, this dissertation addresses the question of whether the ability to understand others' mental states and the ability to control one's own thoughts and actions are linked to possible behavioral problems at preschool age. To answer this question comprehensively, the investigation will include children from three different age groups to gain evidence from different stages of early development. Results of this study will provide important information about possible correlates of behavioral problems which create the foundation for conceptual designs of successful prevention programs.

1.1 Theoretical Background

For this study, which investigates possible correlates of preschoolers' behavioral problems, I will provide a brief theoretical background about the constructs involved. I will first define the term 'Theory of Mind' and give an overview of the different dimensions of mental state understanding by introducing four important mental states, including common methods of measurement and empirical findings for specific age groups. I will continue with clarification about the processes of 'Executive Functions' and focus on inhibitory control, as one important domain of behavior regulation mainly developing during childhood. Aside from the social-cognitive constructs, I will discuss two aspects of behavioral characteristics important for this study. First, temperament will be defined, divided into its dimensions and explained in terms of the effects on children's behavior. Second, I outline the characteristics of behavioral problems occurring in childhood, classify specific syndromes and will briefly report the main correlates, which are expected to be of significance.

Theory of Mind – Different Mental States

As social beings, humans communicate as soon as their life begins. Forms of communication start with simple nonverbal gestures and develop into complex verbal scenarios. Regardless of its level, communication is always directed at another individual, at least at one interaction partner. The sender transfers a message to the receiver and the receiver has to decode the incoming signals. Even for verbal messages, the process includes unobservable components, which require not only decoding but also reasoning about the receiver's intended message (Sperber & Wilson, 2002). The ability to reason about others' thoughts is a unique part of human cognition, underlies mental representation processes, and belongs to the mechanism of mental state understanding. The capacity to attribute mental states to others and to oneself is commonly known as the 'Theory of Mind' (ToM) and can be considered as a crucial cognitive development during childhood (Gopnik & Astington, 1988; Premack & Woodruff, 1978; Wellman & Liu, 2004). Depending on the situation, people can hold several mental states. Imagine an everyday situation: A child wants to go to a lake to catch a fish with a net. This seemingly simple act involves a complex structure of internal states, which will lead to a certain action performance. To start with the beginning of the chain, the child holds the *desire* to catch a fish. He *believes* there are fish in the lake. Based on this, an *intention* is created by setting the *goal* to catch a fish and choosing the specific means to do this with the net. This example emphasizes four different mental states: goals, intentions, desires, and beliefs, which all are of significance in cognitive development. An important characteristic of mental states is that they are unobservable but are reflected in people's behavior and, therefore, have to be inferred by interpreting others' actions in a larger context. Possible developmental processes of the ToM continue to be discussed in science. Explanations show two prominent directions, the so-called positive gain accounts and the negative-release accounts (Low & Perner, 2012). The positive-gain accounts suggest that different systems are operating in mental state understanding. The first limited system that is used to explain other's mentalistic actions is active in infants' early years before a second system starts to function when children start to explicitly express false-belief understanding (Apperly & Butterfill, 2009; Low, 2010). Characteristics of the systems vary according to accounts, but there is a general agreement about a conceptual change, which enables children to make use of the second system. This perspective also includes a distinction between different mental states, claiming that children first register goals and intentions, before mental state reasoning in the form of belief understanding evolves (Apperly & Butterfill, 2009; Tomasello et al., 2005). Similarly, another account suggests that ToM tasks are solved by applying behavioral rules and that children only later develop a deeper understanding of behavior (Apperly & Butterfill, 2009; Perner & Ruffman, 2005). Developmental relations are expected, but should appear situation-specific rather than on a general basis. The second direction includes the negative-release accounts. Models of this approach suggest that the ability to understand others' mental states in its full dimension is present from very early on, but the ability to express it, e.g. by the use of language and with the help of inhibition skills, is crucial and needs to be developed for successful operation in certain situations (Baillargeon, Scott, & He, 2010; Carruthers, 2013; Leslie, Friedman, & German, 2004). In this sense, it is suspected that even infants apply psychological reasoning in a mentalistic manner, although it is not excluded that experience might play a role in development (Luo & Baillargeon, 2010). Neither of the two directions has been undoubtedly confirmed and discussions about the developing mechanisms are still ongoing. Within the framework of this dissertation a clarification of this issue cannot be addressed. However, it is undisputable that mental state understanding is a sophisticated operation of social cognition, grounded in the question of asking *why* humans act in the way they do. Recent research reported that signs of an understanding of different mental states appear early in life, long before children's fourth birthday. In the following, these findings and the characteristics of four mental states will be described.

Goals. Humans show the unique ability and aim to cooperate with others by sharing their psychological states (Tomasello et al., 2005). To successfully interact, an understanding of others' actions in terms of their goals is needed. A fundamental and early developing ability is to understand others' actions as goal-directed. By identifying another person's goal, one can answer the question of "*What* a person is doing." (Buttelmann, Carpenter, Call, & Tomasello, 2008). This means that an individual recognizes and connects the behavior of others related to a certain aim. Consequently, the individual represents this aim as a mental representation of the end state, and precisely this mental representation corresponds with the term *goal* which is used for the purpose of this dissertation. Studies have revealed that infants already start to interpret others' actions by inferring another person's goal before their first birthday, e.g. by reproducing only goal-relevant parts of actions (Behne, Carpenter, Call, &

Tomasello, 2005; Hamlin, Hallinan, & Woodward, 2008). Moreover, even by the age of 6 months, infants are able to encode the goal object of a reaching act demonstrated by a model (Woodward, 1998). In a habituation paradigm, infants were shown two objects simultaneously which had been placed next to each other (object A and object B). In a familiarization phase, they observed a human hand repeatedly grasping object A but not object B. For the test phase the positions of the objects were switched. Results showed that infants looked reliably longer when they observed the situation when the hand used the old path but now grasped the new object B, compared to the situation when the hand went the new path and grasped the old object A. Woodward (1998) introduced additional condition, where infants saw the same procedure with a mechanical claw instead of the human hand and could not find the same effect, and interpreted this difference as evidence for actual goal interpretation instead of the allocation of visual attention. Other empirical evidence for a selective process of action interpretation is delivered by various imitation paradigms. Different studies used scenarios of simple actions, with an underlying complex structure of multiple goals, demonstrated by a model (i.e., gestures implying more than one goal). If cognitive capacity is limited early in the development, children are not able to consider all the goals presented by an actor. Therefore, they break down the complexity by imitating the goals which appear as the most relevant ones. An action is then separated into various goals. For instance, in a study, 18-month-old children observed an actor presenting different actions and tended to copy these actions selectively, depending on what they identified as the actor's goal (Carpenter, Call, & Tomasello, 2005). Further results demonstrate that goals are organized hierarchically, existing on multiple levels, and that the capacity for goal understanding improves with age (Bekkering, Wohlschläger, & Gattis, 2000).

Intentions. More evidence for an early interpretation process of actions and specifically, a differentiation between the concepts of goals and intentions, is given by children's rational imitation of actions. Infants at 14 months of age imitated actions in a goal-directed way by considering the context in which this action took place (Gergely, Bekkering, & Király, 2002). It is suggested that infants took the actor's intention into account (Buttelmann, Carpenter, Call, & Tomasello, 2007). In the study of Gergely et al. (2002) infants saw a model who performed an unusual, irrational action to reach a goal (i.e., the model sat at a table and turned on a lamp, which was placed on the table, by using her head instead of her hand). In one condition the model was physically unconstrained and performed the unusual, irrational action willingly (i.e., had her hands free while demonstrating the act). In another condition the model was physically constrained and, therefore, was forced to perform the unusual, irrational action. (i.e., held a blanket wrapped around her body). When infants themselves got the chance to turn on the lamp they produced the unusual, irrational action, if they previously saw the model in the unconstrained condition; whereas infants did not reproduce the unusual, irrational action but performed the rational act to reach the goal (i.e., infants used their hands to turn on the lamp), if they previously saw the model in the constrained condition. A clear distinction between a goal and an intention appears challenging. Goals can be organized in a hierarchical structure. This means that a person can have a goal, which consists of several sub-goals, however, the specific chosen means to achieve a goal can be considered as the intention. To distinguish between both concepts, intention can be defined as "a plan of action the organism chooses and commits itself to in a pursuit of a goal" (Tomasello et al., 2005, p. 676). This definition of intention will be used for the following work and refers to the question "*How* a person is trying to achieve a goal." With this description in mind, an intention always is connected to a certain goal and is involved in forming and implementing plans (Bratman, 1987). In the work of Heckhausen and Gollwitzer (1987), within the field of motivation psychology, this formation process was also depicted in the Rubicon Model, which describes action formation and execution as a process of transition from a motivational state of mind to a volitional state of mind. According to this model, in the motivational state of mind people desire something and set a certain goal, whereas in the volitional state of mind people form actual intentions and commit themselves to a specific way to reach this goal. In contrast to a goal, which can be achieved in different ways, an intention has to be fulfilled in the exact way it was intended to be carried out (Schult, 2002; Searle, 1983). Even if the aforementioned definition separates the terms from each other, for some cases it remains difficult to maintain a distinction between them and the observer's perspective is decisive. An action itself could also be a person's goal, for instance if it does not cause a special effect on the environment or is directed at an end-state, e.g., a dancer's goal is simply to perform a bodily movement (Tomasello et al., 2005). However, if there might be a higher goal, e.g., to entertain an audience, then the former goal (bodily movement) transforms into the specific means to achieve that higher goal.

In line with the study of Gergely et al. (2002), further evidence of an rudimentary intention understanding was found for infants from 12 months of age onwards (Buttelmann et al., 2008; Zmyj, Daum, & Aschersleben, 2009). Infants imitated the tool use of a model more often when this model freely chose to use the tool compared to when the model was forced to do so due to physical constraints. This indicates that children have an insight beyond surface behavior. Recognizing a failed attempt reflects these abilities. By the age of 18 months infants gained information from failed attempts and showed an interpretation of observed actions (Meltzoff, 1995). Later in development a more sophisticated understanding of intentions emerges that concerns the causal relation between an intention and an action in terms of a goal and which includes the concept of intentionality. At this point it is important to note that there is a clear differentiation between *intentions* and *intentionality*. Whereas an intention refers to the mental state, and to the specific means created to achieve a certain goal, intentionality refers to the awareness to perform an action (Malle & Knobe, 1997). To simplify, one could say an act is performed intentionally or by accident. Namely, it is possible that an intention is created, but not fulfilled because an unintended action happened, which nevertheless might have led to the same goal. To give an example, imagine that a person creates an intention, which is driven by the goal to feed fish and the specific means of putting one piece of food after another into the water. If this person accidentally drops all the food at the same time into the water, the goal to feed the fish is fulfilled but not by the intended means. If conditions allow the existence of a fulfilled goal and at the same time that of an unfulfilled intention, children at the age of 3 years struggle with distinguishing between the two mental states (Schult, 2002). In an experiment, a situation was created where participants made a plan of action as to how to achieve a certain goal (i.e., children chose one of two boxes to hit with a ball, because they suspected a reward was in this box). They indeed achieved their goal (i.e., they found the reward), but not by the planned action they previously set (i.e., they accidentally hit the other box and surprisingly found the reward in this box). The fulfilled goal seemed to dominate children's representations, because they ignored their prior intention and alternated it in favor of matching their goal. When asking for their prior intention (i.e., "Which one were you trying to hit?"), they answered by naming the actual event that happened and brought success, not by stating their former plan of action. Only later, by the age of 4 and 5 years were the answers correct. Further supported by other findings, it is suggested that during later preschool age, children build up the awareness of the commitment that intentions entail and that children start to distinguish between the mental states *goal* and *intention* (Astington, 2001; Schult, 2002).

Desires and preferences. Desires are the start of the chain by determining goal and intention formation, since they cause the general motivation for certain actions (Heckhausen & Gollwitzer, 1987; Moses, Coon, & Wusinich, 2000). Desires can be classified in an objective or in a subjective concept. The objective concept represents the relation between a person and an object in the physiological sense of approaching, grasping or referring to an object in various manners, whereas the subjective concept represents the attitude of a person towards something, e.g., a desired end state or a specific object a person is longing for (Doherty, 2009). For the purpose of this work the subjective concept of desires, which refers to attitudes that form the basis of behavior, will be of significance. Attitudes are psychologically generated and built upon experience, hence, they vary between individuals. They are relatively consistent over time and can be expressed by emotional signals such as showing pleasure towards or aversion to a certain object. As a result of their continuity, attitudes become apparent in preferences that people create over the course of their development. Physiologically-generated desires (e.g., being thirsty and therefore desire something to drink), however, show no consistency over time (Moses et al., 2000) and change or disappear as soon as the specific desire is fulfilled. For the following investigations, the focus will lie on attitude-generated desires in the sense of preferences. Experiments revealed an early understanding of preferences. Thus, the ability to interpret another person's emotional communication was already seen for infants at 14 months of age (Repacholi, 1998). Infants were able to understand emotional signals, which were demonstrated by a model and referred to an object. Later, by the age of 18 months, infants were able to recognize that other people hold preferences that differ from their own and, furthermore, were able to act in accordance with this knowledge (Repacholi & Gopnik, 1997). Infants observed an experimenter who showed a preference for a food they themselves rejected, and who showed a rejection for a food they themselves preferred. In this situation the experimenter showed preference for a piece of vegetable and rejection for a cracker. When the experimenter requested after his demonstration that infant give him one of the two foods, infants were able to identify the correct one and give him the piece of vegetable. This indicates a decoupling process, because children had to restrain their own preference and perceive the deviating preference of someone else. However, complexity of mental states rises with extended scenarios. Taking a situation of decision-making as an example with the possibility of various choices, it is conceivable that a person holds multiple desires or even that desires conflict within this person's mind. Bennett and Galpert (1993) found 5-year-old children understand the existence of multiple simultaneous desires, e.g., wanting two things at the same time. In their study, children listened to a story of a person who wanted to go to two birthday parties, which were held at the same time. They were asked if they think that the person could be at one of the birthday parties but wants to be at the other birthday party at the very same time, and the significant majority of children answered with yes. On the other hand, desires that cause an internal conflict, e.g., wanting something and not wanting it at the very same time because of certain circumstances, are understand only by 11 years of age. Described in more detail, in the cited study only 11 year old children understood that a person wants to go to a party, but also does not want to go because there is another party at the very same time that he wants to join (Bennett & Galpert, 1993).

Beliefs. The last dimension of mental state understanding, which is significant for the following study, is beliefs. To understand that people hold beliefs requires the imagination that other people can mentally represent any possible content of the world (Wellman, Cross, & Watson, 2001). To symbolize this, one can imagine that humans represent each other with thought bubbles containing representations of objects, situations or any other aspects of the world. One characteristic of beliefs is that they do not necessarily reflect reality. Beliefs are supposed to be true, since people's behavior is based on their knowledge about certain circumstances. However, mistaken assumptions due to lack of knowledge are part of everyday life and accordingly, beliefs also appear to be false. The ability to read others' false beliefs is a sophisticated cognitive performance, because it requires the reconciliation of own knowledge, which is necessary to detect deviations. To investigate preschoolers' false-belief understanding, two standard paradigms, based on verbal constructs were created during the 1980s and have subsequently gained wide acceptance. Firstly, I will describe the unexpected-transfer task (Wimmer & Perner, 1983). Children observe a scene with two covered containers (A and B) positioned next to each other. The first agent enters the scene carrying an object, places the object into container A and covers the container. The agent leaves the scene and in his absence the second agent appears, transfers the object from container A to container B, covers the containers and leaves as well. The test question for this task is where the agent will search for his object when he returns. Secondly, I will describe the unexpected-content task which is another widely used measurement (Hogrefe, Wimmer, & Perner, 1986). Children observe a scene with a closed box. This box has to be familiar to the children (e.g., a box of well-known chocolate). After presenting the closed box to children they are asked what they think is inside of the box. If they answer by naming the expected content (e.g., 'chocolate' or 'sweets') the box will be opened and an unexpected content (e.g., pencils) will be revealed. Afterwards, the box will be closed again and children receive the test question. The question here is about what a person, who was absent when the unexpected content was revealed, would think is in the box. Both paradigms investigate an explicit false-belief understanding by demanding verbal answers and requiring the explicit statement about the false belief of another person. Children around the age of 4 to 5 years are able to significantly pass these traditional tasks. Further research showed successful performances by 3-year-olds by reducing the task demands, especially the verbal requirements (Hansen, 2010; Rubio-Fernandez & Geurts, 2012). During the last two decades non-verbal paradigms were invented to provide evidence for an early implicit understanding of others' false beliefs. Experimental study designs for infants often involve looking time paradigms. For investigating an implicit understanding of false beliefs, the violation-of-expectation paradigm is widely used. Participants observe a scene where an agent holds a false belief, e.g., children observe the procedure of the unexpected-transfer task, and look reliably longer at the scene in the condition where the agent acts according to a true belief (e.g., the agent approaches the location, where the target object is actually hidden currently), compared to the condition where he acts according to his false belief (e.g., the agent approaches the location, where he previously put the target object. Infants from 13 months onwards showed attribution of beliefs in such a study design (Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007). Another research method is given by anticipatory looking tasks, in which children visually anticipate the agent's action according to his belief. This means for example, children watch a scene, in which an agent observes an object moving from one container A to another container B. Afterwards the agent disappears while both containers (covered) remain present. After a short pause, the agent reappears and looks into container B. Since the agent was able to observe the transfer of the object, he acts according to a true belief. To make sure the children show an anticipatory looking behavior the location of the children's eye gaze is analyzed shortly before the agent reappears. If children show an anticipatory looking behavior, they will fixate on container B. This procedure is repeated and serves as a familiarization phase. In the test phase, children watch a scene in which the agent is distracted and is not watching an additional transfer of the object. This time, the object does not stay at location B but moves back to location A. Here again, children's eye gaze is analyzed shortly before the agent reappears. If they consider the agent's false belief, they fixate location B. Infants at the age of 18 months showed this pattern of anticipatory looking (Thoermer, Sodian, Vuori, Perst, & Kristen, 2012). Crucial evidence that goes beyond the mere registration of false beliefs was delivered via active helping paradigms (Buttelmann, Carpenter, & Tomasello, 2009; Buttelmann, Over, Carpenter, & Tomasello, 2014). Children at the age of 18 months were able to identify a person's goal based on his belief. Described more precisely, children observed an agent in a typical procedure of the unexpected-transfer task or the unexpected-content task. They had to recognize the agent's false belief by observing his actions and had to assist accordingly by helping him to achieve his goal. To gain insight into the structure of the paradigm, I will describe the procedure of the unexpected-transfer task (Buttelmann, Carpenter, & Tomasello, 2009). Two containers A and B were present and children observed an agent who put one object into container A and then left the scene. In the absence of the agent children watched a second agent transferring the object from container A into container B and covered both containers. When the first agent returned, he went to container A but was not successful in opening this container and hesitated, and the children were encouraged to help him. If children were able to correctly identify his false belief, they opened container B. Considering the rich body of research, it is suggested that children before the age of 4 years start to gain knowledge about others' beliefs and that the ability to verbally express themselves referring to others' mental states increases with age (Wellman et al., 2001).

By applying ToM humans are able to react appropriately according to others' expectations, can adjust own actions, or are even able to manipulate others in accordance to their own interests. To execute actions one's own behavior and thoughts have to be controlled and regulated as well. These processes belong to the area of executive functions and will be discussed in the following.

Executive Functions – Inhibitory Control

There is no generally accepted definition for the term 'executive function', but the majority of literature describes executive functions as a set of higher-order self-regulatory processes, which contribute to the adjustment of own thoughts and behavior (Miller & Marcovitch, 2012; Miyake & Friedman, 2012). It involves mechanisms that coordinate the operation of cognitive sub processes, regulate the dynamics of cognition, and are located in the prefrontal cortex (Miyake, Friedman, Emerson, & Wager, 2000; Verhaeghen, 2011). Three major domains are affected by executive functions: working memory, inhibitory control, and flexibility in rule use (Miller & Marcovitch, 2012; Miyake et al., 2000; Müller, Liebermann-Finestone, Carpendale, Hammond, & Bibok, 2012). Namely, processes like memory updating, accessing recall, comprehension, resisting to interference, controlling motor responses, delaying gratification and flexibility in shifting can be captured under the umbrella term 'executive functions'. Although these processes do not provide an exact definition for executive functions, they highlight the fields where these functions work and by this display possible cognitive processes, which may serve for investigations (Zelazo, Müller, Frye, & Marcovitch, 2003).

For this dissertation the domain ‘inhibitory control’ (IC) will be of significance. IC can be described as the ability to inhibit a response to a dominant, but irrelevant, stimulus in order to focus on the less dominant, but relevant stimulus, which is important for coping with a wide range of tasks (Carlson & Moses, 2001). Deficits in IC show direct impact on motor control and affect other neuropsychological abilities like motivation and arousal, working memory, internalization of speech and reconstruction (Barkley, 1997). Even though different components of executive functions may influence each other to some extent, they can be clearly considered separate and do not contribute in the same manner to performances in complex tasks (Miyake et al., 2000). IC abilities continuously develop over the early years of childhood and display significant improvement between the ages of 3 and 6 years (Carlson & Moses, 2001; Gerstadt, Hong, & Diamond, 1994; Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996). They can be measured with tasks that require the inhibition of a dominant prepotent response in favor of a subdominant response, or with tasks that require the inhibition of a dominant prepotent response for a certain amount of time (Mischel, Shoda, & Peake, 1988). The most established designs to measure IC are the inhibition-of-conflict paradigms (conceptual conflict and spatial conflict) and the delay-of-gratification paradigm. These paradigms are used to create tasks to successfully measure early inhibitory abilities of children from 2 years onwards. An example for a spatial-conflict paradigm is the so called A-not-B task. Participants are habituated to search for an object at location A, but later have to switch and search at location B. By doing so, they have to overcome the impulse to search at the habituated location. Children by the age of 2 years start to perform correctly, although there is a clear improvement with increasing age (Espy, Kaufmann, McDiarmid, & Glisky, 1999). Therefore, it is particularly important to determine age-appropriate demands when creating novel tasks. An extended meta-analysis, for instance, indicated that the reduction of habituation trials (i.e., the amount of trials searching in location A) helps children to perform better (Marcovitch & Zelazo, 1999). An example for a conceptual-conflict paradigm is the bucket task, which requires the restriction of a dominant response, and the execution of another response which is in conflict (Carlson, Mandell, & Williams, 2004). Children are presented with two buckets at the same time, one large and one small bucket. Furthermore, they receive large and small blocks and are familiarized by sorting the large blocks into the large bucket and the small blocks into the small bucket. By this means, they build up a conceptual association. Later they are instructed to sort the blocks in a reversed manner, thus, they have to overcome and suppress the conceptual association. An example for the delay-of-gratification paradigm is the waiting task (Mischel et al., 1988). Children are presented with a tempting reward and are promised a desired second one, but only if they resist eating the reward and are able to wait for a certain duration. From 2 years onwards, children are able to follow simple task instructions and start to successfully resist the temptation by suppressing a dominant prepotent response (Carlson et al., 2004). In this type of task, setting an age appropriate waiting time is important, as well as a reward that elicits desire, since the attractiveness of the reward is crucial for performance (Golden, Montare, & Bridger, 1977).

For older preschoolers, the same paradigms are used, but vary according to the task demands. Children by the age of 3 and 4 years are able to follow more complex instructions and to handle stronger stimuli. A widely-used example of a conceptual-conflict paradigm is a Stroop-like task. Participants are shown cards, one by one, with either a day symbol or a night symbol on them (e.g., a sun or a moon). They are requested to name the opposite concept to the one they associate with the symbol shown (e.g., say “night” when the day symbol is shown and vice versa). Stroop-like tasks are manageable from 3 years onwards and performance significantly improves with age (Carlson & Moses, 2001; Gerstadt et al., 1994). An example of a more advanced spatial-conflict paradigm is the windows task. Preschoolers participate in a sticker winning game, playing against an opponent. The aim of the game is to win as many stickers as possible. A child and their opponent are sitting at a table, facing each other. The opponent behaves neutrally and does not show any emotional reactions. Children are shown two identical boxes with windows opened by an experimenter in only in their direction. Thus, the opponent is not able to see the content of the box. Children are explicitly instructed to win as many stickers as they can. One sticker is hidden by the experimenter in one of the boxes. To win, the sticker children are instructed to point at the box that the opponent should receive. Therefore, children have to resist pointing to the desired object, which appears as a dominant response. Obviously the task demands are high as they require a certain level of rule understanding and memory besides IC.

Regardless of children's age, there are significant individual differences found in IC, which show persistence during the first 4 years of life and even up until adulthood (Carlson et al., 2004; Kochanska et al., 1996; Miyake & Friedman, 2012). Other findings provide support for the correlation of IC to other competences, for example IQ, verbal fluency, internal state language and the ability to concentrate (Carlson et al., 2004; Golden et al., 1977; Mischel et al., 1988). For a more detailed report on relations between children's inhibitory abilities and diverse behavioral characteristics see Section 1.2. The capacity of regulation and control, inter alia, is also to be found in the concept of child temperament. It is directly reflected in different temperament dimensions and contributes to the individual way of experiencing and coping with things and situations of everyday life.

Temperament

Humans vary in their intensity to react to certain situations, persons or stimuli. Emotional and physical reactions to an event might be strong and intense for one child, but moderate or low for another in the exact same situation. These reactions and their regulation are set by temperament characteristics. Temperament can be defined as a construct, which includes "emotional, motor, and attentional reactivity and self-regulation" (Rothbart & Bates, 1998, p. 109), differs among individuals, and shapes social interactions as well as behavior (Rothbart & Bates, 1998). Temperament traits can be considered as biological features and in this sense count as a subset of personality traits, without including cognition-like concepts about the self and others (Rothbart, 2011). It is considered to be relatively stable over the life span, even though temperament traits are not completely constant (Bates & Pettit, 2007). Considering children's temperament as biological anchored, its impact on developmental processes is of interest in clinical and personality psychology, as well as in pedagogy. There is no universally accepted classification of temperament dimensions, so instead I will briefly summarize the five most popular accounts. A well-known approach is suggested by Goldsmith, Lemery, Aksan, and Buss (2000), who describe temperament as a basic behavioral level consisting of reaction modes, which are shaped by environmental influences and apply mainly to emotional expression and regulation. This indicates a genetic factor but extends the view to possible changes shaped by individual experiences. Buss and Plomin (1984) on the other hand, focused on characteristics which show continuity over time, and highlighted the aspect of heritability. They include the dimensions of emotionality, sociability, activity and impulsivity. Similar categories are to be found in the work of Chess and Thomas (1996) who emphasize that temperament traits should be clearly distinguished from personality traits because they refer to the specific style of behavior, not to the motivation or content of the behavior. With the help of nine categories (activity level, rhythmicity, approach or withdraw, adaptability, threshold or responsiveness, intensity of reaction, quality of mood, distractibility, attention span and persistence), children can be assigned to the constellations 'easy child', 'slow-to-warm-up child' and 'difficult child' (Chess & Thomas, 1996). Considering the positive and negative connotation, this ranking goes beyond a description of temperament traits, because it evaluates children's behavior with respect to the demands of the environment. Notably, these terms do not serve as psychopathological criteria, but rather indicate a wide range of behavior styles among normally developing children. An approach which focuses more on the level of reactivity and self-regulation is the neurobiological approach of Rothbart (2011). She provides a structure of three main dimensions with four to six subordinated scales. The first dimension ‘surgery’ contains children’s activity level, approach, high-intensity pleasure, impulsivity, shyness, and laughter. The second dimension ‘negative affectivity’ covers, for example, anger and frustration, discomfort and fear. The third dimension ‘effortful control’ refers to attention focusing and IC, to name a few. Support for this dimension formation is delivered by multi-cultural studies using the Children’s Behavior Questionnaire (Ahadi, Rothbart, & Ye, 1993; Rothbart, Ahadi, & Hershey, 1994). Finally, to combine the well-examined aspects of these aforementioned approaches, an integrative perspective is given by Zentner and Bates (2008), which reduces the dimensions to social inhibition/shyness, frustration, positive emotionality, activity level, and attention, and adds sensory sensitivity. The dimension of inhibition/shyness describes a child’s behavior when he or she meets unknown people or is confronted with unknown situations. Some children react in a rather inhibited and shy manner, whereas others are open-minded and respond without hesitation when meeting the unknown. Positive emotionality, as the second dimension, captures behaviors like positive anticipation, smiling and laughing, as well as novelty-seeking. The dimension of frustration covers the general level of irritability and frustration tolerance, which for example becomes apparent if expectations are violated. Activity level, as offers as a dimension, describes a child’s drive for movement. Some children seem to be full of energy, whereas others appear to be rather calm and relaxed. Within the dimension of attention, a child’s ability to concentrate and to stay focused, even if challenges arise, is depicted, and the dimension sensory sensitivity characterizes to what extent a child reacts to visual, auditory or tactile stimuli. Some children are very sensitive to temperature, noise or taste, whereas others are rather insensitive. Finally, Zentner (2011) argues that it remains unclear whether positive emotionality is really an independent characteristic or rather a system of associated features. Activity and social inhibition, for instance, are influenced by positive emotions. That is, a high level of positive emotion pushes the level of activity, whereas a low amount might enhance social inhibition. Based on the integrative perspective (Zentner & Bates, 2008), the Integrative Child Temperament Inventory (Zentner, 2011) was invented and will be used for the present study. It contains five dimensions excluding positive emotionality.

Since temperament characteristics shape humans' behavior, it is not surprising that temperament stands in direct relation with behavioral problems and also in indirect relation, if environmental influences promote certain development (Rothbart, 2011). A high level of anger and frustration at 10 to 11 years of age, for example, predicted externalizing and internalizing behavioral problems a few years later (Ormel et al., 2005). Early impulsivity and high activity were also identified as predictors for externalizing behavioral problems at preschool age (Hagekull, 1994), whereas children high in fear, shyness and effortful control showed lower externalizing behavioral problems (Lengua, 2003; Morris et al., 2002; Ormel et al., 2005; Rothbart & Bates, 2006; Russell, Hart, Robinson, & Olsen, 2003). However, shyness seems not to serve a general protective function against the development of behavioral problems, since a shy temperament in childhood was identified as a risk factor for anxiety problems in adolescence (Prior, Smart, Sanson, & Oberklaid, 2000). As an example for the indirect relation between temperament and behavioral problems, the variable 'parenting' is of major interest. Whereas a well-developed effortful control may protect children from the effects of poor parenting (Morris et al., 2002; Rubin, Burgess, Dwyer, & Hastings, 2003), children and their parents could also enter a coercive cycle of interaction. Mothers of distress-prone children are more likely to use aversive discipline and their children, in return, are more likely to resist their mothers' attempts. Thus, aggression increases in both interaction partners (Patterson & MacCoby, 1980). Thinking about this pattern of behavior, it is conceivable that temperament could be surely affected by environmental influences, which promote or suppress individual characteristics and could lead to behavioral problems. The classification, prevalence and impacts of behavioral problems in childhood will be discussed in the following section. The aforementioned findings demonstrate a sizeable body of research on the relation between temperamental factors and behavioral problems. A relationship, which was rarely studied, is the link between temperament and social-cognitive abilities, more precisely ToM. There is a small body of research, which suggests a connection between both constructs in the way that certain temperamental factors may influence children's participation in social interactions, which in turn could enhance or reduce the possibilities to learn about the others' minds (Suway, Degnan, Sussman, & Fox, 2012; Wellman, Lane, LaBounty, & Olson, 2011). This relationship is of major interest for the current study and will be further discussed in Section 1.2.

Behavioral Problems in Childhood – Classification, Prevalence and Possible Correlates

Any abnormality in behavior or health is associated with subjective symptoms or objective measurable signs. A collection of concurrent symptoms and signs is called a syndrome (Cullinan, 2004). The Achenbach System of Empirically Based Assessment (ASEBA) is used in clinical practice, as well as in research for diagnosing forms of maladaptive behavior and provides different syndrome scales aligned on the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition). For preschool children, different syndrome scales examine internalizing and externalizing behavioral problems. Internalizing problems include the symptoms of emotional reactivity, anxiousness and depression, somatic complaints and withdrawal (Achenbach & Rescorla, 2000). These types of syndromes tend to be introversive or intrapersonal in nature and children do not necessarily show disconformities or problems in discipline (Webber & Plotts, 2008). Due to these characteristics, internalizing problems are less striking than externalizing problems and might remain unnoticed for a longer time by caregivers or parents. Different internalizing symptoms often co-occur and do not always allow a clear distinction between different syndromes. For instance, early signs of depression could be sadness, social withdrawal, loss of appetite, headaches, stomach-aches and other physical symptoms (Webber & Plotts, 2008). Anxiety disorders also show comorbidity with depression, and measurements of the two syndromes were found to be highly correlated (Brady & Kendall, 1992). Furthermore, cognitive symptoms like irritability or the inability to concentrate may be present and, hence, overlap with externalizing behaviors (Webber & Plotts, 2008). Achenbach and Rescorla (2000) classify the symptoms of aggression and attention problems under the score of externalizing problems. Externalizing disorders demand the attention of children's environment because they are extroversive or interpersonal in nature and stand in conflict with social requirements (Webber & Plotts, 2008). Attention-deficit hyperactive disorder (ADHD), which includes symptoms of inattention, impulsivity and hyperactivity (American Psychiatric Association, 2013), or conduct problems, like aggressive and anti-social behavior, rank among them. With respect to the issue of the unclear boundaries between the individual syndromes, for the present study I will focus on summarized total scores of internalizing and externalizing behavioral problems provided by Achenbach and Rescorla (2000). If a child scores for example above the 93th percentile of the norm group on interval-scaled variables, their behavior is assigned to a borderline range, or if a child scores above the 97th percentile of the norm group, their behavior is assigned to a clinical range. For diagnostic and therapeutic practice, both cases should lead to further assessment with adequate measurements, and suitable intervention should be considered if necessary.

The necessity of investigating the correlates of behavioral problems is substantiated by high international prevalence rates among preschool children, ranging from 7% to 20% (Campbell, 1995; Egger & Angold, 2006). Even among children of 2 years of age rates of 11.8% for subclinical and clinical ranges were found, and 32% of these children lacked social-emotional competences (Briggs-Gowan, Carter, Skuban, & Horwitz, 2001). Early diagnosed behavioral problems show a relatively high stability from early childhood up to elementary school years, or into adolescence, and indicate a higher risk for later academic problems (Campbell et al., 1994; Hinshaw, 1992; McGee, Partitdge, Williams, & Silva, 1991). Over the previous decades, several variables have been investigated, which are suspected to contribute to different extents to behavioral problems. Genetic factors can only serve to explain a certain percentage. They account partially for the temporal stability and etiology of internalizing traits and anxiety disorders (Franić, Middeldorp, Dolan, Ligthart, & Boomsma, 2010; Gregory & Eley, 2007). For externalizing behavioral problems, heritability of 50% was indicated (Moffitt, 2005). Aside from the genetic component, physical and environmental factors also correlate with children's behavioral problems. As a physical factor, body weight is taken into account. Infants with a very low birth weight, for instance, later showed high levels of hyperactivity (McCormick, Gortmaker, & Sobol, 1990). As an environmental factor, parents' behavior towards their child turned out to be a correlate. Parenting and education style were found to be related to behavioral problems (Ermisch, 2008; Mash & Johnston, 1983). Special attention should also be given to cognitive abilities and their relation to behavioral problems. If one takes into account that the understanding of others' minds and the inhibition of own behavior are two major components involved in daily social interactions, it seems logical to assume that both abilities should contribute to humans' emotional states and to their behavior. More evidence for the assumption of a connection among ToM, IC and behavioral problems will be discussed in Section 1.2.

1.2 Relations Between the Constructs

This section will review recent findings about the relations between the concepts important for the work of this dissertation. To start with, the link between ToM and IC will be discussed taking previous investigations into consideration. Reported correlations and possible explanations for the connection between these two constructs will be followed by a broader view of their relationships to language abilities, behavioral problems and temperament traits. A comprehensive view of important empirical findings will be outlined and analyzed with controversial results. During the course of this, I will display current shortcomings and submit initial proposals essential for contributing to clarifications and filling existing research gaps.

The Relationship Between Theory of Mind and Inhibitory Control

The hypothesis of an existing link between children's ToM and executive functions (EF) is widely accepted, but research during recent decades is still discussing explanations and causality. Opinions about developmental precursors and similar underlying concepts are considered from different perspectives and discussed based on empirical findings in research. There are two main directions of interpreting the positive correlation between ToM[1] and EF[2], expression accounts and functional dependency accounts (Kloo, Perner, & Giritzer, 2010). Expression accounts refer to the same task demands implemented in EF tasks as the ones in and false-belief tasks (Carlson, Moses, & Hix, 1998; Russell, Mauthner, Sharpe, & Tidswell, 1991). The focus is on IC, which is required for solving false-belief tasks. Children have to suppress the dominant response (e.g., the current more salient reality) and instead have to provide a less dominant response (e.g., the less salient previous state). Support for this

[1] When reporting results of sighted studies, I refer to task batteries of ToM or to single ToM tasks (e.g., false-belief task).
[2] When using the term 'EF' in this section, I refer to task batteries of EF or to single EF tasks (e.g., IC tasks).

explanation is delivered by studies, which report a general positive correlation between IC and performance in false-belief tasks (Cole & Mitchell, 2000; Ozonoff, Pennington, & Rogers, 1991; Russell et al., 1991). However, studies that compared performances in different false-belief task designs, which varied in inhibitory demands, delivered contradictory findings. On one hand, a meta-analysis revealed that children performed better on false-belief tasks with reduced saliency of the current reality than on false-belief tasks of the standard design, indicating that inhibitory demands challenge children to pass false-belief tasks (Wellman et al., 2001). On the other hand, another study compared both false-belief task designs with respect to measurements of IC and could not confirm this assumption. Children's performances on a false-belief task with reduced saliency of reality and children's performances on a standard false-belief tasks were examined and revealed no differences in correlation to measurements of IC, which indicates that inhibitory demands cannot be responsible for the existing link (Perner, Lang, & Kloo, 2002). A meta-analysis of international studies further questioned that IC is solely responsible for the link, since all tasks out of EF batteries were positively correlated with false-belief tasks (Devine & Hughes, 2014). Samples from the United States, United Kingdom, Continental Europe, East Asia, Canada and Australia/New Zealand were included into the analyses investigating the link between composite scores of EF and false-belief batteries, as well as the link between single EF and false-belief tasks. Even though the composite scores showed the largest effects sizes for the relation between EF and false-belief understanding, all single EF tasks (i.e., IC, rule shifting, working memory and planning) showed positive links to false-belief understanding. Additionally, cross-cultural investigations of children from China and the United States showed similar links between EF and false-belief composite scores for the samples of both nationalities, but interestingly, when comparing the task performances of children from the two nations, Chinese children outperformed children from the United States on EF tasks but not on false-belief tasks (Sabbagh, Xu, Carlson, Moses, & Lee, 2006). These results further challenged whether executive demands in false-belief tasks are solely responsible for the existing link. Other authors refer instead to the general ability to reason, which incorporates both concepts (Frye, 1999; Frye, Zelazo, & Burack, 1998). They argue that false-belief task designs and EF tasks incorporate embedded rules, which are more complex in hierarchy than a simple rule (i.e., "If a, then b."). Children have to switch between two possible sets of conditions before they can apply the simple rule. In the case of solving a false-belief task, the two different perspectives (i.e., the own perspective and the perspective of the other person) have to be considered before applying the simple rule. In the case of solving the dimensional card sorting game (Zelazo, 2006), which measures flexibility in rule use, two perspectives (i.e., the dimensions of color and shape) also have to be considered, before applying the simple rule. The perspective of embedded rules could be confirmed by a study which showed stronger relationships between EF and false-belief performances when EF tasks included reasoning about rules (e.g., conceptual conflict) instead of simple inhibition (e.g., go no-go task) (Perner et al., 2002). On the other hand, this perspective can be weakened by studies, which revealed correlations between simple delay tasks and ToM tasks (e.g., Carlson et al., 2004).

The functional dependency accounts argue one step further and refer to cognitive capacities as incorporating both concepts, thus, are building a mutually dependency. Based on shared cognitive complexity and control mechanisms, both concepts might influence each other in a developmental sense. Empirical evidence supports the assumption that EF might be a prerequisite for developing ToM, because both concepts are found to be minimally connected in early years, and instead correlations start to appear from 3 years onwards. Furthermore, it was shown that early EF predicted later ToM (Carlson et al., 2004; Carlson & Moses, 2001; Hughes, 1998; Müller et al., 2012). These findings refer to children between 2 and 5 years of age and results consistently led to the same direction of prediction, although the reverse pattern could not be found. For instance, children's performances on EF tasks at 2 years of age positively predicted children's performances on ToM tasks at 3 years of age, but not vice- versa (Carlson et al., 2004). A similar pattern of prediction was found for slightly older children. Performances on EF tasks of children from 3 to 4 years of age predicted their performances on ToM tasks 13 months later. Authors suspect an effectively operating executive system could promote the emergence of mind understanding (Carlson et al., 2004) and that self-control may offer children wider possibilities to gain rich experience of others and their minds (Hughes, 1998). More support for a functional dependency is given by a training study, which showed a transfer between the card-sorting task and false-belief performances (Kloo & Perner, 2003). Three to 4-year-old children received training for the dimensional card-sorting game and showed improvement on a false-belief task, for which performance was assessed before and after the training. Likewise, a second group was trained for false-belief tasks and showed improvement on the dimensional card-sorting game. Since only one type of task was used in this experimental design for measuring ToM and EF, it would be useful to implement more training studies using tasks for different dimensions of ToM and EF.

Finally, other variables like language abilities, social interactions or temperament might also be involved in developmental interactions between ToM and EF. Hence, the relationship of the two concepts should be examined from a broader perspective.

Language Abilities – an Important Correlate of Theory of Mind and Inhibitory Control

The link between language abilities and ToM is discussed by several researchers, and investigations revealed close relationships. Sperber and Wilson (2002) suggested similarities between the pragmatics of verbal communication and mental state understanding. They point out that for a successful operation between communicators, verbal messages have to be encoded not only concerning the observable signals, but also concerning non-observable signals. This process of pragmatic interpretation contains reasoning about the sender's intentions and also a relevance-guided inferential comprehension. The authors claim that the receiver of a message attributes different levels of relevance to linguistic utterances they hear, and, therefore, pay more or less attention to them accordingly. The speaker expects the receiver to find the utterance relevant enough to pay attention to it. Both processes, reasoning about intentions, as well as presumptions about relevant parts of communication, are required in meta-cognitive abilities. More support for a link between language and ToM is given by studies on emotion words and mental state talk. Children from 2 years of age onwards start to use words referring to mental states and try to manipulate the behavior of others by using emotional language, which allows them to reach a level of intersubjectivity and ensures mutual understanding (Bretherton, Fritz, Zahn-Waxler, & Ridgeway, 1986; Bretherton, McNew, & Beeghly-Smith, 1981). For example, conversations were transcribed in which a child said that it loves her mother and wants to hold her mother, after the mother had scolded her. The child's words can be interpreted as an attempt to regain the mother's affection (Bretherton et al., 1986). Furthermore, a correlation between pretend play and the frequency of mental state talk among 4-year-old children strengthens the assumption of a positive link between language and ToM (Hughes & Dunn, 1997). Language abilities of preschool children between 3 and 4 years, especially sentence understanding and morphological rule construction, significantly predicted their performances on ToM tasks (Astington & Jenkins, 1999; Lockl, Schwarz, & Schneider, 2004). Thus, it is assumed that language provides the basis for developing ToM abilities. Predictions were not found in the opposite direction, namely that ToM would predict later language abilities. By taking into consideration that sentence structures of a language, as well as morphological characteristics, require a representation of grammatical and syntactic features, one can assume that capacities in this area might foster other representational skills as well. A training study, including a control group, indicates that sentential complements might play a major role in the development of ToM abilities (Lohmann & Tomasello, 2003). Sentential complements are parts of the complex structure of sentences. For clarification and related to the aforementioned study, a sentence contains a main clause with a mental state verb, which embeds another clause referring to this mental state verb (e.g., Suzie *thinks* the cat is in the garden. The last part of the sentence is the complement referring to the first part.). Children at 3 years of age showed an improvement on ToM performances if they had received training in the syntax of sentential complements (Lohmann & Tomasello, 2003). Aside from this, evidence for a positive effect of verbalizing the change of perspective was delivered as well. Children, who received training on general perspective-shifting discourse, improved their ToM performances. The evidence that language training and an increased use of mental state talk by mothers in early years enhance later ToM abilities in children underpins the idea of a connection between ToM and language development and the aforementioned mentioned findings (Guajardo & Watson, 2002; Hale & Tager-Flusberg, 2003; Symons, Fossum, & Collins, 2006).

Regarding the link between language abilities and IC, previous findings point to the same direction. A positive relationship between IC (e.g., delay-of-gratification tasks, conceptual conflict tasks) and language abilities was found by several studies (Carlson et al., 2004; Mischel et al., 1988; Slade & Ruffman, 2005; Wolfe & Bell, 2004). A possible explanation could be found in regulation capacities, which are needed for language processing, as well as for the inhibition of thoughts and behavior. Regulation is needed for example, for the selection process when recalling the correct word for an activated lexical concept, or when language production is affected by interfering stimuli, which have to be suppressed for producing the correct word (Green, 1998). In line with this idea, bilingual children showed better IC abilities compared to monolingual children, because more demanding inhibition processes are needed for regulating two competing languages (Poarch & van Hell, 2012). Similar findings are presented in a study about language switching, where a link between high IC and the switch costs of trilingual speech production was found (Linck, Schwieter, & Sunderman, 2012). English native speakers, who learned French and Spanish as their second and third languages received an IC task. Results showed that better IC performances were related to faster reaction times when switching into or out of the native language. Moreover, language abilities were not only identified as important correlates of ToM and IC, children's behavior also seems to be linked to their level of language skills. Associations between language impairment and behavioral problems were revealed repeatedly (Helland, Lundervold, Heimann, & Posserud, 2014; Maggio et al., 2014; McGee et al., 1991; Moffitt, 1990; Stevenson & Richman, 1978).

The Relationship Between Theory of Mind and Children's Behavior

The view of existing literature offers a picture of mixed findings concerning ToM and its possible connections to behavioral problems. Evidence for a positive relationship was delivered by a clinical study involving adults diagnosed with borderline personality disorder, a disorder, which shows characteristics of low IC and distorted mental state understanding (DSM-IV). One group of participants was treated with special focus on mental state understanding, whereas another group was treated without this focus. The group, trained with the special focus, showed greater improvement on the reduction of suicide attempts and global functioning compared to the group trained without the special focus (Bateman & Fonagy, 2008). More positive effects were delivered by studies focusing on children. Investigations of ToM at the early age of 2 years, harsh parenting and behavioral problems suggested that high ToM could provide a protective effect against negative environmental conditions (Hughes & Ensor, 2006, 2007). Children with low and medium ToM abilities who were exposed to high levels of harsh parenting, showed a high level of behavioral problems, whereas children with high ToM abilities who were exposed to high levels of harsh parenting, showed no increase in behavioral problems (Hughes & Ensor, 2006). Another positive effect of advanced ToM abilities was found for 4-year-old children. A positive connection between ToM and children's communication with friends was detected (Dunn & Cutting, 1999).

Children with good performances on ToM tasks experienced less conflict situations during peer interactions and better communication than children who performed poor on ToM tasks. These results were independent of children's language abilities, but notably only children who already were involved in firm friendships were selected for the study. Based on these results one can speculate that the quality of social interactions might be influenced by ToM abilities, but on the other hand, firm friendships existed for all children, hence, it cannot be concluded that children with lower ToM abilities are not successful in creating stable social relationships. Further findings, however, point out the influence of language abilities. Examination of a random sample of children investigated their ToM in relation to their social popularity among peers and found that although a positive link between both variables was detected, it was strongly influenced by children's language abilities (Slaughter, Dennis, & Pritchard, 2002). A similar finding was delivered by Badenes, Clemente Estevan, and Garcia Bacete (2000), who could not detect a positive link between social popularity and typical ToM tasks, but for a white-lie task containing figurative language. The aforementioned findings refer rather to the link between ToM and qualitative characteristics of social interactions, than to actual behavioral problems. Even if behavioral problems might be reflected in social interactions, the direct link between ToM and specific behavioral problems is of major interest for the purposes of this dissertation and will be considered in the following. Evidence for a negative relationship between ToM and children's behavioral problems is delivered by several studies, which revealed negative correlation between ToM and aggressive behaviors (Capage & Watson, 2001; Harvey, Fletcher, & French, 2001; Lane et al., 2013; Wellman et al., 2011). In contrast, other studies could not confirm this negative correlation, and instead found positive correlations. Performances on ToM were positively linked to aggressiveness or high sensitivity to criticism (Dunn, 1995; Renouf et al., 2010; Walker, 2005). Notably, there were also studies which could not find correlations between

ToM and behavioral problems at all (Hughes, White, Sharpen, & Dunn, 2000; Slaughter et al., 2002; Yiwen, Chongde, & Wenxin, 2004).

On first view these seemingly contradictory findings appear puzzling. However, further detailed analysis of the study designs reveals that there are a number possible explanations for the different findings on the relationship between ToM and behavioral problems. Differences in the specific tasks, age groups, and control measures can be identified, which might explain the different outcomes. Most studies that revealed negative links between ToM and behavioral problems investigated children's ToM from 3 years of age onwards (Capage & Watson, 2001; Lane et al., 2013; Wellman et al., 2011). For younger age groups, results varied and studies with children below the age of 3 years are rare. In one of these investigations, a well-developed ToM by the age of 2 years prevented children from developing behavioral problems by the age of 4 years despite the influence of harsh parenting (Hughes & Ensor, 2007). However, the majority of previous studies focused mainly on false-belief understanding, which is only one dimension of ToM and may not provide a comprehensive profile of children's understanding of others' mental states (Tomasello et al., 2005; Wellman & Liu, 2004). Furthermore, as discussed in Section 1.2, language is considered an important correlate of ToM and of behavioral problems, and consequently, this variable should be controlled for. However, not all of the aforementioned studies assessed children's language abilities (Hughes et al., 2000; Suway et al., 2012; Walker, 2005).

The Relationship Between Inhibitory Control and Children's Behavior

For a successful cooperation with others, children have to make use of their executive functions, which includes the conscious control of thoughts and behavior (Miller & Marcovitch, 2012). Especially when learning how to get along with others in the sense of commitments, consideration, and courtesy, IC abilities can affect whether an individual functions well or badly in socialization processes (Kochanska et al., 1996). One could assume that for children with higher IC their popularity in a group could increase and simultaneously their problematic behavior decrease. Investigations on the relationship between IC[3] and behavioral problems are fairly consistent. A large number of studies indicate the advantages of a high IC. Negative correlations between IC and behavioral problems (e.g., angry and antisocial behavior, attention-deficit/hyperactivity disorder, aggressive and delinquent behavior) were found (Berlin, Bohlin, & Rydell, 2004, Espy, Sheffield, Wiebe, Clark, & Moehr, 2011; Hughes & Ensor, 2008; Hughes et al., 2000; Lewis, Dozier, Ackerman, & Sepulveda-Kozakowski, 2007; Morris, Keane, Calkins, Shanahan, & O'Brien, 2014; Olson, Schilling, & Bates, 1999; Oosterlaan & Sergeant, 1996).

Longitudinal studies also showed the positive effects of a well-developed IC among older preschoolers on various competences in adolescence, particularly emotion regulation, social competences, academic success, verbal fluency, and success in coping with frustration (Carlson & Wang, 2007; Mischel et al., 1988). IC was also considered as a possible mediator. Children with high regulation skills showed less externalizing behavioral problems and more social competence compared to children with lower regulation skills. Interestingly, children's regulation skills mediated mothers' positive and negative emotional expressions on children's externalizing behavioral problems and social competences (Eisenberg et al., 2001). IC also mediated the relationship between language abilities at 2 years of age and behavioral problems at 4 years of age (Hughes & Ensor, 2008). IC as a correlate to behavioral problems should receive particular attention, because individual differences in IC are common and mostly persist during the developmental years (Carlson et al., 2004; Kochanska et al., 1996; Miyake & Friedman, 2012). It is therefore possible that differences in IC could have long-term effects.

[3] When reporting results of cited studies, I refer to task batteries of IC or task batteries of executive functions, which include IC tasks.

The Relationship Between Theory of Mind, Inhibitory Control and Children's Temperament

The correlation between ToM, IC and temperament dimensions is rarely examined. A lack of aggressiveness and a shy-withdrawn attitude to social interactions, as well as low levels of activity were found to be related with better false-belief abilities (Lane et al., 2013; Wellman et al., 2011). The emotional-reactivity hypothesis (Hare & Tomasello, 2005) could supply the first explanation for these findings by taking into account self-domestication in human phylogeny. This approach refers to the formation of social groups, which is the base for the development of cooperative cognition. Individuals who are low in reactivity and aggressiveness might have been more accepted by a group and fostered their social cognition by participating in social interactions. In contrast, individuals who are high in reactivity and aggressiveness might have been less accepted by a group, were rejected from social interactions, and limited in developing their social competences. It seems reasonable that certain temperament traits promote participation in social cooperation. This, in return might causally influences the development of ToM. Which temperament dimensions precisely would be involved in this process has not been clarified. Support for a link between temperament traits and later ToM comes from longitudinal studies. A shy temperament at 18 months and 3 years of age was positively related to children's ToM scores by the age of 3 years (Mink, Henning, & Aschersleben, 2014). This result could favor the assumption that temperament is involved in ToM development. Authors suggest that a shy-withdrawn and observant stance towards others could lead to a better understanding of interpersonal processes (Mink et al., 2014). However, investigations are rare and the logic of this argumentation can easily be reversed by suggesting that shy-withdrawn personalities participate less in social groups, and thereby gain little experience in interaction, negotiation and in taking others into consideration. Suway et al. (2012), for instance, found a negative connection between behavior inhibition (e.g., latency to approach novel stimuli, proximity to mother in novel situations) and ToM. Three-year-old children displayed lower ToM scores if they showed high levels of behavioral inhibition combined with mainly negative peer interactions at the age of two. No such connection was found for children with few negative peer interactions, suggesting that temperament traits might enhance social interactions, which then equip children with possibilities to learn about the mental states of others. Furthermore, a growing body of research indicates a higher risk of difficulties in social contexts for children with a particularly shy temperament (Coplan & Armer, 2007).

Studies on the link between IC and temperament reported positive correlations between performances on IC tasks and the temperament dimensions of focusing and shifting attention (Gerardi-Caulton, 2000; Wolfe & Bell, 2004). These results appear reasonable, because self-regulatory processes like attention focusing are typically demanded in IC tasks. A finding that goes beyond the related features of both concepts is given by a study investigating adults. A difficult temperament was connected to aggressive and antisocial behavior, and this link was mediated through EF, including IC components (Giancola, Mezzich, & Tarter, 1998). Considering the relatively low number of empirical investigations, the necessity to include the assessment of temperament as a possible correlating factor on ToM or IC becomes apparent.

2 Dissertation Project

The major goal of this dissertation is to identify the link between ToM, IC, and behavioral problems in preschoolers. Studies on early behavioral problems and their relation to social-cognitive competences and inhibitory abilities are underrepresented in research and show a narrow range of investigated mental states, informants and age groups. The relatively high rate on indicated behavioral problems at the age of 2 years (see Section 1.1.) emphasizes the demand for reliable information about early stages of life. Importantly, I want to clarify if a connection is present from very early on or if it gains importance with increasing age. With respect to previous research, I assume a negative relationship between IC and behavioral problems. Due to the ambiguous findings on the relationship between ToM and behavioral problems, the purpose of this study is to clarify the direction of correlation. Since language was an important correlate of all three variables (ToM, IC and behavioral problems, see Section 1.2.), it has to be included as a control variable in this theoretical model (Figure 1). Furthermore, I want to consider temperament. The relationship between temperament and behavioral problems was investigated repeatedly (Caspi, Henry, McGee, Moffitt, & Silva, 1995; Prior et al., 2000; Rothbart & Bates, 1998), therefore focus will be on the link between ToM, IC and temperament in early childhood.

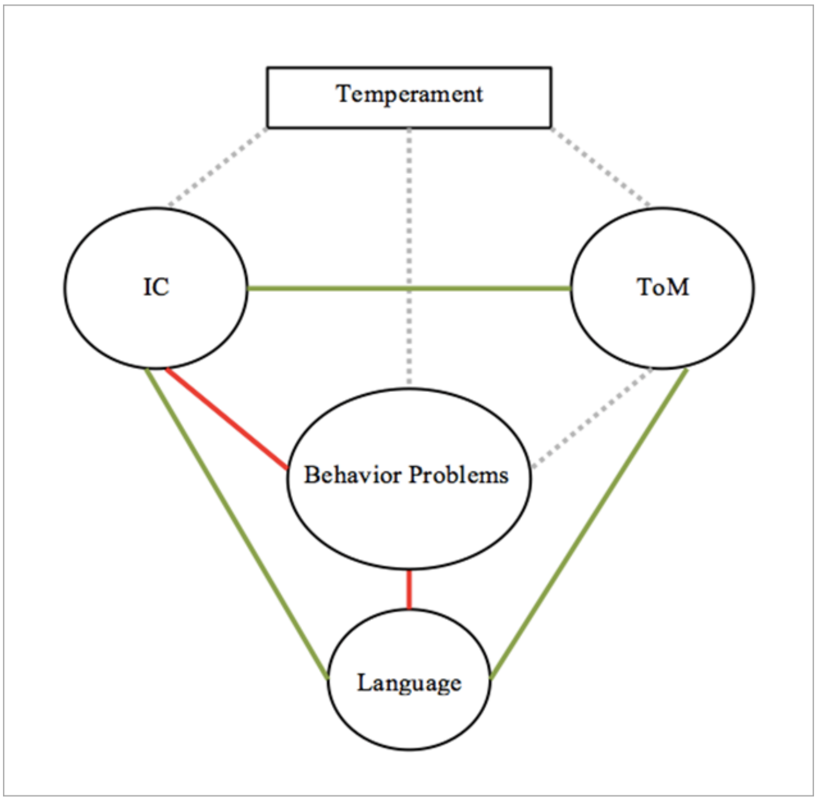


Figure 1. Theoretical model of suspected correlations between ToM, IC, behavioral problems, language and temperament (green line: positive correlation, red line: negative correlation, grey dotted line: positive or negative correlation).

The suspected direction of correlations between IC, ToM and temperament varies depending on the dimension of temperament. With respect to previous findings (Olson et al., 1999; Oosterlaan & Sergeant, 1996) and the similarity between temperament dimension characteristics and symptoms of behavioral problems, I assume a positive relationship between IC and social inhibition, as well as a positive relationship between IC and attention, whereas I assume a negative relationship between IC and activity. Little is known about the relationship between ToM and temperament. Controversial approaches suggest, either that high inhibition and shyness could enhance ToM, because children might learn about others' mental states by watching and observing others (Wellman et al., 2011); or that a less inhibited behavior could enhance children's social interactions and by this they might learn about others' mental states (Suway et al., 2012). To clarify the direction of correlation, the link between ToM and social inhibition will be of particular interest. Since attention focusing is a necessary feature for observing the environment and is helpful for a sensitive perception of others, high attention could be also an advantage for understanding others' minds and a positive correlation to ToM is suspected. By contrast, a high activity level could be assumed to distract a child from observing others precisely, and therefore, a negative correlation to ToM is suspected.

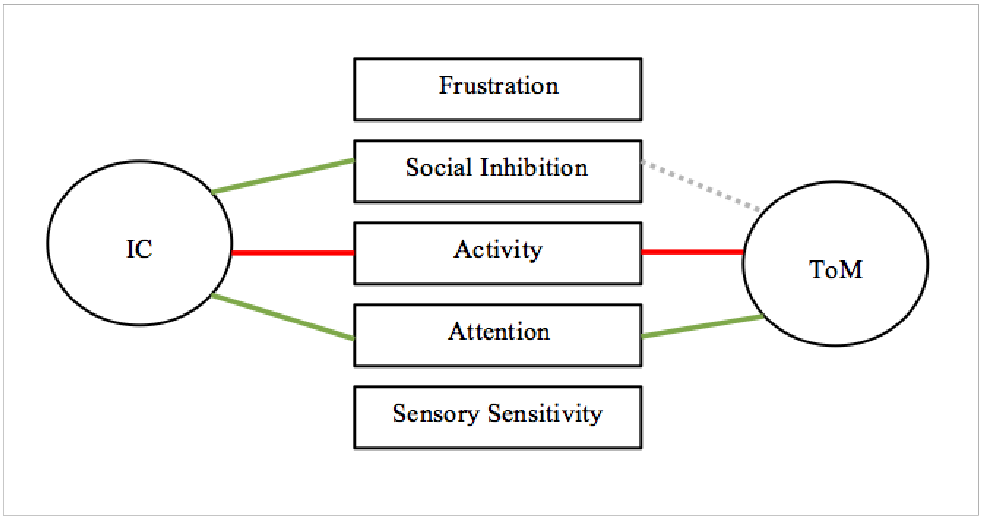


Figure 2. Theoretical model of suspected correlations between ToM, IC and temperament dimensions (green line: positive correlation, red line: negative correlation, grey dotted line: positive or negative correlation).

In this dissertation project, three age groups will be included: Children of 2, 3 and 4 years of age respectively. They will be tested with comparable measurements to gain insight into different developmental stages. To create a complex profile of abilities, it is necessary to include different dimensions of ToM and IC assessed by age-appropriate tasks. Previous studies focused mainly on false-belief understanding and operated on traditional task designs, which require a certain level of language comprehension and production, because children have to explicitly answer test questions (e.g., 'What does Person A think?' or 'What does Person B believe?'). Thereby, children have to understand the meaning behind the mental state terms and to which mental constructs they refer. The explicit understanding of mental states is usually tested from 4 years of age onwards, where an extensive mother tongue vocabulary is developed and used for complex conversations (Menyuk, Liebergott, & Schultz, 1995). However, this method is not suitable for younger children. Research provides new paradigms to examine ToM already before a child's second birthday, by creating nonverbal designs, which measure either gaze behavior or children's active helping (Buttelmann, Carpenter, & Tomasello, 2009; Buttelmann et al., 2014; Kovacs, Teglas, & Endress, 2010; Repacholi & Gopnik, 1997). IC can also be assessed from very early on, for instance with age-appropriate waiting tasks or simple Stroop task designs (Carlson et al., 2004; Golden et al., 1977).

To multidimensionally investigate ToM and IC, extensive batteries for 2-year-old, 3-year-old and 4-year-old children were created. The ToM batteries included tasks to measure goal understanding, preference understanding, intention understanding, and false-belief understanding, comparably for all age groups. The IC batteries included tasks to measure the inhibition of a conflicting concept, the inhibition of a spatial conflict, and the delay of gratification, comparably for all age groups. A major advantage of the current study lies in the use of implicit task designs: Instead of verbally answering test questions, 2-year-old children could solve the tasks completely nonverbally by imitating, assisting or helping the experimenter. The 3-year-old children could solve the majority of tasks nonverbally as well, as only two ToM tasks and one of the IC tasks required simple language production (e.g., naming colors or symbols, and reporting the predicted actions of someone else regarding their beliefs). Furthermore, to strengthen reliability of measurements, test-retest reliability was assessed for the separate measures. To put children's extensive profiles in relation with behavioral and temperamental factors, parents completed the Child Behavior Checklist 1 ½ - 5 (CBCL, Achenbach & Rescorla, 2000) and The Integrative Child Temperament Inventory (IKT, Zentner, 2011). To gain multi-informant data about children's behavior in other contexts, caregivers filled out the Caregiver-Teacher Report Form 1 ½ - 5 (C-TRF, Achenbach & Rescorla, 2000). To include language as a potential correlate of social-cognitive abilities, receptive language abilities of all children were measured with subtests of the SETK 2 (Grimm, Aktas, & Frevert, 2000) and SETK 3 – 5 (Grimm, 2001). This study extends the existing literature on the relationship between ToM, IC, behavioral problems and temperamental factors with a unique approach, investigating children with extensive batteries consisting of tasks appropriate for three developmental stages in preschool age.

2.1 Method

Participants

In total, 252 2-, 3- and 4-year-old children were investigated for this study. The sample of 2-year-olds consisted of 82 children (mean age = 24.0 months, SD = 0.5; range: 23 months and 2 days to 24 months and 25 days; 41 girls). Information letters were sent to families with at least one child. Subsequently, children were registered by their parents for participation in studies on child development and were invited via telephone. Testing took place in a university laboratory in a mid-sized German city and consisted of two sessions, each lasting 30 minutes with a 10 minute warm-up phase at the beginning. The mean interval between both sessions was 6.95 days. The sample of 3-year-olds consisted of 90 children (mean age = 36.1 months, SD = 0.57; range: 35 months and 1 day to 37 months; 47 girls). Children in this sample were recruited via lists in daycare centers and via telephone, because children had been registered by their parents for participation in studies on child development earlier. Testing took place in a university laboratory in a mid-sized German city or in daycare centers and consisted of two sessions, each lasting 40 minutes with a 10 minute warm-up phase at the beginning. The mean time interval of the testing sessions was 7.6 days. The sample of 4-year-olds consisted of 80 children (mean age = 48.0 months, SD = 0.6; range: 46 months and 29 days to 49 months and 5 days; 44 girls). Children in this sample were recruited via lists in day-care centers. Testing took place in day-care centers of a mid-sized German city and consisted of two sessions, each lasting 50 minutes with a 10 minute warm-up phase at the beginning. The mean time interval of the testing sessions was 8.0 days.

One additional child (3-year-old) participated in the study but had to be excluded from data analyses because of developmental delay. Another eight children (four 2-year-olds, two 3-year-olds, two 4-year-olds) attended the first testing session but were absent from the second testing session. A varying minor number of children refused to participate in some of the single tasks within the batteries.

Procedure

Testing was split into two sessions with a delay of approximately seven days in order to reduce the duration of testing. The ToM session included the assessment of ToM tasks whereas the IC session focused on IC tasks. The order of sessions was counterbalanced between participants (i.e., half of the participants received the ToM session first and the IC session second, whereas the other half of children received the sessions in the opposite order). At the end of the first session, parents and caregivers were given the questionnaires, accompanied by the instruction to complete and return them at the second session. To investigate test-retest reliability, every child received the first task of the first session again at the beginning of the second session. For the assessment of 2-year-old children, one trained experimenter was involved in the data collection. For the assessment of 3-year-old children, three trained experimenters were involved in the data collection. For the assessment of 4-year-old children, seven trained experimenters were involved in data collection. Furthermore, for all age groups, an assistant was involved in data collection.

Measures

Theory of Mind Batteries

The ToM batteries for each age group contained five established and partially modified tasks, examining four dimensions of mental states: the understanding of others' goals, intentions, preferences and false beliefs. False-belief understanding was assessed by two experiments, an unexpected-content task and an unexpected-transfer task. Tasks for the three age groups are comparable concerning the mental states investigated and rise in complexity with increasing age.

Goal understanding – 2-year-old children. To investigate 2-year-olds' goal understanding, an imitation task was modified (Carpenter et al., 2005). The experimenter sat in front of the children, facing them, at a table. The task consisted of two conditions, a *house condition*, where a physical final location was present, and a *no-house condition*, in which an imaginary final location was present. In the house condition, the experimenter placed two small houses (16 cm x 9 cm x 11.5 cm) on the left and right side of the middle of the table. The houses were placed within the children's and experimenter's reach approximately 20 cm away from each person's side of the table. Within each condition, the experimenter demonstrated four actions directed at one of the final locations. Actions were demonstrated by holding a small toy animal (i.e., mouse or sheep) in the right hand and by moving it from the center of the table to one of the houses by a specific means (i.e., hopping or flying). The experimenter's action was directed two times at the right house and two times at the left house in an alternating order. The no-house condition resembled the house condition with the crucial difference that no houses were present and all means were performed towards imaginary spots on the table. Every action was accompanied by a sound ("dumdumdum" for hopping, "duuuwummmm" for flying). Before the experimenter demonstrated the action, she directed the children's attention to her action ("[Name of the child], look!"). When the experimenter finished the action, she smiled, gave the toy to the children, and said "Now it's your turn!" Children's behavior was scored as correct if they put the toy animal in the demonstrated final location. Consequently, children could receive one point for each trial they participated in. As they were presented with four trials in each of the conditions, they could receive a maximum score of 8 points. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have reached. The order of conditions was counterbalanced across children. Half of the children received the house condition first, followed by the no-house condition. The other half of the children received the conditions the other way around. The means and the locations within the conditions were counterbalanced across children.

Goal understanding – 3- and 4-year-old children. I adopted an imitation task (Bekkering et al., 2000) to investigate 3- and 4-year-olds' understanding of others' goals. The experimenter sat in front of the children, facing them, at a table. As young children naturally imitate others' gestures by mirroring them (Schofield, 1976), they were instructed "Do as I do!". The task consisted of two conditions (*dot condition*, *ear condition*), which included a physical final location, and one condition (*no-dot condition*), with an imaginary final location. In the dot condition, the experimenter placed two blue dots in front of the children and two identical blue dots in front of herself. The dots were located approximately 20 cm away from the edge of the table within reaching distance. Within each condition, the experimenter demonstrated four uni-manual and two bi-manual hand gestures, i.e. she used either one hand (e.g., right or left) or both hands to touch the dots. When she used only one hand, the movements were either ipsilateral by moving the hand straight forward to the dots, or contralateral by moving the hand diagonally across the body midline to the dots. When she used both hands, the movements were either ipsilateral by moving the hands straightforward to the dots, or contralateral by moving the hands diagonally across the body midline to the dots. The ear condition differed from the dot condition only with regard to the final location. In the ear condition, no dots were present and the experimenter touched her ears instead. The no-dot condition differed from the dot condition also with regard to the final location. In the no-dot condition, an imaginary final location was present and all hand movements were directed at imaginary end points left and right on the table. Six hand movements (un-imanual-ipsilateral right, uni-manual-ipsilteral left, bi-manual-ipsilateral, uni-manual-contralateral right, uni-manual-contralateral left, and bi-manual-contralateral) were shown within each condition (for a total of 18 movements). Children's responses were scored as correct if they imitated the demonstrated hand movement towards the correct final location. With six trials in each of the three conditions (i.e., *dot, ear, no-dot*) children were able to reach a maximum of 18 points. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have reached. The order of conditions and the order of hand movements within each condition were counterbalanced across children.

Intention understanding – 2-year-old children. To evaluate children's understanding of others' intentions as rational choices of action plans, I used the established task of Buttelmann et al. (2008). The experimenter and the children sat at a right angle at the table. An assistant (sitting opposite the children) started a familiarization phase by putting a small bucket on the table, calling the experimenter's name to get her attention, and put a little toy figure into the bucket. The experimenter looked into the bucket and took the toy figure out of the bucket. She smiled afterwards and said "Oh nice!". This familiarization trial was presented three times in a row with different toys. Then the assistant proceeded with a fourth demonstration where the experimenter called the children's name, and gave the bucket to the children. Children were then allowed to take the toy out. Then the demonstration phase began. The assistant removed the bucket and put a seesaw on the table. The slide was directed towards the children and the experimenter. The slide of the seesaw was blocked from below by a pink styrofoam cube. The assistant placed a toy figure on the top of the slide and called the experimenter's name ("[*Name*], look!"). The experimenter explored the seesaw, bent over and looked left and right at each side of the seesaw. Her actions were accompanied by mumble ("Hm a ha."). After inspecting the seesaw, she pushed the pink styrofoam cube to the side with her right hand, pushed the slide down and let the reward roll down onto the table. She picked it up and expressed delight ("Oh nice!"). This procedure was repeated another two times. Then, the test trials for the children started. The assistant repeated the same procedure, looked at the children and called their names while putting the toy figure on the seesaw ("[Name of the child] look!"). Subsequently, children were allowed to have the toy figure.

Two more test trials followed in the same manner. Then three more demonstration trials were directed at the experimenter again, before another three test trials for the children followed. Consequently, children received a maximum of six trials. Children's behavior was scored as correct or incorrect, based on whether they successfully or unsuccessfully used the tool before grasping the reward. If children simply took the reward, they received no point. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have reached. The order of toy figures was counterbalanced across children.

Intention understanding – 3-year-old children. To investigate 3-year-olds' understanding of others' intentions I slightly modified a task by Schult (2002). An intention can be described as a plan of action to achieve a goal, thus it consists of both a goal and an action plan (Tomasello et al., 2005). The basic idea was to investigate, whether children are able to understand that intentions need to be fulfilled by the means that were planned. The experimenter and the children were instructed by the assistant to build a team for playing a game. Two differently colored boxes (blue and red) were put on a table approximately 2 m away from the experimenter and the children. The assistant asked the children to name the colors of the boxes to check whether they were able to identify the correct color. If the children were not able to identify the correct color, the assistant explained how to label the boxes. Then she explained the aim of the game: to collect six stamped images of an animal (i.e., zebra or elephant) to complete a graphic on a sheet of paper. She then explained the rules of the game: In each box, one card was hidden by the assistant. In one box, she hid a card with the animal symbol (symbol card) and in the other box, she hid a blank card. To receive a stamped image for the graphic, the experimenter and the children had to find the symbol card. To find the symbol card, the experimenter had to toss a ball into the box children suspected to hold the symbol card. For every trial the assistant asked the children "I hid the symbol card in one of the boxes. It's now either in here (pointing at the red box) or in here (pointing at the blue box). Into which of the boxes shall [name of experimenter] now toss the ball?" While asking, she held a red card above the red box and a blue card above the blue box. After the children made their choice, she put the card on the floor in front of the box with the same color. At this moment an intention was illustrated by creating the action plan to receive the prize. The experimenter threw the ball and acted like she would try to hit one of the two boxes. The assistant removed the card from the box that was hit. If the experimenter hit the box holding the symbol card, children received a stamped image, but if she hit the box with the blank card, children received no stamped image. The assistant then asked the test question "Which box should [name of experimenter] have hit?". The experimenter's intention was only fulfilled, if she hit the box that children told her before. The intention was unfulfilled, if the experimenter missed the intended box and hit the other one instead. However, the reward could have been obtained in both scenarios. A match condition was either created when the means was fulfilled (the experimenter hit the intended box) and the reward was retrieved, or when the means was unfulfilled (the experimenter missed the intended box) and the reward was missed. A mismatch condition was either created when the means was fulfilled but the reward was missed, or when the means was unfulfilled but the reward was retrieved. Six trials were performed in each condition. Children's behavior was scored as correct if they answered the test question with naming the intended box (e.g., "She wanted to hit the red box."). Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have achieved. To compare the performance between conditions, the total performance score was additionally divided into a match-score and a mismatch-score.

Intention understanding – 4-year-old children. To test 4-year-old children's understanding of intentions, another task of Schult (2002) was modified. The experimenter sat in front of the children facing each other at a table. Two sets of picture stories ('Tom' and 'Maria'), each consisting of three different picture stories, were used (see Appendix A for illustrations). In each of the picture stories, the protagonist had the same intention, i.e. a plan of action to achieve a goal. What varied between stories was the means that were used to reach the goal. In story type A, the end state was achieved by the protagonist himself but with a different means than planned before (e.g., Tom had the plan to jump into a puddle of mud to achieve his goal of having a dirty pair of trousers. While drinking chocolate he poured it himself and, by this, achieved the end state of having a dirty pair of trousers). In story type B, the end state was achieved by the same means but by another person (e.g., another child jumped into the puddle of mud, splattered Tom, and achieved the end state of having a dirty pair of trousers). Finally, in story type C, the end state was achieved by the protagonist using means very similar to the one planned, but these means happened accidentally (e.g. Tom fell into the puddle of mud by accident and achieved the end state of having a dirty pair of trousers). Consequently, although the protagonist always achieved his goal, the protagonist's intention was always unfulfilled because the action plan was not realized. Having two sets allowed the experimenter to present children with two trials of each story type. At the end of each story, the experimenter asked two test questions. As the first, an open question, she asked children what the protagonist's idea was. If the answer included cues to both the means (e.g., 'jump') and the end state (e.g., 'dirty pants') children were scored as correct. The second question was a forced-choice between two pictures, one showing the means the protagonist had planned to perform and the second showing the means that was actually used. The experimenter asked the children whether the protagonist wanted to achieve his goal by doing A or B. Children's responses were scored correct if they pointed at or described the means the protagonist had planned to perform. Scores of mean performance in percentages of trials were calculated separately for story types A, B and C. Additionally, each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have reached. The order of the story sets and the story types were counterbalanced across children.

Preference understanding – 2-year-old children. To test children's understanding of others' preferences I modified a task by Repacholi and Gopnik (1997). During a warm-up period before the test session started, the assistant conducted a pre-test to identify children's preferences. She presented six pairs of objects on a tray, one pair after another. Each pair of objects consisted of one colorful object (e.g., a toy figure) and one colorless object (e.g., a piece of cardboard or a plain empty paper box). On each presentation, children were asked to choose one of the objects. During the test phase, the experimenter and children sat at a table, opposite each other. The assistant placed the tray with two objects (the same pairs as in the pre-test) in front of the experimenter, who then expressed her preference or rejection for each object. She picked up one of the objects and demonstrated delight ("Oh aha!", accompanied by raising her eyebrows and smiling) and put the object down. Then she took the other object, showed expressions of dislike ("Egh uh.", accompanied by pulling a face and wrinkling her forehead) and put the object down again. Subsequently, she turned away and pretended to write a text. Meanwhile, the assistant put the tray closer to the children and out of the experimenter's reach. The experimenter turned back, bent over, and extended her right hand towards the center of the tray with the two objects. Since she was unsuccessful, she asked the children "[Name of the child], give me one!" Two conditions were included. In the match condition, the experimenter preferred the same toy that the children had chosen in the pre-test. In the mismatch condition, the experimenter preferred the object that the children had not chosen in the pre-test. Three trials were performed in each condition, consequently, children received six trials in total. Each child received a total percentage score, which was created derived by dividing the number of correct responses by the maximum number of correct responses that children could have achieved. To compare performance between conditions, the total performance score was additionally divided into a match-score and a mismatch-score. The order of the expressions and the order of pairs of objects were counterbalanced across children.

Preference understanding – 3-year-old children. This task was based on the task of Buttelmann, Call, and Tomasello (2009), which was originally carried out with great apes as test subjects. The experimenter sat opposite the children at a table. Then she turned away pretending to be busy. A tray with two opaque cups, turned up-side down, was placed on her side of the table. The assistant explained that the aim of the game was to find Haribos, a special kind of sweets. She showed the children two Haribos (differing in color and shape), placed an occluder on the table to block children's view, and put one Haribo under each cup. Her actions were accompanied by saying "I put one Haribo under one cup and the other Haribo under the other cup. There is now one Haribo under each cup." After this hiding procedure, she removed the occluder. The experimenter turned around and looked under one of the cups in such a way that the children were not able to see its contents. She expressed delight ("Oh aha", accompanied by raising her eyebrows and smiling). Afterwards, she repeated the procedure with the other cup and expressed disgust ("Egh uh." accompanied by pulling a face and wrinkling her forehead). She put the occluder in front of the cups, bent over and removed the Haribo she desired accompanied by chewing sounds. Children were not able to see which Haribo was removed, but saw and heard her chewing. She removed the occluder and said "I ate one Haribo, but one Haribo is still there." She then put the tray with the two cups closer to the children and asked "Where is the one that is left?". At this point, children had to infer that the experimenter had eaten the reward she liked, and consequently, had to go for the cup holding the reward the experimenter did not like. A total of eight trials was performed. Children's behavior was scored as correct if they chose the cup still holding a reward. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have achieved. The order of the expressions was counterbalanced across children.

Preference understanding – 4-year-old children. To investigate if 4-year-old children understand that individuals act according to their individual preferences, a task of Buttelmann et al. (2009) was modified. The experimenter sat in front of the children at a table, facing them. The experimenter presented children with six picture stories (see Appendix B for illustrations). All stories introduced two protagonists, Teddy and Paul. In each story, the protagonists found a pair of objects, with different objects included in each of the stories. The six pairs of objects were a ball and a toy car; a flower and a mug; pencils and a book; a toy mouse and a helicopter; a duck and a mouse; and a shoe and a cap. In every story, Teddy and Paul liked the same object and disliked the other one. After revealing these preferences, they placed a blanket over the objects so that the objects were no longer visible. Teddy then left the scene. Paul then took one of the objects without the children being able to see which object he took. Consequently, only one object remained under the blanket. At the end of the story, Teddy returned and the children were asked which object was left under the blanket for Teddy. Children's responses were scored correct if they inferred that Paul acted according to his preference and, thus, took the preferred object, and so it was the disliked object, which was left for Teddy. After each trial children were asked which object they would have taken. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have achieved. The order of the stories and the objects liked by the protagonists were counterbalanced across children.

False-belief understanding – 2-year-old children. To investigate children's false-belief understanding an unexpected-content and an unexpected-transfer task were administered. For the unexpected-content task, I used an interactive helping paradigm created by Buttelmann et al. (2014). The experimenter sat down onto a cushion at the right side of the children, the assistant sat down onto a cushion on the left side of the children. Four identical looking cardboard boxes with pictures of colorful toy blocks at the front were placed in a cabinet opposite to the children. The experimenter mentioned that she would like to play with blocks. The assistant asked if she should go and get some. The experimenter showed appreciation

("Oh yes!") so that the assistant went to the cabinet, took one of the boxes and sat down again. The assistant explained that this is a toy block box whilst she pointed at the pictures, then opened the box, revealed the block, and handed it over to the experimenter. The experimenter took the block and placed it in front of her. Afterwards, she said she wanted more. This procedure was repeated two times, so that the experimenter built a tower of three blocks. After the third demonstration, the experimenter said she had to leave for a while and left the room. In her absence, the assistant went and took the last box. She opened the box and found a spoon inside. She showed surprise ("Oh there is no block inside! Look, there is a spoon inside! But see, there are blocks on the box, this is a block box, but there is a spoon inside. This is strange."). After both, the assistant and the children, examined the unexpected content, the assistant closed the box and put it back on the shelf in the cabinet. As soon as she sat down again, the experimenter entered the room. She brought a small bowl and mentioned that she found this bowl outside. She sat down, looked at the blocks, looked at the bowl, and touched both at the same time. She then looked up at the box on the shelf, performed a reaching gesture and called for the assistant's attention ("I want this!"). The assistant lifted a large piece of cardboard, which laid in front of the cabinet, and revealed a spoon and a block. She said to the children "Look [name of the child], I have it here as well", sat down again, and asked the children to go and get the experimenter what she wanted ("[Name of the child] go and get it for [name of experimenter]"). Children's behavior was scored as correct if they gave the block. The object children touched first was coded as well. The positions of the objects placed under the piece of cardboard were counterbalanced across children.

For the unexpected-transfer task, I used another interactive paradigm (Buttelmann, Carpenter, & Tomasello, 2009). Children sat on a cushion next to the assistant, and the parents sat behind their children. Two wooden boxes (a pink box and a yellow box) were placed in the middle of the room approximately 1 m away from each other and 2 m away from the children. The boxes were equipped with lids, which could be locked with pegs. The boxes were unlocked when the experimenter approached. The pegs were placed in front of the two boxes. The experimenter approached the boxes, one after another, and lifted their lids. She demonstrated delight, smiled and said “Oh nice!”, then left the room. In her absence, the assistant told the children to explore one of the boxes together. She opened the lid, put a toy into the box, closed the lid, and locked it with the peg. She shook the lid a little to demonstrate that the box was locked. Afterwards, she pulled out the peg to unlock the lid, opened the box, and removed the toy from the box. She described each step of action. After this demonstration, she put the toy back into the box and asked the children to unlock the box alone. This practice trial was repeated until the children learned how to unlock the boxes on their own. They had to successfully open the box at least two times. If children still required help on the third try, the practice procedure was ended and the assistant continued with the task. The same practice trials were repeated for the other box. After finishing the practice trials, the assistant locked both boxes and - together with the children - returned to the cushions. As soon as they sat down again, the experimenter returned and showed her favorite toy to them. She then tried to open one of the boxes to put her toy in. Since she had been absent while the functions of the boxes had been explained, she was not able to open the locked box and hesitated. The assistant requested the children to help the experimenter. While children unlocked the box, the experimenter turned away, thus, did not pay attention. If children were not able to open the box on their own, the assistant helped. The experimenter then put her toy into the opened box, closed the lid and left the room again. In her absence, the assistant removed the toy from the box and put it into the other box. During the transfer of the toy she giggled in a sneaky way, commented every action and mentioned that the experimenter was outside and could not see what they do. After she finished the hiding process, she returned to her cushion. As soon as she sat down again, the experimenter returned, went straight to the box into which she had initially placed her toy, and tried to open the lid (with three short pulls). Again, she was not able to open the box. In resignation, she sat down in the center between both boxes. The assistant encouraged the children to help the experimenter ("Come on [name of the child], go and help [name of experimenter]."). If the children were not assisting after the second request, the experimenter asked the children to help her. If the children still hesitated, the assistant offered to join while helping the experimenter. If the children still refused to help, the parent was instructed to request the children to go and help the experimenter. If the children still hesitated, the assistant allowed the parent to go and help together with their children. Parents were instructed to let children take the lead and decide on their own what to do. Children's behavior was scored as correct if they chose the box that now contained the toy. The box children touched first was coded as well. The location of the yellow box and the box in which the experimenter hid her toy were counterbalanced across children.

False-belief understanding – 3-year-old children. For the unexpected-content task, I modified a task from Hansen (2010). The experimenter sat in front of the children at a table, facing them. She introduced a small hand puppet, a white duck ("Oh, see who is here! It's the little duck. Isn't she cute?"). Then, the experimenter put a box of chocolate ('Kinderschokolade', a well-known brand in Germany) on the table. The duck looked at the box and expressed delight ("Oh…ah!"). Subsequently, the experimenter explained "The little duck is very busy and has to go right now" and put the puppet under the table. After the duck disappeared, the experimenter asked the children what they thought was inside the box (reality question). If children were not able to answer the reality question, and therefore, did not demonstrate a belief, the task was stopped. If the children were able to answer the reality question, the experimenter opened the box and revealed that pencils were inside. The experimenter showed surprise about the content ("Oops, there is no chocolate in it, there are pencils in it!"). After this demonstration, she closed the box and asked the children about the actual content of it, to ensure they were aware of the actual content. If children answered incorrectly or did not answer at all, the actual content was revealed once again, followed again by the reality question. If children answered incorrectly three times, the task was stopped. If children named the current content, the experimenter asked two questions. One question referred to the children's former belief (self-question: "You and I know there are pencils inside. And what did you think earlier was inside this box?"). The other question referred to the belief of the duck (other-question: "You and I know there are pencils inside. If now the duck would come back, what would she think was inside the box?"). The children's answer was scored as correct if they answered both questions with 'chocolate' or 'sweets'. For the final analysis, the answer to the question about the hand puppet's belief was used, because all tasks of the ToM battery measured the understanding of others' mental states. The order of the questions was counterbalanced across children.

A task created by Southgate, Chevallier, and Csibra (2010) was modified and used as the false-believe unexpected-transfer task. The experimenter sat in front of the children at a table, the two facing each other. There were two containers on the table, which were centered on the left and right side (a basket with a lid on the left side, and a cardboard box with a lid on the right side). The experimenter showed two novel objects (object A and object B) and placed them in the middle of the table. Children were then allowed to get familiar with the objects. After 20 seconds had elapsed, the experimenter put object A in the basket and object B in the cardboard box. She commented her actions ("I put this one in here, and I put this one in here."). After she finished, she left the room. In her absence, the assistant switched the two objects' locations in a sneaky manner ("[Name of the experimenter] cannot see or hear us. I have an idea! I take this one out and put it in here (putting object A into the cardboard box). And now, I take this one out and put it in here (putting object B into the basket).") After she finished the exchange, she told the children not to inform the experimenter ("But [name of the child], shhhh!"). As soon as she sat down again, the experimenter returned and sat down at the table. She looked at both containers, pointed to the basket and said: "Now, I want to have the Modi!" By this, she demonstrated her knowledge of the object names, and that she wanted to get object A as she put it in the basket at the beginning. She pushed both containers closer to the children simultaneously and centered her hand between the containers to receive the object. Their behavior was scored as correct, if children gave the object from the box the experimenter did not point at. The container children touched first was coded as well. After finishing the task, children were asked to point at the object they liked most, to identify their preference. The location of the containers and objects were counterbalanced across children.

False-belief understanding – 4-year-old children. For the unexpected-content task (Perner, Leekam, & Wimmer, 1987), a package of a well-known chocolate (i.e., 'Kinderschokolade') and colorful pencils served as material. During the warm-up procedure, children were asked to name their best friend. At test, after presenting the closed package of chocolate to children, they were asked what they thought was inside. If they were not able to label the content as chocolate or sweets, and hence, did not demonstrate a false belief, the procedure was not continued. If they answered 'chocolate' or 'sweets', the experimenter opened the box and revealed the unexpected content, i.e. pencils, showing surprise. The experimenter put the pencils back into the box, and closed the box. She then asked children the reality question, that is, what actually was inside the box, to ensure that children were aware of the real content. If children answered incorrectly, the experimenter revealed the actual content once more, followed by the reality question. The maximum number of demonstrations was three times. If children answered incorrectly three times, the procedure ended. If children answered correctly, the experimenter asked two test questions: One about the children's former belief, the 'self-question' ("What did you first think was inside this box?") and one about the belief of the best friend, the 'other-question' ("If [best friend's name] saw this box, what would [she/he] think is inside?"). Children's answers were scored as correct if they answered the test questions with 'chocolate' or 'sweets'. For the final analyses, I only used the other-question because all tasks of the ToM battery measure the understanding of others' mental states. The order of the questions was counterbalanced across children.

The unexpected-transfer task consisted of three puppet games which were based on the Maxi-and-the-chocolate task (Wimmer & Perner, 1983). Children sat at a table with the experimenter sitting opposite them. The experimenter played out each story with two hand puppets, which were well-known characters from a German television show for children. At the beginning of each trial, the experimenter told children the puppets' names and subsequently tested whether they knew the names of the puppets (i.e., "Show me, who is Pittiplatsch!"). In one of the stories, Pittiplatsch owned a footlocker and Schnatterinchen owned a basket. Both characters put their containers onto the table and took a position behind their containers. Next, Pittiplatsch put a ball in front of his footlocker, opened the footlocker, put the ball inside, closed the lid, and left the scene. In his absence, Schnatterinchen opened the footlocker, put the ball in front of it, closed the lid, put the ball in front of her basket, opened the basket, put the ball inside, closed the lid, and left as well. All actions were narrated by the experimenter. After this demonstration, the experimenter asked the test question "Where will Pittiplatsch look for his ball when he returns?" Children's responses were scored correct, if they indicated that Pittiplatsch will look for the ball in the footlocker, hence, were able to infer the false belief of Pittiplatsch. Subsequently, the experimenter asked the reality question "Where is the ball really now?", and the memory question "Where did Pittiplatsch put the ball in the first place?". The answers for the reality and the memory questions were coded as correct if children indicated the correct locations. If memory or control questions were answered incorrectly, children were excluded from the final analysis. Two other versions of the story were presented with different protagonists and objects but with an identical course of action. The second story was presented with Moppy and Sandmann who owned a pot and a small bucket and one rubber duck. The third story was presented with Mr. Fox and Ms. Magpie who owned a bowl, and a cardboard box, respectively, and one red block. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have achieved. The sides of protagonists and objects were counterbalanced.

Inhibitory Control Batteries

The IC batteries were designed to measure inhibitory control in various dimensions. Two tasks of each battery tested the inhibition of a dominant prepotent response in favor of a subdominant response. The dominant prepotent responses were either directed to a strongly associated concept, or a salient location. The third task of each battery tested the inhibition of a dominant prepotent response for a certain period of time, commonly referred to by the term 'delay of gratification'. Tasks for the three age groups are comparable concerning the dimensions of IC investigated and become increasingly difficult with increasing age.

Inhibition of concept – 2-year-old children. A reverse categorization task from Carlson et al. (2004) was adopted to test children's ability to inhibit a response referring to conceptual processes. The experimenter sat in front of the children at a table, the two facing each other. She put a big bucket (height = 27 cm, Ø = 25.5 cm) on the right side and a small bucket (height = 18.5 cm, Ø = 16.5 cm) on the left side on the table. In the familiarization phase, she told the children to sort the blocks into the buckets matching their sizes ("The small blocks have to go into the small bucket and the big blocks have to go into the big bucket."). She performed six trials in the familiarization phase, three trials with small blocks (3 cm x 3 cm x 3 cm), and three trials with big blocks (6 cm x 6 cm x 6 cm). After the familiarization trials she offered the children a big block and asked them to sort the block into the correct bucket. If they put the big block into the small bucket, the experimenter corrected them by demonstrating the correct action, i.e. she sorted the big block into the big bucket. After six trials the experimenter removed the blocks from the buckets and started the test phase. She explained that they will play a fun game. This time, the small blocks have to go into the big bucket and the big blocks have to go into the small bucket. After two demonstration trials, children undertook twelve test trials in a pseudo-randomized order. No feedback was given during the test trials. Children's behavior was scored as correct when they placed a block into the reversed sized buckets. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have achieved. The position of the big bucket and small bucket was counterbalanced across children.

Inhibition of concept – 3-year-old children. For the sample of 3-year-old children, a Stroop-like task (Carlson & Moses, 2001) was adopted. The experimenter sat in front of the children at a table, facing each other. She put two cards (i.e., a green and a white card) on the table next to each other with a distance of 15 cm between them. She asked the children to point to the white card if she says 'grass' and to point to the green card if she says 'snow'. Two trials were performed, one trial for each of the cards. If children did not point to the correct card, the experimenter explained the rule again. A maximum of 3 rounds was given. If children did not perform correctly at the third round, the experimenter ended the task. As soon as children were able to complete one round (i.e., performing correctly in both trials), 16 test trials followed in a pseudo-randomized order. Children's behavior was scored as correct for each trial where they pointed to the appropriate card. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have achieved. The positions of the green and white cards were counterbalanced across children.

Inhibition of concept – 4-year-old children. The day-and-night task (Gerstadt et al., 1994) was conducted to test 4-year-old children's inhibitory abilities concerning a conceptual conflict. Sixteen cards with a day or a night symbol were used. Eight day cards showed a light blue sky with a white cloud and a big yellow sun. Eight night cards showed a black sky with little yellow stars and a crescent moon. The experimenter started the task by telling children that they would play a fun game. The experimenter placed the pile of cards upside-down in the middle of the table, took the first card (a night card) showed it to children and asked them to always say the word ‘day’ whenever seeing this card. Then the experimenter proceeded with the second card (a day card) and told children to say the word ‘night’ whenever seeing this card. The two cards were put back on the pile and one training round followed, to control if children were able to apply the rule. The experimenter showed the night card and asked ‘What do you say to this card?’ and proceeded in the same way with the day card. If children gave correct responses to both cards, 16 test trials started in a fixed order (see Gerstadt et al., 1994). If children gave incorrect response to one of the cards, the training round was repeated. A maximum of three training rounds was given. Children only entered the test phase if scoring correctly on at least one training round. During the test phase no feedback was given. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have achieved. Furthermore, test trials were separated into two blocks and summed to obtain scores of mean performances of trials in the first block (trials 1-8) and of the second block (trials 9-16). A division between a conservative scoring and a non-conservative scoring was made. For the conservative scoring, only the answers ‘day’ and ‘night’ were scored as correct. For the non-conservative version variations which still referred to the opposite concept were also scored as correct, for example, ‘dark’ or ‘light’.

Spatial inhibition – 2 –year-old children. In a modified A-not-B task (Diamond, Prevor, Callender, & Druin, 1997), children played a searching game. The experimenter presented the children with a music box with a xylophone inside, put two little blocks (3 cm x 3 cm x 3 cm) one after the other into it and thereby produced a descending sound. She then asked the children to throw two blocks into the box themselves. Then, she took a seat in front of the children at a table and put a tray on it. She performed three trials to familiarize the children with the searching procedure of the task. Therefore, she presented another block, put it on the tray, put an occluder in front of it to block the children’s view and removed the occluder after a delay of 2 seconds. She put the tray closer to the children and asked them to take the block and put it into the music box. She then put a cup, which was turned upside-down, in the middle of the tray, put a block under the cup, pushed it to the right side, put the occluder in front of the children, removed it after a delay of 2 seconds, and put the tray closer to the children while saying "It's your turn to search for the block!". The last trial was identical, except that she pushed the cup to the left side on the tray. In the test trials, she presented a block, put it under the cup, pushed it to the left side, put the occluder in front of the children, and put an identical cup on the right side of the tray. After a delay of 2 seconds she removed the occluder and put the tray closer to the children while saying "Now you are allowed to search for the block!" Thus, children were presented with two cups. The same trial was repeated until children lifted the cup containing the block two times in a row. If so, the following trial was performed by pushing the cup to the other side (i.e., to the right side). The procedure was repeated for a maximum of 10 trials. Consequently, they received a maximum of four switches. Children's behavior was scored as correct for each successful retrieval of the block. If children were lifting the cup containing no block five times in a row, they received a trial with only one cup to keep their motivation high. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have achieved. The side to which the first cup was pushed was counterbalanced across children.

Spatial inhibition – 3-year-old children. In a modified version of the windows task (Russell et al., 1991), children could win stickers in a competitive game against an assistant. The assistant sat in front of the children at a table, the two facing each other. The experimenter sat at the head of the table and instructed the children to win as many stickers as possible by playing against an opponent (i.e., the assistant). She started a familiarization phase and put two identical opaque boxes (12 cm x 12 cm x 12 cm) in the middle of the table, between the children and the opponent. In each trial, the experimenter instructed the children and the assistant to turn around. Meanwhile, as both were turned away from the table, she told them she would now hide a sticker under one of the boxes. After this procedure, children should point at the box, in which their opponent should search for the sticker. If this box contained the sticker, the opponent received it and children received nothing. If this box was empty, the opponent received nothing and the children received the sticker from the other box. After the children pointed at one of the boxes, the experimenter lifted it and exposed the content. If the sticker was inside, she gave it the opponent. If no sticker was inside, she lifted the other box and gave it to the children. A total of four trials were performed to familiarize children with the procedure. The experimenter controlled the familiarization phase and arranged for the children and their opponent to win two stickers each. That is, instead of actually hiding the sticker at the beginning of the trial, she secretly let it slip under the box when lifting the box. In this way, she was able to coordinate who should win the sticker. On the fifth trial, the training phase started. After children pointed at one of the boxes, the experimenter lifted both boxes, paused, and asked the children who received the sticker. This was done to check if they understood the rules. If children were not able to answer according to the rule, the experimenter explained the rule again. A maximum number of nine trials was presented. If children were not able to pass three trials of the training phase, the game ended. If children passed three trials, the experimenter started the test phase. She replaced the boxes with two similar boxes with an open side that faced the children. The opponent could not see into the boxes. The procedure of the test trials was identical to the procedure of the familiarization trials, except that the location of the sticker followed a pseudo-randomized order. A maximum of 15 trials was conducted. Children’s behavior was scored as correct for each trial in which they pointed at the empty box. Each child received a total percentage score, which was derived by dividing the number of correct responses by the maximum number of correct responses that children could have reached.

Spatial inhibition – 4-year-old children. To test 4-year-old children's ability to inhibit a prepotent spatial impulse, the same task and scoring system was used as for the 3-year-old children. The task differed slightly only in two aspects. The first was that children only received two trials instead of four trials in the familiarization phase. The second change concerned the procedure of the 15 pseudo-randomized test trials. The experimenter used an occluder to block children's view of the boxes for 3 seconds at the beginning of each trial. Before every test trial the experimenter repeated the rule and afterwards lifted the occluder.

Delay of gratification – 2-year-old children. To measure the ability to resist sweets, a waiting task was administered (Golden et al., 1977). The experimenter sat in front of the children at a table, the two facing. A tray and a bell were put in the middle of the table. For one training trial, the experimenter put a wooden block on the tray. Then she said that she will turn around and that the children should wait and should not take the block until she turned back and rang the bell. Children had to wait for 20 seconds until the experimenter rang the bell. If children did not wait, the trial was repeated. A maximum of five training trails were given. If children did not follow the rule, the task ended. If children waited, the experimenter continued with the first test trial. She put three identical cups upside-down on the tray and presented a gummy bear on a spoon ("Look what I have!"). She put the gummy bear under one of the cups and said "Wait until I ring the bell, afterwards you are allowed to take the gummy bear!" Ten test trials were performed, ranging in duration from 5 to 50 seconds in length and alternated via cups, which were chosen to hide the gummy bear. I coded the time until children touched the sweets, because this was always followed by the children eating it. For every child a mean total score was calculated and converted into percentages of mean performance.

Delay of gratification – 3-year-old children. To measure children's ability to delay a gratification I modified the task by Mischel et al. (1988). Materials consisted of a paper plate and two kinds of sweets; either two small bags of gummy bears or two small bars of chocolate. Children were first asked if they preferred gummy bears or chocolate. Children were then asked if they preferred to have one package or two packages of their favored sweet, to test whether they really liked the sweet. If children were shy and answered they would just take one package, the experimenter smiled and asked if two of them would not be much better. When the experimenter repeated the question, all children said they wanted to have two of them. The experimenter opened one of the packages while saying that this is the first bag of gummy bears or the first bar of chocolate. The gummy bears were scattered on the plate. The bar of chocolate was broken into pieces. The experimenter smelled it and mentioned that it looked really good and smelled delicious. The experimenter then mentioned she had to leave the room for a moment and instructed children that they could either eat the treat right now or they could wait until she got back and would be then given the second package as well. The experimenter then left the room and waited outside for a maximum of 5 minutes. In order not to leave the children unattended, an assistant stayed in the room, sat in a corner out of children's view and pretended to write something in order to make children believe that they were unobserved. The duration of the period before the children touched, licked, or ate the sweets during the 5 minutes response period was measured. Children received a total score for the time that they waited before touching the sweets, before licking the sweets, and before eating the sweets. All scores were converted into percentages of waiting time (of a total time of 5 minutes). For the final analyses the scores until children touched the sweets were used, because the touch-scores revealed greater variance than the ate-scores, since children were very good at resisting eating the sweets. Only 22 out of 90 children actually ate the sweets.

Delay of gratification – 4-year-old children. To test 4-year-old children's ability to delay gratification, the same task and scoring system was used, as for the 3-year-old children. The task differed only with regards to the waiting time. Children had to wait for a maximum of 7 minutes. For the final analyses the scores for the periods until children touched the sweets were used, because the touch-scores revealed greater variance than the ate-scores, since children were very good at resisting eating the sweets. Only nine out of 61 children ate the sweets.

Language Assessment

To assess children's language abilities, sentence comprehension and word comprehension were assessed with the SETK 2 and SETK 3 – 5 (German language development test for 2, and 3- to 5-year-olds, Grimm, 2001; Grimm et al., 2000). Children received the SETK at the end of the first test session. The 2-year-old children were shown a selection of pictures. The experimenter read words and sentences to the children and they were asked to identify the picture representing what had been described (e.g., "Show me the picture in which children are sitting under the table."). The 3-year-old children and 4-year-old children were presented with different objects arranged in a given order on the table. They were asked to follow the experimenter's instructions, which always referred to simple actions to be performed with the objects (e.g., "Put the bag between the pencils."). The raw values were transferred into T-values for the final analysis.

Behavior and Temperament Assessment

To assess behavioral problems and temperament traits parents were invited to complete two questionnaires, the CBCL (Child Behavior Checklist for Ages 1 ½ - 5, Achenbach & Rescorla, 2000) and the IKT (The Integrative Child Temperament Inventory, Zentner, 2011). To increase reliability and check whether children would be evaluated similarly by parents and caregivers, children's kindergarten teachers were asked to complete the C-TRF (Caregiver – Teacher Report Form, Achenbach & Rescorla, 2000). The CBCL consists of 100 questions for which parents estimate the frequency of their children's behavior over the last two months, deciding among 'not true', 'sometimes true' and 'often true'. The items are assigned to eight scales of syndromes. Four scales relate to internalizing factors, which are Emotionally Reactive, Anxious/Depressed, Somatic Complaints, and Withdrawn. The externalizing factors include the syndrome scales Attention Problems and Aggressive Behavior. Both scores are combined and combined with the remaining scales for Sleeping Problems and Other Problems to obtain a total problem score. For the final analysis, internalizing ($\alpha = .89$) and externalizing scores ($\alpha = .92$) were used, as well as the scores from the Other Problems scale. The C-TRF can be considered equivalent to the CBCL, but there is no syndrome scale for sleeping problems. For the final analysis I used the internalizing ($\alpha = .89$) and externalizing scores ($\alpha = .96$), as well as the scores of the Other Problems scale. The IKT offers 30 questions on a 6-point Likert-scale, ranging from 'not true' to 'always true'. The raw values are summed up to five subscales of temperament dimensions containing Frustration ($\alpha = .81$), Social Inhibition/Shyness ($\alpha = .80$), Activity ($\alpha = .85$), Attention ($\alpha = .81$) and Sensory Sensitivity ($\alpha = .70$). For the final analysis values were converted into percentile ranks.

Data Reduction and Reliability

Data Reduction. Since more than one main experimenter was involved in administering the tasks, I used Kruskal-Wallis tests to check for possible influences. I found group effects in the sample of the 3-year-old children for the spatial-inhibition task ($\chi^2(2) = 7.89$, $p = .019$) and the delay-of-gratification task ($\chi^2(2) = 7.47$, $p = .024$). Post-hoc Mann-Whitney's *U*-tests revealed differences between Experimenter 1 and Experimenter 2 for the spatial-inhibition task ($U = -2.40$, $p < .017$) and the delay-of-gratification touch score ($U = -2.256$, $p < .024$). For these two tasks, videos where rechecked and no differences in the testing procedures were found. Since protocols did not deviate from each other and differences appear only between two experimenters, no data reduction for this sample was performed. However, for the sample of the 4-year-old children data reduction was necessary. For the delay-of-gratification task, the data from one experimenter had to be excluded from the analyses due to a significant group effect (Kruskal-Wallis test, $\chi^2(6, N = 72) = 13.8$, $p = .032$).

Post-hoc tests using Mann-Whitney's *U*-tests showed that the results from this experimenter differed from that of the four other experimenters (all p-values $\leq .046$). For the preference understanding task, the data from another experimenter was excluded from the analyses due to a significant group effect (Kruskal-Wallis test, $\chi^2(6, N = 78) = 13.9$, $p = .031$). Post-hoc tests using Mann-Whitney's *U*-tests showed that the results from this experimenter significantly differed from two other experimenters (all p-values $\leq .012$). For all other tasks, the analyses did not show any differences in children's performances across experimenters (all p-values $\geq .078$).

Reliability. To determine interrater-reliability, two naïve independent persons coded 25 % of the videos again. Agreement for the samples of 2-, 3- and 4-year-olds was excellent ($\alpha > .91$ for the false-belief tasks and $r_s > .88$ for the other tasks). Wilcoxon tests also revealed no differences between coders (all p-values $\geq .157$).

Test-Retest Reliability – 2-year-old children. Because the first task for each child was re-administered at the beginning of the second session, I analyzed test-retest reliability. For the sample of the 2-year-old children I found correlations between test and re-test values for goal understanding ($r_s = .685$, $N = 8$, $p = .061$), intention understanding ($r_s = .802$, $N = 8$, $p = .017$), preference understanding ($r_s = .808$, $N = 8$, $p = .015$), inhibition-of-concept ($r_s = .939$, $N = 8$, $p = .001$), spatial-inhibition ($r_s = .807$, $N = 10$, $p = .005$) and delay-of-gratification tasks ($r_s = .641$, $N = 10$, $p = .046$). The false-belief unexpected-content task showed acceptable reliability ($\alpha = .73$, $N = 4$), whereas the false-belief unexpected-transfer tasks showed no reliability ($\alpha = .17$, $N = 7$).

Test-Retest Reliability – 3-year-old children. Test-retest reliability for the sample of the 3-year-old children showed high correlations between test and re-test values for the spatial-inhibition ($r_s = .900$, $N = 7$, $p = .006$) and delay-of-gratification task ($r_s = .868$, $N = 9$, $p = .002$), moderate correlations for the goal understanding ($r_s = .536$, $N = 7$, $p = .215$), and only small correlations for the intention understanding ($r_s = .216$, $N = 9$, $p = .578$) and inhibition-of-concept task (r_s = .171, N = 10, p = .637) . The false-belief unexpected-transfer task showed no reliability (α = .36, N = 10). Furthermore, the sample size for the preference understanding task was too small for a reliable analysis, as well as the sample for the false-belief unexpected-content task because most children remembered the content revealed in the first session.

Test-Retest Reliability – 4-year-old children. Test-retest reliability was confirmed with strong correlations for the intention understanding (r_s = .815, p = .004), spatial-inhibition (r_s = .718, p = .172) and delay-of-gratification tasks (r_s = .500, p = .391), and with moderate correlations for the preference understanding (r_s = .395, p = .333) and inhibition-of-concept tasks (r_s = .316, p = .408). The false-belief unexpected-transfer (r_s = .216, p = .607) and the goal understanding tasks (r_s = .284, p = .397) only showed small correlation coefficients. The false-belief unexpected-content task was excluded from reliability analyses, because most children remembered the unexpected content from the first test session.

2.2 Results

Descriptive Statistics

To display the overall multidimensional understanding of mental states, each child received a total score of ToM, composed of the total scores of the goal understanding, the preference understanding, the intention understanding task and a collapsed score of the two false-belief understanding tasks, which were transformed into z-values for the correlational analyses. In the same way, each child received an IC score, composed of the total scores of the inhibition-of-concept task, the spatial-inhibition task, and the touch scores of the delay-of-gratification task, which were also transformed into z-values for the correlational analyses. Table 1 shows the descriptive statistics for the total scores of ToM, IC and language.

Table 1.

Descriptive Statistics for Scores of Language, ToM and IC of 2-, 3- and 4-Year-Old Children

	Mean	SD	Range
2-year-old children			
Language	49.7	7.5	34.0 - 66.5
ToM	46.0	18.5	0.0 - 91.7
IC	56.4	16.2	24.0 - 100
3-year-old children			
Language	50.0	10.1	23.0 - 71.0
ToM	53.5	11.8	20.8 - 90.0
IC	44.9	27.4	0.0 - 100
4-year-old children			
Language	52.0	11.5	32.0 - 77.0
ToM	55.6	17.8	25.0 - 100
IC	54.4	27.9	0.0 - 100

Gender Differences

For all tasks, Mann-Whitney's *U*-tests were ran to check for gender differences. Two-year-old girls scored better in the false-belief unexpected-content task than two-year-old boys ($\chi^2(57) = 3.932$, $p = .047$, $r = .521$). For the sample of the 3-year-old and 4-year-old children, there were no differences between girls' or boys' scores within the task batteries (all ps $\geq .079$).

Regarding gender differences in behavioral problems and temperament, Mann-Whitney's U-tests were run for the CBCL and C-TRF internalizing and externalizing scores, the scores of the Other Problem scale, and the IKT dimension scales. For parents' ratings of 2-year-old children, it was found that boys received higher scores than girls on the CBCL Other Problem scale ($U = 533.5$, $N = 40$, $N = 39$, $Z = -2.429$, $p = .015$, $r = 0.273$), concerning the CBCL externalizing scores ($U = 532.5$, $N = 40$, $N = 39$, $Z = -2.432$, $p = .015$, $r = 0.274$) and on the IKT Activity scale ($U = 487.5$, $N = 40$, $N = 39$, $Z = -2.436$, $p = .015$, $r = 0.274$). For parents' ratings of the 4-year-old children, it was found that boys received lower ratings than girls on the IKT Activity scale ($U =390$, $N = 38$, $N = 32$, $Z = -2.557$, $p = .010$, $r = 0.306$). No statistically significant results were revealed for the sample of the 3-year-old children and for caregiver ratings in general.

Session Order Effects

It was also checked whether children scored differently depending on the type of session they received first. With regard to the large amount of tasks, only a few effects were revealed. Two-year-old children who received the IC session second, passed the inhibition-of-concept task better than children who received the IC session first (U = 479.5, N = 32, N = 41, Z = -1.986, p = .047, r = 0.232). Analyzing the IC score in relation to the session order, no difference was found (p = .996). Four-year-old children who received the ToM session first passed the goal understanding task better than children who received the ToM session second

(U =566, Nfirst = 40, Nsecond = 38, Z = -1.971, p = .049, r = 0.223). For the delay-of-gratification task it was found that 4-year-old children, who received the IC session first showed shorter times on waiting to touch the sweets than children who got the task on the second occasion (U =332, Nfirst = 32, Nsecond = 29, Z = -2.063, p = .039, r = 0.264). Analyzing the ToM and IC scores in relation to the session order, no difference was found (all ps ≥ .053). No other effects of session order were revealed (all ps ≥ .055).

Correlation and Comparison of Parents' and Caregivers' Ratings

Sample of 2-year-old children. Parents of the 2-year-old children completed and returned 96% of questionnaires of the CBCL and 96% of questionnaires of the IKT. Caregivers of the 2-year-old children completed and returned 57% of questionnaires of the C-TRF. Pearson correlations were run and showed positive correlations between the CBCL externalizing and C-TRF externalizing scores. The following syndrome scales were positively correlated concerning parents' and caregivers' rating: Emotionally Reactive, Withdrawn, Aggressive Behavior and Other Problems scales (all *r*s ≥ .392, all *p*s ≤ .007). Additionally, caregivers' ratings of C-TRF externalizing scores showed positive correlations to parents' ratings on the IKT Frustration scale and negative correlations to the IKT Attention scale. Caregivers' ratings on the Other Problems scale showed negative correlations to parents' ratings on the IKT Attention scale. For correlation coefficients see Table 2. To enable a direct comparison of informants' judgments, I further compared the internalizing scores, externalizing scores and the other-problems scores of parents and caregivers. A significant difference concerning mean scores of the scales was detected for the Other Problems scale and the externalizing scores. On average, parents rated children higher than caregivers did on the Other Problem scale (Wilcoxon test, $N = 47$, $Z = -3.479$, $p = .001$, $r = .507$) but lower with regards to externalizing scores (Wilcoxon test, $N = 47$, $Z = -1.984$, $p = .047$, $r = .289$).

Sample of 3-year-old children. Parents of the 3-year-old children completed and returned 96% of questionnaires of the CBCL and 96% of questionnaires of the IKT. Caregivers of the 2-year-old children completed and returned 93% of questionnaires of the C-TRF. Pearson correlations were run and revealed no correlations for the internalizing and externalizing scores, or for the scores on the Other Problem scales. Positive correlations were found between the following syndrome scales concerning parents' and caregivers' rating: Anxious/Depressed and Somatic Complaints scales (all rs $\geq .382$, all ps $\leq .001$). Caregivers' ratings of C-TRF externalizing scores correlated positively with parents' ratings on the IKT Frustration scale. For correlation coefficient see Table 4. Regarding the direct comparison of informant's judgment, a difference was found for parents' and caregivers' ratings for the internalizing and externalizing scores. On average, parents rated children lower on the internalizing and externalizing scores than caregivers (Wilcoxon test, internalizing scores: $N = 80$, $Z = -2.793$, $p = .005$, $r = .312$; externalizing scores, $N = 80$, $Z = -3.866$, $p < .001$, $r = .432$).

Sample of the 4-year-old children. Parents of the 4-year-old children completed and returned 89% of questionnaires of the CBCL and 88% of questionnaires of the IKT. Caregivers completed and returned 95% of questionnaires of the C-TRF. Pearson correlations were run and revealed no correlations for the internalizing and externalizing scores, or for scores on the Other Problem scale. Notably, only one statistically significant correlation was found between parents' and caregivers' questionnaires for behavioral problems. The CBCL Aggressive Behavior scale correlated positively with the C-TRF external scores ($r = .251$, $p = .040$). The direct comparison of informants' judgments detected a significant difference in scores on the Other Problems scale, as parents rated children higher than caregivers did (Wilcoxon test, $N = 67$, $Z = -3.417$, $p = .001$, $r = .415$).

Comparison of behavior ratings across the different age groups. To compare the ratings of behavioral problems across all age groups, Kruskal-Wallis tests were run. A significant group effect was only detected for caregivers C-RTF internalizing scores (χ^2(2, $N =$

206) = 9.0, p = .011). Post-hoc tests using Mann-Whitney's U-tests showed that 3-year-old children received higher internalizing scores from caregivers compared to 4-year-old children (U = 2577.5, N = 83, N = 76, Z = -1.992, p = .046, r = 0.158). In addition, 3-year-old children received higher internalizing scores from caregivers compared to 2-year-old children (U = 1338.5, N = 83, N = 47, Z = -2.972, p = .003, r = 0.261). Even if the difference between 2- and 3-year-old children is statistically significant, it should be noted that the sample size of the 2-year-old children is fairly small with an N of 47. For mothers' ratings of behavioral problems, no statistically significant difference was found between the age groups.

Correlations Among Age, Language, Theory of Mind, Inhibitory Control, Behavioral Problems and Temperament of 2-Year-Old Children

For the final correlations of the 2-year-olds, total scores of language, ToM and IC, CBCL and C-TRF internalizing and externalizing scores, as well as the scores of the CBCL and C-TRF Other Problem scale were entered into Pearson correlations. Because the IKT does not offer a total score, all dimension scales were entered into the correlation. Language showed a positive correlation to IC scores and a negative correlation to the IKT Sensory Sensitivity scale, but showed no correlation to the any of the CBCL or C-TRF scores. The IC scores showed a significant negative correlation to the IKT Activity scale and at trend level a negative correlation to the C-TRF Other Problem scale. CBCL scores were correlated to various IKT scales. Within the IKT scales, Frustration was positively correlated to Activity and negatively correlated to Attention. Furthermore, age was negatively correlated to C-TRF internalizing scores. For correlation coefficients see Table 2.

To control for language and gender as possible influencing factors, partial correlations were run to further investigate the link between IC, ToM, behavioral problems and temperament. The negative correlation at trend level between IC and the C-TRF Other Problem scale disappeared. All other correlations remained stable.

Table 2.

Pearson Correlation Coefficients Between Age, Language, ToM Scores, IC Scores, CBCL Internalizing and Externalizing Scores, Scores on the Other Problems Scale, C-TRF Internalizing and Externalizing Scores, Scores on the Other Problems Scale, and Scores on the IKT Dimension Scales for the Sample of 2-Year-Old Children

	Age	**SETK**	**ToM**	**IC**	**CBCL o.p.**	**CBCL int.**	**CBCL ext.**	**C-TRF o.p.**	**C-TRF int.**	**C-TRF ext.**	**IKT frust.**	**IKT inhib.**	**IKT activ.**	**IKT atten.**	**IKT senso.**
Age	-														
SETK	.058	-													
ToM	-.102	.047	-												
IC	.092	.234 *	-.035	-											
CBCL o.p.	-.012	-.056	.010	-.020	-										
CBCL int.	-.014	.024	-.074	.128	.596 **	-									
CBCL ext.	.027	-.039	-.113	-.013	.709 **	.565 **	-								
C-TRF o.p.	-.130	-.203	-.196	-.260 ^{t}	.439 **	.119	.329 *	-							
C-TRF int.	-.322 *	-.051	-.035	.120	.202	.211	.232	.558 **	-						
C-TRF ext.	-.230	-.183	-.093	-.193	.351 *	.120	.383 **	.727 **	.506 **	-					
IKT frust.	-.027	-.054	-.106	-.083	.522 **	.580 **	.605 **	.223	.080	.319 *	-				
IKT inhib.	.045	.087	-.074	.175	.207	.270 *	-.027	.087	.203	-.092	-.046	-			
IKT activ.	-.022	-.094	-.136	-.246 *	.301 **	.148	.407 **	.135	-.106	.067	.426 **	-.224	-		
IKT atten.	-.033	.045	.118	.104	-.330 **	-.262 *	-.493 **	-.323 *	-.086	-.365 *	-.550 **	.014	-.168	-	
IKT senso.	-.179	-.311 **	-.046	-.044	.009	.286 *	-.007	-.087	-.066	.073	.203	.033	.052	.050	-

Note: ** p < .01, * p < .05, ^{t} p < .10, all ps two-tailed.

ToM = ToM total score, IC = IC total score; CBCL o.p. = CBCL Other Problem scale; CBCL int. = CBCL internalizing problems score; CBCL ext. = CBCL externalizing problems score; C-TRF o.p. = C-TRF Other Problem scale; C-TRF int. = C-TRF internalizing problems score; C-TRF ext. = C-TRF externalizing problems score; SETK = Language; IKT frust. = Frustration; IKT inhib. = Social Inhibition/Shyness; IKT active. = Activity; IKT atten. = Attention; IKT senso. = Sensory Sensitivity

Multiple Regression Analyses for the Sample of 2-Year-Old Children

To investigate the relative importance of the temperament dimension Activity on IC at 2 years of age, a standard linear regression was performed, entering the IC scores as the criterion, and IKT Activity as the predictor variable. The model showed statistical significance ($F(1,73) = 4.620$, $p = .035$) with $R^2 = .060$ (adjusted $R^2 = .047$). A hierarchical regression was performed to control for other influences. At step one, language and gender were entered to account for potential confounding effects with the variable of interest, then at step 2 IKT Activity was entered. The overall model reached statistical significance only at trend level ($F(3,68) = 2.511$, $p = .066$) with $R^2 = .104$ (adjusted $R^2 = .062$). Only language contributed significantly to the prediction of IC within the model. For a summary, see Table 3.

Table 3.

Summary of Hierarchical Regression Analysis Predicting Inhibitory Control at 2 Years of Age

Predictor	Inhibitory Control at 2 Years of Age			
	β	t	F change	R^2 change
Step 1				
Gender	.061	.511	2.126	.061
Language	.240 *	2.009		
Step 2				
Activity	-.215 ^{t}	-1.773	3.143 ^{t}	.043

Note: * $p < .05$, ^{t} $p < .10$, Gender is a dichotomous variable (1 represents female, 2 represents male)

Correlations Among Age, Language, Theory of Mind, Inhibitory Control, Behavioral Problems and Temperament of 3-Year-Old Children

For the final correlations of the 3-year-olds' total scores for language, ToM and IC, CBCL and C-TRF internalizing and externalizing scores, as well as the scores of the CBCL and C-TRF Other Problem scale were entered into Pearson correlations. Language was significantly positively related to age and IC scores, and showed negative correlations to the C-TRF Other Problem scale and C-TRF externalizing scores only at trend level. Language was also positively correlated to IC, but only at trend level to ToM. No correlation between the ToM or IC scores and any of the behavior or temperament dimension scales were found. Various correlations were found between parents' ratings on the CBCL and IKT. Within the IKT scales, Frustration was positively correlated to Activity and Sensory Sensitivity, and negatively correlated to Attention. Activity was negatively correlated to Attention. For correlation coefficients see Table 4.

To further investigate the links between ToM, IC, behavioral problems and temperament, a partial correlation was run, controlling for language and gender. All correlations among behavior and temperament scales remained stable. No correlations appeared between the ToM or IC scores and any of the behavior or temperament dimension scales.

Table 4.

Pearson Correlation Coefficients Between Age, Language, ToM Scores, IC Scores, CBCL Internalizing and Externalizing Scores, Scores on the Other Problems Scale, C-TRF Internalizing and Externalizing Scores, Scores on the Other Problems Scale, and Scores on the IKT Dimension Scales for the Sample of 3-Year-Old Children

	Age	**SETK**	**ToM**	**IC**	**CBCL o.p.**	**CBCL int.**	**CBCL ext.**	**C-TRF o.p.**	**C-TRF int.**	**C-TRF ext.**	**IKT frust.**	**IKT inhib.**	**IKT activ.**	**IKT atten.**	**IKT senso.**
Age	-														
SETK	.220 *	-													
ToM	.050	.184 ^{t}	-												
IC	-.076	.316 **	.041	-											
CBCL o.p.	-.085	.118	-.116	.047	-										
CBCL int.	-.014	.110	-.165	-.019	.690 **	-									
CBCL ext.	-.047	.087	-.052	.071	.578 **	.402 **	-								
C-TRF o.p.	-.018	-.211 ^{t}	-.024	-.054	.123	.121	.036	-							
C-TRF int.	-.020	-.156	-.107	.124	.163	.170	.028	.709 **	-						
C-TRF ext.	-.085	-.209 ^{t}	.011	-.036	.100	-.055	.162	.696 **	.600 **	-					
IKT frust.	-.069	-.005	-.125	.043	.393 **	.285 **	.638 **	.104	.160	.210 ^{t}	-				
IKT inhib.	-.034	.053	.045	-.002	.129	.267 *	-.099	.121	.209 ^{t}	-.157	.009	-			
IKT activ.	.101	.013	.083	.167	.176	-.014	.439 **	-.152	-.061	.078	.503 **	-.173	-		
IKT atten.	-.094	-.107	-.068	-.028	-.300 **	-.137	-.465 **	-.111	-.009	-.097	-.444 **	-.079	-.374 **	-	
IKT senso.	.161	-.031	-.169	.006	.345 **	.387 **	.214	-.076	.113	-.143	.231 *	.158	.200	-.055	-

Note: ** $p < .01$, * $p < .05$, ^{t} $p < .10$, all ps two-tailed.

ToM = ToM total score, IC = IC total score; CBCL o.p. = CBCL Other Problem scale; CBCL int. = CBCL internalizing problems score; CBCL ext. = CBCL externalizing problems score; C-TRF o.p. = C-TRF Other Problem scale; C-TRF int. = C-TRF internalizing problems score; C-TRF ext. = C-TRF externalizing problems score; SETK = Language; IKT frust. = Frustration; IKT inhib. = Social Inhibition/Shyness; IKT active. = Activity; IKT atten. = Attention; IKT senso. = Sensory Sensitivity

Correlations Among Age, Language, Theory of Mind, Inhibitory Control, Behavioral Problems and Temperament of 4-Year-Old Children

For the final correlations of the 4-year-olds, total scores for language, ToM and IC, CBCL and C-TRF internalizing and externalizing scores, as well as the scores of the CBCL and C-TRF Other Problems scale were entered into Pearson correlations. Age was not correlated to any of the variables. Language was positively correlated to ToM and IC, but did not correlate with CBCL and C-TRF scores, or with IKT scales. IC was negatively correlated at trend level to C-TRF externalizing scores. ToM was negatively correlated to CBCL externalizing scores and the CBCL Other Problems scale, and positively correlated to C-TRF internalizing scores. Various correlations were found between parents' ratings on the CBCL and IKT. However, caregivers' ratings showed no links to any of the IKT scales. Within the IKT scales, Frustration was positively correlated to Social Inhibition/Shyness and Sensory Sensitivity, and negatively correlated to Attention. Notably, no correlation for any of the IKT scales and ToM or IC was found. For correlation coefficients see Table 5.

To further investigate the links between ToM, IC, behavioral problems and temperament, a partial correlation was run, controlling for language and gender. The positive correlation between ToM and IC remained significant ($r\,(73) = .243$, $p = .035$). IC remained negatively correlated at trend level to C-TRF externalizing scores ($r\,(71) = -.220$, $p = .062$) and a negative relation to the CBCL Other Problems scale appeared at trend level ($r\,(71) = -.212$, $p = .072$). ToM scores were still negatively correlated to CBCL externalizing scores ($r\,(66) = -.231$, $p = .058$) and the CBCL Other Problem scale ($r\,(66) = -.209$, $p = .076$), however, only at trend level, but reached significance in being positively correlated to C-TRF internalizing scores ($r\,(66) = .281$, $p = .016$). Additionally, ToM showed a positive correlation at trend level to the C-TRF Other Problem scale ($r\,(71) = .209$, $p = .076$). The correlations between parents' ratings on CBCL scales and IKT dimension scales remained significant for Frustration, Social Inhibition/Shyness, Attention and Sensory Sensitivity (all $p < .05$).

Table 5.

Pearson Correlation Coefficients Between Age, Language, ToM Scores, IC Scores, CBCL Internalizing and Externalizing Scores, Scores on the Other Problems Scale, C-TRF Internalizing and Externalizing Scores, Scores on the Other Problems Scale, and Scores on the IKT Dimension Scales for the Sample of 4-Year-Old Children

	Age	**SETK**	**ToM**	**IC**	**CBCL o.p.**	**CBCL int.**	**CBCL ext.**	**C-TRF o.p.**	**C-TRF int.**	**C-TRF ext.**	**IKT frust.**	**IKT inhib.**	**IKT activ.**	**IKT atten.**	**IKT senso.**
Age	-														
SETK	.014	-													
ToM	.112	.385 **	-												
IC	.045	.333 **	.329 **	-											
CBCL o.p.	-.048	.004	-.270 *	-.184	-										
CBCL int.	-.202	-.010	-.131	-.179	.735 **	-									
CBCL ext.	.051	-.113	-.258 *	-.006	.686 **	.567 **	-								
C-TRF o.p.	.104	-.005	.184	-.169	.033	.113	.128	-							
C-TRF int.	.007	-.079	.220 ^{t}	-.095	-.005	.049	.203	.673 **	-						
C-TRF ext.	-.029	-.035	.080	-.207 ^{t}	-.006	.017	.227	.617 **	.473 **	-					
IKT frust.	.057	-.034	-.139	.018	.449 *	.558 **	.542 **	-.061	.173	.052	-				
IKT inhib.	.053	.067	.022	.162	.257 *	.379 **	.085	.045	.086	-.163	.306 **	-			
IKT activ.	.045	-.031	-.033	-.079	.112	.050	.246 *	-.098	-.117	.064	.139	-.215	-		
IKT atten.	-.014	.124	.082	.086	-.183	-.267 *	-.392 **	-.070	-.182	-.125	-.437 **	.031	-.061	-	
IKT senso.	-.006	.095	-.105	-.021	.271 *	.453 **	.048	-.107	-.045	-.219	.424 **	.223	-.006	-.069	-

Note: ** $p < .01$, * $p < .05$, ^{t} $p < .10$ all ps two-tailed.

ToM = ToM total score, IC = IC total score; CBCL o.p. = CBCL Other Problem scale; CBCL int. = CBCL internalizing problems score; CBCL ext. = CBCL externalizing problems score; C-TRF o.p. = C-TRF Other Problem scale; C-TRF int. = C-TRF internalizing problems score; C-TRF ext. = C-TRF externalizing problems score; SETK = Language; IKT frust. = Frustration; IKT inhib. = Social Inhibition/Shyness; IKT active. = Activity; IKT atten. = Attention; IKT senso. = Sensory Sensitivity

Multiple Regression Analyses for the Sample of 4-Year-Old Children

To consider the contribution of cognitive factors to behavioral problems, standard multiple regressions were performed for ToM and IC related to CBCL and C-TRF externalizing and internalizing scores, as well as to scores on the CBCL and C-TRF Other Problem scale. For all models, IC and ToM were entered simultaneously as the predictors, CBCL and C-TRF scores were entered as the criterion. The model for C-TRF internalizing problems reached statistical significance ($F(2,74) = 3.332$, $p = .041$) with $R^2 = .085$ (adjusted $R^2 = .059$). ToM predicted internalizing behavioral problems rated by caregivers, whereas IC showed no direct influence (see Table 6). The model for C-TRF Other Problems reached statistical significance ($F(2,74) = 3.606$, $p = .032$) with $R^2 = .091$ (adjusted $R^2 = .066$). ToM predicted C-TRF Other Problems to a positive direction, whereas IC showed a negative prediction (see table 6.). The model for C-TRF externalizing problems reached trend level ($F(2,74) = 2.486$, $p = .090$) with $R^2 = .065$ (adjusted $R^2 = .039$). Only IC contributed to children's C-TRF externalizing problems (see Table 6).

Table 6.

Summary of Multiple Regression Analysis Predicting C-TRF Internalizing Problems, C-TRF Externalizing Problems and C-TRF Other Problems at 4 Years of Age

	Model: C-TRF Internalizing Problems score			Model: C-TRF Externalizing Problems score			Model: C-TRF Other Problems score		
Predictors	B	SE B	*β*	B	SE B	*β*	B	SE B	*β*
ToM	4.566	1.871	.289 *	2.407	1.869	.154	2.115	.950	.263 *
IC	-2.381	1.838	.126	-.327	1.537	-.253 *	-1.649	.781	-.249 *

Note: * $p < .05$

The model for CBCL Other Problems reached trend level as well ($F(2,68) = 2.911$, $p = .061$) with $R^2 = .081$ (adjusted $R^2 = .053$). Only ToM contributed to CBCL Other Problems ($\beta = -.238$, $p = .070$). All other models were statistically not significant.

For the statistically significant models, hierarchical regressions were assessed, controlling for other influences. At step one, language and gender were entered to account for potential confounding effects with the variables of interest, then at step 2 ToM and IC were entered. Neither language nor gender explained a significant portion of the variance of behavioral problems in any of the models. The model for predicting C-TRF internalizing problems reached statistical significance ($F(4,73) = 2.928$, $p = .027$) with $R^2 = .145$ (adjusted $R^2 = .096$), but still only ToM was related to internalizing problems ($\beta = .365$, $p = .004$). Also the model for C-TRF Other Problems reached statistical significance ($F(4,73) = 2.734$, $p = .036$) with $R^2 = .137$ (adjusted $R^2 = .087$). ToM predicted C-TRF Other Problems to a positive direction, whereas IC showed a negative influence. The model for predicting CBCL Other Problems reached statistical significance as well ($F(4,67) = 2.855$, $p = .031$) with $R^2 = .153$ (adjusted $R^2 = .100$); there again, ToM was related to behavioral problems but only at trend level ($\beta = -.251$, $p = .070$). For a summary see Table 7. Only the model for predicting C-TRF externalizing problems did not reach statistical significance.

Table 7.

Summary of Hierarchical Regression Analysis Predicting C-TRF Internalizing Problems, C-TRF Other Problems and CBCL Other Problems at 4 Years of Age

	Model: C-TRF Internalizing Problems				Model: C-TRF Other Problems				Model: CBCL Other Problems			
Predictor	*β*	*t*	*F* change	R^2 change	*β*	*t*	*F* change	R^2 change	*β*	*t*	*F* change	R^2 change
Step 1			.990	.027			.666	.018			1.234	.037
Gender	.140	1.188			-.026	-.217			.192	1.567		
SETK	-.108	-.912			.137	1.152			-.012	-.095		
Step 2			4.761	.118*			4.743	.118*			4.349	.117*
ToM	.365 **	2.938			.296 *	2.375			-.251 ^{t}	-1.845		
IC	-.191	-1.550			-.300 *	-2.419			-.214	-1.577		

Note: ** p < .01, * $p < .05$, ^{t} $p < .10$ (SETK = language ability)

Since strong positive correlations appeared between language, ToM and IC, multiple linear regression models were run to regress the relative importance of language for predicting ToM or IC. In the first model, ToM was entered as the criterion measure, and language and IC were entered as the predictors. The model reached statistical significance ($F(2,76) = 8.705$, $p < .001$) with $R^2 = .190$ (adjusted $R^2 = .169$) and revealed that both language and IC predicted ToM. The second model was composed to consider the relative importance of language and ToM on IC. For this model, IC was entered as the criterion measure, and language and ToM were entered as the predictor variables. The model was significant ($F(2,76) = 7.122$, $p = .001$) with $R^2 = .161$ (adjusted $R^2 = .139$) and revealed that both language and ToM predicted IC. For a summary, see Table 8.

Table 8.

Summary of Multiple Regression Analysis Predicting ToM and IC at 4 Years of Age

	Model: ToM				Model: IC		
Predictors	B	SE B	*β*	Predictors	B	SE B	*β*
SETK	.016	.006	.298 **	SETK	.015	.007	.241 *
IC	.198	.094	.235 *	ToM	.288	.136	.243 *

Note: ** p < .01, * *p* < .05, (SETK = language ability)

2.3 Discussion

The aim of this study was to investigate ToM and IC as possible correlates of behavioral problems and temperament in children's early years of life. In accordance with previous findings, the positive relationship between ToM and IC can be confirmed, although this link was only detected for children of 4 years of age. Language abilities were not correlated to behavioral problems, but showed a positive link to IC from 2 years onwards and to ToM at the age of 4 years. ToM and IC, in turn, predicted behavioral problems at the age of 4 years to different extents. The higher the ToM scores, the more internalizing behavioral problems and various behavioral problems (refers to the scores of the Other Problems scale) children showed. In contrast, the higher the IC scores, the fewer various behavioral problems children showed. Contrary to previous findings (Hughes & Ensor, 2007; Hughes & Ensor, 2008), the results do not support the assumption of a robust relationship between ToM, IC and behavioral problems at the early ages of 2 and 3 years. Considering temperament, it was found that children high in activity performed weakly on IC tasks at the age of 2 years. However, this link disappeared when taking language into account. Based on the results, I suggest that behavioral problems in 2- and 3-year-old children occur largely independently of ToM and IC abilities, whereas behavioral problems in 4-year-old children show significant relationships to their ToM and IC. Thus, it can be assumed that mental state understanding and inhibition at 4 years of age contribute to children's behavioral problems.

The Relationship Between ToM, IC, Behavioral Problems and Language

The main purpose of the current investigation was to extend the limited amount of research on the relationship between ToM, IC and behavioral problems in younger preschool children and to clarify the nature of the proposed connection (Capage & Watson, 2001;

Harvey et al., 2001; Lane et al., 2013; Wellman et al., 2011). Children's ToM and IC were assessed comprehensively, and language abilities – as an important potential correlate – were also controlled for. Language showed a positive correlation to IC for all three age groups. Thus, it is suggested that language is a meaningful component for the development of regulation processes in preschool years (Martin-Rhee & Bialystok, 2008; Wolfe & Bell, 2004). Referring to Vygotskiĭ and Cole (1978), language is a necessary tool for self-regulatory processes like planning and executing own actions, as well as for controlling spontaneous impulses. An early promotion of verbal abilities would therefore be an essential element for fostering successful cognitive development. On the other hand, IC might also be a component of developing language abilities. Studies on adults suggested that the activation and also the suppression of information are required for language processing (Faust, Balota, Duchek, Gernsbacher, & Smith, 1997; Gernsbacher & Faust, 1991).

Regarding the link between ToM and language, no statistically significant correlation was found for 2- and 3-year-old children, only for 4 year-old children. One explanation could be that the correlations simply reflect task demands. Thus, the ToM batteries for 2- and 3-year-old children mainly included tests of implicit understanding. Two-year old children received tasks which they were able to solve exclusively non-verbally. They were always prompted to perform an action and by this, revealed their understanding of the other's mental state. Three-year-old children were partially able to solve the task non-verbally. Only two tasks required a very rudimentary vocabulary. In the intention-understanding task, children had to indicate one of two buckets by naming the colors blue and red, and for the false-belief unexpected-content tasks, they had to state explicitly what they think the other character suspects is in a box. The verbal demands of both age groups are verifiably lower than the demands for the 4-year-old group, which could account for the missing link. Notably, this line of argument is based on the different requirements on productive language abilities.

However, receptive language abilities were required in the tasks of all age groups. Thus, an alternative explanation for the findings must be taken into account. Taking previous results into consideration, which indicated a positive relationship between both concepts in studies of longitudinal and training designs, a promoting effect of language on ToM might exist in a developmental perspective (Astington & Jenkins, 1999; Hale & Tager-Flusberg, 2003; Müller et al., 2012). Similar to the present results, authors of the cited studies could not show cross-sectional correlations before a mean age of 3 years and 7.5 months. Based on the reported findings, I assume that there might be no direct observable relationship between language and ToM abilities at younger preschool age, but the connection gains more importance with increasing age, and children with high levels of language abilities at earlier stages of life may benefit from this advantages later. Sperber and Wilson's (2000) theory about the common features of pragmatics of verbal communication and mental state understanding explained in Section 1.2., contributes to this assumption. It is conceivable that the process of pragmatic interpretation rises in complexity and achieves higher quality over the course of language acquisition. Children start with two-word combinations at 2-years of age, but expand their skills remarkably from 3 to 4 years of age. They reach major progress in vocabulary and in creative language use, they start to apply past tense, and develop an extensive mother tongue (Menyuk, 2000; Menyuk, Liebergott, & Schultz, 1995). Hence, the link between language and ToM should increase with age, when both abilities mature.

If longitudinal studies and correlational data converge, an early promotion of language skills is advised because it could enhance regulative and social-cognitive competences. Regular documentation of children's language development is essential for detecting potential deficiencies at an early stage. If one considers the usual practices at German day-care centers and the provisions of language assessment, a heterogeneous environment appears. The age children are tested at, the diagnostic instruments that are used, and also the consequences that would follow are the responsibility of the government of each federal state. In Thuringia, for instance, there is no provision for a statewide language assessment or a language training at all (Thüringer Ministerium für Bildung, Wissenschaft und Kultur, 2008). If there are no formal requirements that professional caregivers have to fulfill, the observation and support of children's language development as an integral part of the daily routine remains an individual responsibility. This situation, however, demands a high level of attention and sensitivity, if early interventions should be introduced. With respect to the relationship of language to other cognitive competencies described above, the focus on adequate language skills should be of major importance. The present study investigated receptive language abilities, which can be easily trained in various activities. Picture book reading, asking searching questions about pictures, and listening to storybooks are possibilities which can be easily integrated into daily routine. Storytelling and promoting children's participation (e.g., by answering questions or creating the ending of the story) could enhance receptive and productive competences at the same time. Since there is no reliable information about whether specific aspects of language are responsible for advances in ToM or IC, syntax, semantics and pragmatics could be included. Astington and Jenkins (1999) suggested that the syntactic structure of language is the basis for representing false beliefs. Likewise, other studies found a promoting effect of training sentential complements on false-belief understanding for children around 3- and 4-years of age (Hale & Tager-Flusberg, 2003; Lohmann & Tomasello, 2003). Even if emotion understanding is rather considered as a precursor of ToM, it is important to note that training on mental state talk at 3-, 4- and 5-years of age also promoted children's insight to others‘ emotions, with the largest effects at 4 years of age (Gavazzi & Ornaghi, 2011).

As I investigated the concepts of ToM and IC simultaneously in three age groups, I consider it an important aspect to briefly discuss the relationship found between the two concepts. In this study, a positive connection was only detected for 4-year-old children. The link between ToM and IC is still the topic of lively debate. Previous works also showed positive relations between the concepts, but the underlying mechanisms responsible for the connection could not have been clearly proven so far. As described in Section 1.1., advocates of the expressional accounts rely on the idea that tasks, within which both concepts are tested, include similar requirements. More precisely, typical false-belief understanding tasks also demand a high level of IC (Carlson et al., 1998; Russell et al., 1991). Müller et al. (2012) expanded on this argument, as they did not find a direct correlation between IC tasks and a ToM battery of 2-year-old children. They referred to the fact that the tasks included in the ToM battery for the 2-year-old children did not require IC. In a visual perspective taking task, children received a picture and afterwards were asked by an experimenter if they could show him the picture. The task was passed if children took the experimenter's perspective into consideration and turned the picture around. In the second visual perspective taking task, children saw their mothers covering their eyes with both hands. Children received the instruction to show a toy to their mother. They passed the tasked if they showed attempts to unblock their mothers' view. The last task included in the ToM battery for the 2-year-old children was a pretense task. Children simply had to imitate a demonstrated action of pretense performed with small toys by the experimenter. Certainly, the missing inhibitory demands could be one reason for the results, but there is yet another shortcoming to consider. Viewed from a critical perspective, the abilities they tested (i.e., pretense and level 1 visual perspective taking) can be seen as precursors to ToM (Flavell, Everett, Croft, & Flavell, 1981; Leslie, 1987). Our results cannot support the assumption that later occurring relations between ToM and IC are solely due to inhibitory task demands (Perner et al., 2002), otherwise the link would have occurred not only for the group of the 4-year-old children. Missing correlations between the task batteries of 2 and 3 years old children contradict the hypothesis that inhibitory demands cause the connection because a view on the tasks of the ToM batteries reveals indeed a certain level of IC. For example, on the intention-understanding task for 2-year-old children, the impulse of just using the easiest and most efficient way to reach the goal has to be suppressed in favor of operating the seesaw correctly. That is, children had to resist the spontaneous impulse to simply grasp the toy, and instead had to push the slide down to properly use the seesaw. In addition, on the preference-understanding task of the 3-year-old children, the spontaneous impulse to reach for the cup, which was positively highlighted by the experimenter with vocal and facial expressions, had to be overcome to solve the task correctly, since children had to choose the negatively labeled cup to pass the task. To give a last example, false-belief understanding tasks also demand the ability to inhibit behavior, such as the false-belief unexpected-transfer tasks of the 2-year-old children. In the final sequence of this task, the experimenter mistakenly tried to open box A to get the toy, whereas to pass the task, container B needed to be opened, since the target object was transferred from container A to container B in the absence of the experimenter. Thus, children had to suppress the impulse to open the highlighted container A, in favor of opening container B. If commonalties in inhibitory processes are responsible for a positive link between ToM and IC, then the correlations should have been present for all age groups. Until now, investigations of younger preschool children have been rare, but the existing experiments could not show robust correlations between ToM and IC for 2-year-old children and indicated stronger links from 3 years of age onwards (Carlson et al., 2004; Müller et al., 2012). Interestingly, they revealed positive long-term effects, as a well-developed IC at 2 years of age positively predicted ToM at 3 years of age, and a well-developed IC at 3 years of age positively predicted ToM at 4 years of age. As the present study cannot provide longitudinal data, it is not possible to draw causal conclusions at this point. But taking longitudinal findings and the present results together, a functional dependency account is favored. The identification of the cognitive processes, which might explain the relationship, is still an ongoing research topic. For example, Frye, Zelazo, and Palfai (1995) referred to the ability to reason, which is a key element shared by most tasks that ToM and IC are tested with. According to this account, the same structure of reasoning is found in inhibition-of-concept as well as in false-belief understanding tasks. On the inhibition-of-concept task, children have to act according to one dimension while ignoring another dimension. This is similar to the typical false-belief understanding task, where children have to reason from one perspective (i.e., the agent's) while ignoring another perspective (i.e., their own one). Based on the findings of this dissertation project, it is not possible to endorse this theory, since batteries of all age groups contain tasks comparable in design. On the inhibition-of-concept tasks, for instance, children of 2, 3 and 4 years of age had to act according to a certain concept by ignoring the associated concept, and the ToM tasks also followed the principle explained above. If the reasoning structure is responsible for the link, the question remains why there was no correlation found for the batteries of the 2- and 3-year-old children. In summary, the study reported here cannot clarify any causality, but with its novel approach of using a design of tasks appropriate and comparable for three different age groups, it contributes to the clarification of possible underlying mechanisms responsible for links occurring during the preschool years. The basic finding indicates a positive connection between ToM and IC at the age of 4 years. By integrating all four mental states in a battery for 2- and 3-year-old children, it can be suggested that a direct connection between ToM and IC is not present at these early years. However, the present study, with its broad design, provides the foundation for future investigation concerning the relationship of ToM and IC comprehensively on a longitudinal perspective.

The Relationship Between IC and Behavioral Problems

Significant relationships between behavioral problems and IC were revealed for 4-year-old children. In the younger samples of 2- and 3-year-old children no significant links were found. The majority of previous studies examined children from 4 years onwards and demonstrated that children with well-developed IC abilities showed less attention problems, anger, and antisocial or aggressive behaviors (Espy et al., 2011; Hughes et al., 2000; Raaijmakers et al., 2008). For a similar age group, the current dissertation project revealed a robust negative link to teachers' ratings on the Other Problem scale and a negative link on trend level to teachers' scores on externalizing behavioral problems. Children who were better at resisting a strong desire and in controlling dominant impulses, showed less behavioral problems of different nature. Externalizing behavioral problems refer to attention deficits and aggressive behavior. The items clustered on the Other Problem scale resemble behaviors, which are not clearly correlative to specific syndromes. Thus, it is not possible to provide a specific label or classification. To give a better idea of symptoms, it can be emphasized that the 34 items describe maladaptive behaviors, which reflect imbalance in socio-emotional wellbeing such as jealousy, constantly seeking help, or frequent crying. One interpretation of the findings is that a well-developed IC might be one of the necessary features to reacting appropriately in numerous everyday situations. To regulate behavior and emotions in an appropriate manner, the inhibition of responses is an essential component. Barkley (1997) already illustrated the operation of inhibition processes on other executive functions in his hybrid neuropsychological model. According to this model, the ability to inhibit prepotent responses or to stop ongoing responses are not just directly connected to the system of motor control, but also supports the performances of working memory, self-regulation of arousal, motivation and affect, internalization of speech, and reconstitution. Thus, it is plausible to assume that IC affects numerous actions of children's everyday lives and deficits in this area become visible in behavioral problems. If children, for example, are limited in their working memory due to a reduced IC, they are expected to experience difficulties in completing complex tasks or problem-solving. Distraction or any interference may cause serious issues to pursuing goals. Considering the demands of Western society, children are confronted from very early on with high standards of education. Precisely those syndromes of behavioral problems, which are positively linked to IC in the current study, might be a logical response if children may not meet the demands described above. They are conspicuous, for example, by the display of attention deficits or various other problems. If one draws the connection to self-regulation and regulation of arousal, it seems likely that deficits in both domains due to low IC could be apparent by externalizing behavioral problems. A lack of control, reflected in the inability to modulate impulsive expressions, was found to be linked to externalizing behavioral problems (Caspi et al., 1995).

The question is why there was no relationship between IC and behavioral problems detected for the younger age groups. A previous study indicated less behavioral problems for 3-year-old children with high scores on an EF battery (Hughes & Ensor, 2008). The reason for a difference in results could lie in the use of different measurements and in the different methods of data collection used. Hughes and Ensor (2008) created an aggregate score of different behavior questionnaires (Bayley scales of infant development, Bayley, 1993; Goodman, 1997; SSRS, Gresham & Elliott, 1990, SDQ) and integrated data of observations gained by the experimenter from video-based ratings of behavior. The experimenter completed the SSRS and the SDQ during face-to-face interviews with mothers and teachers. Researchers completed the Bayley rating scales. Taken together, both the assessment of behavior ratings on questionnaires and the video-based observations indicate a greater involvement of the experimenter in the data collection compared to the procedure of the current dissertation. It is possible that the closer involvement could have affected the behavior evaluation (e.g., observer effects, response bias) and might be responsible for different results. In the present study, parents and teachers completed the questionnaires independently of the experimenter, because an unaffected evaluation of parents and teachers was considered an important criterion. Furthermore, the focus of the scales included in the aggregate score is to a large extent on externalizing behaviors like ADHD or conduct disorders, thus, behavioral evaluation was more restricted compared to the dimensions included in the current study. Most importantly, the EF battery included only a single IC task among four other tasks examining rule learning, rule shifting and working memory. Thus, IC was only a minor element in the battery of Hughes and Ensor (2008). The present study assessed IC by three different tasks and took paradigms into account, which were comparable for all three age groups. One potential reason as to why there was no relationship between IC and behavioral problems for the 2-and 3-year-old children could be due to altered expectations of adults towards children of different ages. It is likely that expectations of adults on children's regulated expressions and behavioral impulses rise with increasing age of the children (Hughes & Ensor, 2008). This relates to a general increase of demands concerning a behavior, which is adapted to the social expectations of society (e.g., complying with rules and standards of everyday life). Likewise, the social settings and forms of play change over time and expand in terms of the number of persons who are directly involved in interactions and confront each other with different individual needs. The group sizes in daycare increase and consequently the potential for conflicts within these groups also increases. Hence, it seems reasonable that children with weaker IC might be at a disadvantage in terms of dealing with conflicts. Children with little ability to regulate their responses might react deviantly and express their discomfort through different psychological and physiological symptoms (Hughes & Ensor, 2008). An alternative possibility could be that connections become apparent later because IC itself improves remarkably in later preschool ages (Gerstadt et al., 1994). It is conceivable that a guiding relationship only becomes visible when IC becomes more meaningful, specifically when IC actually can be effectively employed to regulate behavior in conflict situations in everyday life or to fulfill the demands in respect of desired behavior rules. If children at 2 and 3 years of age have a generally low level of IC and are not able to apply these rather rudimentary abilities independently in challenging situations, the missing link would be the logical explanation. Since the tasks of the batteries were adjusted to the performance level of the particular age group (i.e., the IC tasks of 2-year-olds had substantially lower requirements than the IC tasks of 4-year-olds), it is not possible to draw conclusions concerning differences in performances between the age groups.

Due to the cross-sectional design of the present study, causal interpretations can only be discussed. It also must be considered that the weak performances on the IC tasks could be the result of behavioral problems. It is also possible that children with behavioral problems might not have the chance to develop a good IC. If children see themselves as confronted by psychological issues and due to this are impaired in behavior, it is conceivable that this is reflected in the impaired regulation of actions and impulses. However, studies with a long-term design argue against this hypothesis. For example, although no cross-sectional correlations between IC and behavioral problems were found for children between 7 and 11 years, clear long-term effects were detected 2 years later (Riggs, Blair, & Greenberg, 2004). Impressive effects of a well-developed IC at the age of 4 and 5 years were identified for later school performance and general social skills (Mischel et al., 1988). Children who were highly capable at delaying gratification at preschool age were later rated as being highly competent in dealing with stress and frustration, were indicated as being less stubborn, and also showed less jealousy (Mischel et al., 1988). Taken together, the current findings strongly suggest that a good IC can have a positive impact on behavior, especially at later preschool age. Notably, the present study revealed only a single significant correlation for various behavioral problems, besides a trend for externalizing behavior problem scores. Whether these relationships would strengthen with increasing age could be provable by repeated measurements of the same sample, and thus forms an important task for future research.

Clarity about underlying working mechanisms could be gained from studies with experimental training designs. Using results of a training of IC as the independent variable would contribute to identifying possible causal relationships between IC and behavioral problems. There are forms of training, which were already used to successfully enhance young children's IC abilities. For instance, a significant improvement in inhibiting a response was achieved if delay times were integrated into the task, or if training on tasks with executive demands was practiced before (Dowsett & Livesey, 2000; Simpson et al., 2012). Children were better able to inhibit their responses if they were interrupted in various ways (Simpson et al., 2012). In a go/no-go task children saw 16 boxes and received the instruction to find the stickers hidden in some of these boxes. A sticker on top of each box indicated whether a sticker was hidden inside. Children learned that a square on top is the sign for a sticker inside the box (go trial), and a triangle is the sign for no sticker inside the box (no-go trial). Simpson et al. (2012) designed three conditions: In the immediate condition, children had to respond as soon as they saw the boxes; in the delay condition, the experimenter waited 2 seconds before placing the signs on the box; and in the distraction-during-delay condition, children were requested to guess in which hand the experimenter was holding the signal before he placed it on the box. As expected, accuracy of performance on no-go trials was higher in the delay and distraction-during-delay condition. A positive effect of practicing inhibitory skills was provided by Dowsett and Livesey's (2000) study. Children of 3 and 4 years of age had to solve a go/no-go task. An apparatus with a red and a blue light released marbles as a reward. Children were instructed to press a bar in order to receive a marble when the red light appeared, and not to react if the blue light appeared. Children were better solving the task if they had the chance to practice the tasks beforehand and furthermore, if they were trained on other tasks requiring executive demands (e.g., a card sorting game). These findings are essential to creating the basis for testing the cause of the discovered link. If IC training would actually lead to reduced behavioral problems, training in IC in early years would be crucial to exploiting the optimal existing potential, since individual differences in IC are stable in development (Kochanska et al., 1996; Miyake & Friedman, 2012), and thus, could have long-term effects. An exemplary effort into this direction is done by a program which practices self-control for children in second and third grade of primary school (PATHS Curriculum, Riggs, Greenberg, Kusché, & Pentz, 2006). Authors aimed to enhance children's IC by providing them with strategies to stop and calm down, to promote alternative thinking, and to articulate problems. The program effectively promoted IC, and moreover IC was identified to mediate the positive effect of the program on behavioral problems. As highlighted in the present study, there are direct relationships between IC and behavioral problems determined already at preschool age, which strongly recommend earlier support for children. One program for training on general EF at preschool age (Tools of the Mind, Bodrova & Leong, 2007) was already evaluated and indicated a benefits for the trained group compared to the untrained control group in passing EF tasks (Diamond, Barnett, Thomas, & Munro, 2007). To set the course for a positive development of children at an early stage, it would be useful to design and implement such programs for German children from 4 years of age onwards to scientifically verify their effectiveness, and to introduce them to best practice at daycares. Facilities for promoting early skills should also be an integral part of the education of professional caregivers.

The Relationship Between ToM and Behavioral Problems

Considering the relationship between ToM and behavioral problems, the findings from the 4-year-old children replicate previous research that suggested that a well-developed mental state understanding is positively linked to forms of behavioral problems (Renouf et al., 2010; Walker, 2005). High ToM scores predicted statistically significant internalizing and various behavioral problems. To provide more detailed information about the problems children showed, it is helpful to consider the syndrome scales for internalizing problems, and the characteristics of the items on the Other Problem scale. The syndrome scales of internalizing problems include maladaptive behaviors like anxiousness and depressions, social withdrawal, emotional reactivity, and somatic complaints. These are all behaviors, which indicate a disturbance in intropunitive emotions and moods (Zahn-Waxler, Klimes-Dougan, & Slattery, 2000). Children develop negative feelings like fear, worry or guilt, which causes them to react within the self, rather than acting in the direction of others. The items on the Other Problem scale also do not indicate tendencies to hurt or offend others. As a sample, items like “seeks often for help”, “picks skin”, “rocks head and body” or “is over-conformed” are included. Previous studies showed relationships of the same positive direction but rather to forms of aggressive behavior, a syndrome assigned to externalizing behavioral problems (Renouf et al., 2010; Walker, 2005). Walker (2005) found a positive link between aggressive behavior and ToM for boys. Three to 5-year-old boys, who were rated by their teachers as aggressive in social interactions with peers, showed good performances on false-belief tasks. She explained her results by referring to a general difference in the way boys and girls may deal with social interactions and reach personal social goals. Walker (2005) suggests that boys might try to reach their social goals with a dominant approach rather than a conciliatory approach. This could explain the positive link between ToM and aggressive behavior for boys, as well as the positive link between ToM and prosocial behavior she found for girls. Further, Renouf et al. (2010) indicated a positive link to indirect aggression (i.e., trying to harm others by using indirect means like talking badly about others behind their backs). It is noteworthy that this finding was only present for children with an average or low level of prosocial behavior, which in this case allows for the assumption that the positive link between ToM and aggressive behavior cannot be generalized. However, even if different syndromes are affected there might be a single explanation, which could account for the findings: Exhibiting behavioral problems, regardless of the specific character, reflects concerns children are trying to cope with. Having a well-developed ToM implies a good understanding about others' thoughts and behaviors, but at the same time it implies also the recognition of discrepancies when others' mental states are incongruent with own mental states. This discrepancy in knowledge or interest could create some difficulties on an interpersonal level that children have to cope with. The finding of the positive link between ToM and internalizing behavioral problems may reflect these coping attempts. If we note from previous results that children with high ToM scores also displayed more negative perceptions of future events, and that estimated caregivers' judgments on own performances were more critical than children with lower ToM scores (Dunn, 1995), it seems likely that recognizing others' mental states and possible divergences may raise uncertainty or fear, which in turn could lead to internalizing behavioral problems. If children evaluate these everyday situations as being threatening or irritating, it could negatively affect their emotional well-being. Indeed, previous literature found a link between experiencing difficulties in emotional responses and internalizing behavioral problems (Jellesma, Rieffe, Terwogt, & Westenberg, 2009). Children in middle childhood showed more internalizing behavioral problems, when they had difficulties in handling negative situations and were not able to define internal feelings properly. The question remains as to why this link appeared only at the age of 4 years and not for the younger sample. Why might older children be affected by recognizing others' mental states but younger children are not? One possibility could be found in the difference between implicit and explicit ToM understanding. An explicit understanding of ToM is often tested by tasks requiring verbal judgments from children and is commonly found from 4 years onwards, whereas younger children show an implicit understanding of ToM (for an overview see Low & Perner, 2012). If children reach another level of knowledge by increasing age, in the sense that they not only represent other people's thoughts and behaviors but are also able to make judgments about them (Clements & Perner, 1994), it is conceivable that only this higher level of knowledge might affect children's behavior and thus, relationships become more apparent from 4 years of age onwards. As a further option, it is possible that a critical self-reflective perspective is responsible for a raise in uncertainty and irritation, when noticing incongruent mental states. Along with the development of mental representational capacities in later preschool age, self-evaluative processes evolve and children start to develop a motivation to appraise themselves in relation to others and to seek social approval (Higgins, 1989). The missing correlation between ToM and behavioral problems within the younger samples could strengthen this assumption. If self-evaluative processes are not well-developed by 2 and 3 years of age and the wish for social approval and affiliation is not yet that intense, the perception of divergent mental states may not elicit discomfort with which children have to cope. Since this line of thought cannot be answered by the study, this perspective surely deserves attention in future research. However, the assumption that children are more focused on similarities between others and themselves by increasing age was already raised by studies on over-imitation. With increasing age, children, and even adults, readily imitate irrelevant actions of a model, which are not necessary for achieving a goal (McGuigan, Makinson, & Whiten, 2011). In the aforementioned study, participants imitated the causally irrelevant actions of a model to retrieve a reward from an apparatus with a significant age effect. Furthermore, 4- to 5 year-old children chose to imitate irrelevant actions only if the model, which presented the irrelevant actions, was present. If this model left the room, they did not imitate the irrelevant actions, but instead acted in a rational manner to achieve the goal (Nielsen & Blank, 2011). Authors explained this behavior by humans' natural tendencies to affiliate with others. Since imitation might serve this function (Uzgiris, 1981) it is possible that the older children of the current thesis reflect the mental states of others more critically than the younger children, care about them on another level, and consequently are trying to cope with them when realizing less matches.

The fact that the results of Hughes and Ensor (2007) showed direct correlations between ToM and behavioral problems for 2-year-old children appears puzzling. Though their design varied to large extent from our design concerning ToM assessment, they used a pretense, deception and mistaken-belief task for the ToM battery. The majority of these tasks can be considered as measurements for precursors of ToM (Leslie, 1987). The method the authors applied to evaluate behavioral problems also makes comparison to the present study difficult, because they combined different scales of three different questionnaires (Bayley scales of infant development, Bayley, 1993; Goodman, 1997, SSRS, Gresham & Elliott, 1990, see also the prior section) and additionally included observational behavior ratings to create an overall aggregate score. On a whole, these comprehensive investigations exclusively focus on externalizing behavioral problems. It is feasible that the use of four different behavior measurements and the application of tasks measuring precursors of ToM contributed to the different outcome. For the younger age groups of the present sample, there was no evidence for a link to aggressive behavior, which means precisely that the ToM of 2- and 3-year old children was unrelated to externalizing behavioral problems. It should be noted, however, that statistically significant negative correlations were found between ToM and parents' ratings on externalizing behavioral problems for the 4-year-old children on the first analyses. Partial correlations, implemented to control for language and gender, weakened this result to a large extent and the regression model finally turned out not to be statistically significant. Existing literature demonstrates that positive correlations to aggression were typically found for children whose ToM was tested at around 5 years of age (Capage & Watson, 2001; Lane et al., 2013; Wellman et al., 2011). As outlined in the theoretical background of this thesis, the majority of studies which included younger children could not show these relations (Hughes et al., 2000; Slaughter et al., 2002; Yiwen et al., 2004). Here again, the conclusion can be drawn that relationships between both variables attain more importance with increasing age.

The present results indicate that the relationship between social-cognitive skills and behavior cannot be simplified to a negative relationship, which would indicate that better cognitive skills would occur with fewer behavioral problems. Other theories also questioned this approach, since connections between a well-developed ToM and indirect aggression or bullying in late childhood and adolescent were found (Gini, 2006; Sutton, Smith, & Swettenham, 2001). Authors explained these findings by referring to the Machiavellian intelligence theory (Repacholi, Slaughter, Pritchard, & & Gibbs, 2003; Wilson, Near, & Miller, 1996), which states that humans who follow a general tendency to harm others, have to possess capabilities to estimate the impact of their actions on the others in advance. This includes the capacity to read the thoughts and emotions of others, besides a lower level of moral concerns. The theory of Machiavellianism says that people with these personality characteristics have the tendency to manipulate others in order to pursue their own interest (Wilson et al., 1996). Thus, it could be assumed that a person’s social knowledge might be connected with a person’s social functioning. However, investigations on the direct relationship between ToM and Machiavellianism at preschool and preadolescent age did not indicate a relationship (Repacholi et al., 2003). Findings of this thesis go in line with the above mentioned studies, since results only delivered evidence that children with internalizing and other behavioral problems showed a better ToM, and these syndromes do not aim to harm others.

Referring to the present findings, training solely on ToM might not lead to fewer behavioral problems, because a negative relationship is missing. The positive link between ToM and internalizing behavioral problems demand a sensitive and attentive attitude of caretakers and parents towards their children. Preschoolers with a high understanding of others' mental states might be at risk for internalizing and various other problems. In turn, internalizing problems hold the risk of persistent negative consequences in different situations of life. For example, they can negatively affect school life, as they have proven to be linked to poor academic performance (Lundy, Silva, Kaemingk, Goodwin, & Quan, 2010). In terms of a preventative approach, it could be helpful to make divergent mental states a regular subject of discussion and to provide helpful strategies to deal with them. Basically, it would be essential to investigate if children at the age of 4 years lack strategies for coping with divergences constructively. Interview measures to assess coping strategies are mainly provided for children from primary school age onwards (for a summery see Compas, Connor-Smith, Saltzman, Thomsen, & Wadsworth, 2001). Longitudinal data of children from 4 to 12 years revealed that instrumental coping (to act constructively to improve the situation) and positive cognitive restructuring (to think in a positive way about a problem) related to conflict situations among peers, reported by teachers increases over time (Losoya, Eisenberg, & Fabes, 1998). This indicates that children at 4 years of age may not require appropriate strategies to cope with interpersonal problems but to expand their abilities later on. To support children by introducing constructive coping strategies would be an essential part of elementary education. If one considers that 10-year-old children still naturally favor the strategy of seeking adult counsel (Burgess, Wojslawowicz, Rubin, Rose-Krasnor, & Booth-

LaForce, 2006), it is obvious that preschool children need special support, especially children with the disposition towards avoidance and withdrawn stances who may benefit from these strategies.

No links between the temperament dimension scales to ToM or to any of the dependent variables (C-TRF Other Problems scale and C-TRF internalizing score) were found. Since the basic features of temperament assessed for the present sample represent only a part of the personality, it would be interesting to investigate personality characteristics like neuroticism and prosocial tendencies to control if certain characteristics, which reflect the way children are dealing with interpersonal matters, moderate the relationship between ToM and internalizing behavioral problems. The temperament dimension scales used in this study do not include sociability and do not provide a comprehensive profile of neuroticism, which are determinants that are heavily involved in the maintenance of balanced emotional relationships. Besides anger and irritability, neuroticism also encompasses fear and anxiety (Shiner & Caspi, 2003). The latter are not specifically measured by the IKT, but it taps inner-directed negative emotions, thus reflecting a general sensitivity and emotionally vulnerability in conflict situations. The investigation of these concepts in relation to children's ToM and internalizing behavioral problems may contribute to the identification of children at higher risk, who should receive special attention for possible prevention programs.

An alternative explanation for the correlation of the cross-sectional data should be considered as well. It cannot be ruled out that children who show higher internalizing behavioral problems have a better chance to develop ToM abilities. This alternative approach could be assumed if one argues that children who show more introversive behavior like being withdrawn, shy or fearful may have more chances to observe other people from a different or wider perspective, and by this gain a better insight into their minds. This alternative option raises the question as to whether the results would favor the emotional-reactive hypothesis

(Hare & Tomasello, 2005), which claims that a shy and withdrawn stance to other people could foster ToM abilities. One argument that contradicts this assumption can be found in the characteristics of the dimension scales summarized in the internalizing problem scores. The items of the withdrawn scale are clearly distinctive to shyness. They do not refer to shy reactions to new people and unfamiliar situations but rather indicate avoidance and social withdrawal in everyday life (e.g., „shows little affection“, „is apathetic“, „avoids eye contact“). From a theoretical point of view, the idea that internalizing behavioral problems should promote children's ToM abilities by the age of 4 years in a positive way is difficult to justify. It is hard to believe that internalizing syndromes like emotional reactiveness, anxiousness and depression, or the avoidance of social interactions could draw a children's attention to other people's mind and promote the understanding of them because these children are dealing with intrapersonal problems, directed at themselves. Hence, it seems unlikely that they direct their attention to others to a great extent.

Since robust findings were only revealed for caregivers' ratings and correlations among parents' and caregivers' ratings were missing for the computed scales, I will discuss the issue of differences between informants' reports in the following section.

Caregivers' and Parents' Judgments

ToM and IC only predicted behavioral problems of 4-year-old children with respect to caregivers' ratings. No robust link was found for parents' ratings on the equivalent questionnaire. Positive correlations among the major scales of interest[4] on caregivers' and parents' questionnaires were obtained for 2-year-old children, but not for the sample of 3- and 4-year-old children. The correlations of parents' and caregivers' judgments for the youngest sample on the one hand, and the deviating judgments for the older samples on the

[4] The major scales of interest include the internalizing and externalizing scores, as well as the score of the Other Problem scale.

other hand, would have been predicted by the group socialization theory of development (Harris, 1995). A person's behavior is dominantly shaped through the social environment outside of the home, and personal characteristics that are inherent to a child are permanently modified by the group context. Different behaviors of caregivers or parents and the dynamics of these interactions influence children's behavior even in early years (Harris, 1995; Rubenstein & Howes, 1979). The time children spend in daycare centers increases by age, therefore, a manifestation of context-specific behavior is likely. Research on children's cortisol levels repeatedly showed an increase in cortisol level from morning to afternoon at daycare centers, whereas on days spent at home this effect could not be revealed. In addition, the overall cortisol level was lower at home compared to daycare centers (Dettling, Gunnar, & Donzella, 1999; Watamura, Donzella, Alwin, & Gunnar, 2003). A connection to behavior is given, as cortisol level was found to be positively related to aggressive behavior and poor self-control (Dettling et al., 1999). These findings allow for the conclusion that children behave in a context-specific manner and that this leads to deviating judgments of parents and caregivers. When the mean difference of scores on the major scales of interest was analyzed between both groups of informants, significant differences among caregivers' and parents' ratings were detected across all age groups. However, the revealed deviations do not allow for a clear conclusion about a homogeneous pattern. Lower mean caregivers' ratings compared to parents' ratings were found for 4-year-olds, but the opposite pattern was found for 3-year-olds, and patterns for the judgments of 2-year-olds even varied between different scales. It is not indicated that parents, or caregivers respectively, in general tend to judge children more strictly, but the findings substantiate the fact that informants' judgments do not appear homogenous (see also Miner & Clarke-Stewart, 2008), which should be considered for future study designs.

An alternative explanation could also account for the differences. A prolonged period of time spent in daycare centers could increase the possibility that caregivers of 3- and 4-year-old children have more experience with the children during everyday activities and with regard to inter-peer interactions than parents have. This could facilitate a broader picture of behavior and allow for comparison. In contrast, the parents and caregivers tend to spend a more equal amount of time with the child at 2 years of age. It is likely that caregivers have greater experience in professional evaluation processes and better comparison to other children of the same age. Advantages in terms of experience and objectivity would suggest that caregivers provide more reliable information, and by this delivered stronger relationships between ToM, IC and behavioral problems for the sample of the 4-year-olds. Although I can provide only assumptions which may account for the incongruence in the results, this difference in predictive values highlights the importance of a multi-informant design when investigating children's behavior.

Temperament as a Possible Correlate

Regarding temperamental factors, only a single link was detected, which was found between activity and IC for 2-year-olds. Children with low levels of activity were better at inhibiting their own actions and thoughts, and could resist desired sweets for longer. As temperament is considered as a set of relatively stable early-developing characteristics (Rothbart & Bates, 1998), one might assume that high activity levels hinder children from controlling their impulses. However, this relationship was not found for 3- and 4-year-olds, which indicate that temperamental factors influence children's inhibitory skills only early in life, but may lose in strength with the maturation of the frontal cortex, which causes remarkable improvement of inhibitory skills between the ages of 3 to 7 years (Diamond, 2002). When controlling for other influences, a proportional shift occurred and only language turned out to explain a significant proportion of the variance of the IC scores. Surprisingly no correlation between attention and IC was found for any of the age groups. As reported in Section 1.2, previous studies revealed a positive connection between attention focusing and IC tasks (Gerardi-Caulton, 2000; Wolfe & Bell, 2004). A crucial difference between the current investigation and these studies lies in the study design. The cited authors used a single paradigm for assessing IC and decided either for a spatial-inhibition task or for conflict-inhibition tasks. The present thesis used both paradigms and complemented them by adding a delay-of-gratification task to the IC composite score. The waiting paradigm of the delay-of-gratification task did not require children's ability to focus attention and even facilitated children to pass the task when they were able to distract themselves. Nevertheless, it is an important task for measuring the ability to inhibit a dominant response and by this, is an essential part of investigating IC comprehensively. Considering the findings for all temperament dimensions included in the present study, there is no evidence for a direct link between temperament and IC at preschool age.

The present results cannot confirm the suggested positive link between a shy temperament and children's ToM abilities for the samples of 2-, 3- and 4-year-old children (Lane et al., 2013; Mink et al., 2014; Wellman et al., 2011). No links were detected between behavioral problems and the dimension scale Social Inhibition/Shyness, which refers to the inhibition of behavior in new and unfamiliar situations that is grounded in feelings of insecurity but not fear and is comparable to the shyness scale of the CBQ (Rothbart, Ahadi, Hershey, & Fisher, 2001) ($r = .81$). A major difference between the recent studies and the present study lies in the temperament measurements themselves. In past studies, different syndrome scales and generated composite scores from of two different questionnaires were used (Lane et al., 2013; Wellman et al., 2011). Although the authors report their findings in the sense of temperament traits, I have to highlight that the composite score was calculated from the shyness scale of the CBQ (Rothbart et al., 2001) but also from the withdrawn scale of the CBCL/2-3 (Achenbach, 1992). The combination of items from a questionnaire to assess temperament and items from a questionnaire to assess behavioral problems was justified by the high correlativity of the items. Strictly speaking, it is not advisable to use psychopathology criteria to assess temperament. Temperament traits rather indicate a wide range of behavior styles among normal developing children (Chess & Thomas, 1996). Temperament dimensions refer for instance, to a general shy stance towards unknown situations or people and describe children's first reactions to them. Withdrawn, in terms of internalizing behavioral problems, however, refers to the general withdrawal into the self, which is not related to specific situations. Thus, the syndrome scale includes items like „is apathetic and unmotivated", „is unresponsive to affection", or „avoids eye contact". Correlations among variables may indicate a connection and it is possible that behavior is elicited and shaped by temperament characteristics (Rothbart & Bates, 1998), but with a critical view on the items it cannot be assumed that correlations deliver evidence that identical concepts are measured. With a look at the positive link between ratings on this composite score and ToM in the aforementioned studies, these findings appear relatively close to the present finding for internalizing behavioral problems and ToM. As I used a questionnaire that specifically measures temperament dimensions and could not reveal a single correlation, it is proposed that the link between ToM and temperament might not be as clear-cut as previously suggested. It is quite possible that early temperament might influence children's mental state understanding in the sense of later occurring long-term effects. Accordingly, Mink et al. (2014) confirmed this hypothesis, but also could not reveal significant links for cross-sectional data of 3-year-old children. To clarify the assumption that temperament might predispose children for later ToM development, a repeated data collection of the current sample of 2-year-old children at 3 and 4 years of age would be suggested.

Even though it was not the main concern of this dissertation project, I want to briefly discuss the correlations between temperament and behavioral problems. Previous studies suggest that temperament characteristics are reflected in behavioral problems (Rothbart, 2011; Rothbart & Bates, 1998). For every age group in the current investigation, links between certain temperament dimensions and behavioral problems were revealed. The correlations among parents‘ ratings for all three age groups indicate that more socially-inhibited children showed a high level of internalizing behavioral problems, children with the tendency to get easily frustrated showed a high level of externalizing behavioral problems, children with a high sensory sensitivity showed several internalizing behavioral problems, and children with a lower attention level displayed a high amount of external behavioral problems. At first view, the patterns found here suggest that the previous findings were confirmed, but I have to highlight two aspects that do not allow a simple confirmation of suspected patterns. Firstly, cross-sectional data was assessed, which provide no information about developmental courses. Secondly, a large part of these relationships is shown solely in terms of the parents' ratings between the IKT and CBCL, and hardly between the assessments of parents' and caregivers' judgments on the IKT and C-TRF. Therefore, the scales often only correlated for the ratings of one and the same person. A detailed consideration of the items revealed high commonalities between the temperament dimension scales Attention and Frustration and the scales for externalizing behavioral problems. One might suspect that the correlations of parental assessment may be due to the very similar items on the particular scales of the questionnaires. However, the dimension scales Sensory Sensitivity and Social Inhibition/Shyness bear little resemblance to items of the syndrome scales of the CBCL, but still both showed positive links to the CBCL internalizing scores.

Taken together, the outcome of this dissertation project surely contributes to the topic of potential impacts of preschool children's early temperament and could serve as the basis to explore developmental trajectories in a longitudinal design.

Graphical Representation of the Current Model of Correlations

This interim summary graphically displays the correlations found for all age groups before controlling for language. One major aspect, which is visible in the figures below, is that the number of correlations among the variables of interest clearly increases with increasing age. As discussed comprehensively in the sections above, the relationship between ToM, IC and behavioral problems only became apparent at the age of 4 years. The three variables did not correlate in the early years of childhood. The same conclusion can be drawn for the relationship between language and ToM. However, it is important to note that missing cross-sectional correlations do not exclude the possibility of existing long-term effects over the span of age range observed in this study. A remarkably constant relationship between language and IC is shown from 2 years onwards, which supports the assumption of a close connection between language development and regulation processes (Martin-Rhee & Bialystok, 2008). The relationship found between certain dimensions of child temperament, behavioral problems and social-cognitive constructs have already been discussed in the section above. Even if a correlation between activity and IC appeared at 2 years of age, this connection weakened after controlling for language.

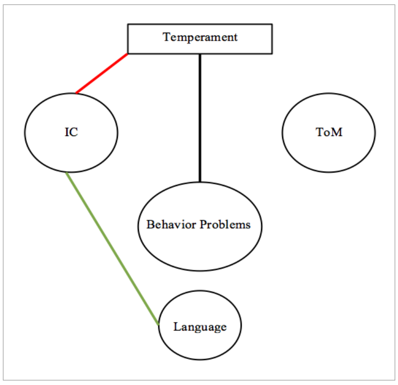


Figure 3 – 2 years of age

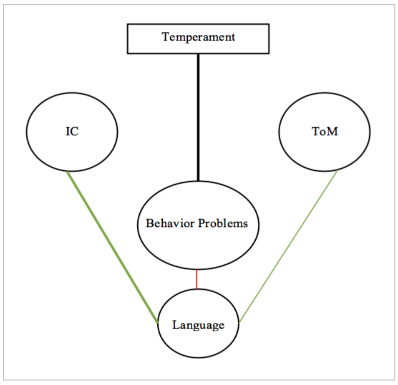


Figure 4 – 3 years of age

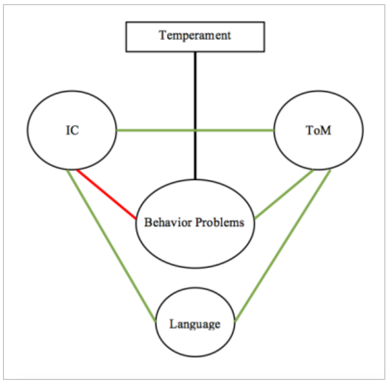


Figure 5 – 4 years of age

Figure 3 – 5. Graphical model of correlations between ToM, IC, behavioral problems, language and temperament for children at 2 years of age (left), 3 years of age (middle) and 4 years of age (right), (green line: positive correlation; red line: negative correlation; black line: positive and negative correlation; thin line: correlation at trend level).

Remarks About Test-Retest Reliability and Task Construction

Before I discuss the relationships between children's social-cognitive, inhibitory abilities and their behavior, I will turn to some particularities concerning the reliability of measurements and special issues of task construction. For all measures included, I assessed test-retest reliability, which is a significant advantage over previous studies. Significant correlations were found for the majority of tasks and for every age group. Moderate and small correlations appeared for the minority of tasks and could be explained by the small sample size. The duration of one test session was approximately 45 minutes. To avoid an overload of children's capacity of concentration, it was only possible to repeat a single first session task at the second session in order to assess the test-retest reliability of this task. Thus, the sample sizes for test-retest reliability ranged only between four and 11 children per task, which surely weakened the results. Notably, all of the false-belief tasks have to be considered carefully concerning re-test reliability, because they involved an unexpected turn in their design, which was recalled by some children at the retest. For instance, on the false-belief unexpected-content tasks of 3- and 4-year-olds, pencils were found in a well-known box of chocolate instead of the sweets, or in the false-belief unexpected-transfer task of the 3-year-olds the assistant played a trick on the experimenter and secretly switched the location of an object in her absence. The design of the false-belief unexpected-content task of 3- and 4-year-olds required children to explicitly state their belief about the content of the well-known box of chocolate. It turned out that most of the children remembered the unexpected turn of the task on the second session and consequently said they actually expected pencils in the box. If this outcome occurred, the task could not be performed. This explicit question was not involved in all other false-belief tasks, hence, it is likely that the memory of the unexpected turn have affected the decision children made when they received the task for the second time, as there were no or only very small correlation coefficients revealed for the test-retest reliability of these tasks. Taken together, for the majority of tasks, test-retest reliability was confirmed. Due to the limitations described above, findings regarding the reliability of the false-belief tasks should be considered cautiously. However, it is an important first step to strengthen research quality criteria. For future study designs, capacities should be expanded to control as accurately as possible for the reliability of the data.

Besides reliability, certain concerns about test validity remain. Thus, a minor amount of tasks will be discussed with respect to their designs. To start with, the distribution of the correct answers on one of the spatial-inhibition tasks appears unusual for a task which measures IC, since children achieved better results on the second block of trials compared to the first block of trials. Regarding the task design of the spatial-inhibition task of 3- and 4-year-old children (i.e., the standard version of the windows task), the question may arise as to whether this task actually measures children's IC, or if it measures children's ability to infer a rule. It appears puzzling that performances of 4-year-olds improved with an increasing amount of trials, which was not the case for 3-year-olds. Simpson, Riggs, and Simon (2004), for example, argue that the standard windows task does not challenge children because of inhibitory demands. They used a modified version for comparison, in which an alternative instruction eliminated the demand to infer the rule (i.e., point to the empty box for receiving the reward) and found that children passed this alternative version significantly better than the standard task. However, a major shortcoming of this version is that the direct instruction 'to point to the empty box' leaves the box with the reward unattended (Simpson et al., 2004), draws the attention exclusively to the empty box, and by this, not only the demand to infer a rule is eliminated, the ability to inhibit and suppress a dominant response is also eliminated. Thus, the original version of the windows task seems to be more appropriate for measuring IC than this modified version. The authors further argue that children might just not be able to discern what to do in the original version of the windows task, which would mean they simply do not understand the rule of the game (Simpson et al., 2004). This indeed could be a possible explanation for weak results. Importantly, to exclude this possibility, I included a training phase and only assessed the task if children passed the training phase by applying the correct rule at least three times. By doing so, I ensured that children understood the rule of the game. Other findings of Carlson, and Moses et al. (1998) contribute to the assumption that the windows task actually measures inhibitory control, by comparing two version of a deceptive pointing tasks, which had a similar structure like the windows task. In both versions of the task, children were sitting in front of two boxes. One box was baited and the other box was empty. An experimenter was naïve concerning the location of the bait. Children were instructed to deceive the experimenter when they are asked about the location of the bait, hence, they had to point to the empty box instead of the baited box. In the first version, children had to point spontaneously with their fingers. In the modified version, children had to point with a large board-game arrow instead of spontaneously pointing with their finger. Children performed better in the modified task compared to the standard task, which leads to the conclusion that the interruption while handling the large board-game arrow helped children to inhibit the prepotent response. This interpretation is further strengthened by findings, which showed that children performed better on IC tasks when receiving a delay time before responding (Simpson et al., 2012). Taken together, the windows task is strongly suspected to measure IC, although a certain level of rule use and working memory is required for it as well. However, despite the wide agreement among research studies, it remains important to keep a critical perspective for interpreting the performance and for discussing the suspected cognitive processes underlying tasks (Zelazo et al., 2003).

All tasks of the ToM batteries were designed to measure children's ability to infer others' mental states. More precisely, others' intentions, preferences or beliefs which were not directly observable and thereby, had to be inferred. In the preference task for 2-year-olds, for example, an actor demonstrated to the children facial expressions of delight and dislike, each towards one of two objects. In the following, children were requested to choose one of the objects to hand to the actor, and by this, had to infer the preference of the actor for an object. To give another example, in the intention-understanding task for 3- and 4-year-olds, children listened to an action plan of an actor and from this they had to infer his intention. An exception, which differs from this task design, is represented by the goal-understanding tasks of all age groups and the intention-understanding task for the 2-year-olds. In these tasks, children were able to observe the actor's goal, respectively his intention, directly by his actions. The final location in the goal-understanding task (i.e., the location reached by the experimenter's hand) was visible to the children. Only the saliency of the final location varied. In the intention-understanding task for the 2-year-olds, the intention of the actor (the action plan and the goal) was also observable. The possibility of observing the goal or the intention of an actor directly, could lead to the assumption that children did not necessarily had to infer the goal or the intention in the sense of an actual mental state. However, imitation requires a certain level of interpretation, since children copy actions in terms of what they think the other person is trying to achieve (Bekkering et al., 2000; Gergely et al., 2002). This cognitive process entails reasoning about the other's mind. Thus, one could argue that children do infer the actor's goal or intention as soon as they are representing it. Evidence that supports the theory of an interpretative act for the intention-understanding task is given by the original study of Buttelmann et al. (2008). Fourteen-month-old children were presented with two different conditions. In one condition the actor was constrained. He saw a reward in a box and wanted to obtain this reward. However, the box was blocked and he was not able to simply take the reward out of the box. Instead, he was forced to use a mounted seesaw. In another condition the actor was unconstrained. The box was not blocked, the actor could have easily grasped the reward, but decided to use the mounted seesaw. Thereby, he freely chose to use the seesaw. Differences in performance revealed that children imitated the use of the seesaw significantly more often in the unconstrained condition compared to the constrained condition, indicating they actually inferred the actor's action as a significant part of his intention. Even if concerns about the design of the intention-understanding task can be eliminated, the deviating task design of the goal-understanding tasks from the other tasks of the ToM battery could be considered as a shortcoming of the study. As previous research does not provide goal-understanding paradigms for children between 2 and 4 years of age, in which the actor's goal is not directly observable, the creation of novel tasks is strongly recommended for future investigations.

In general, all tasks were chosen with respect to results children achieved on previous studies. The aim was to assign children to tasks of an appropriate level of difficulty, to obtain some variance for a successful correlation. In accordance with this, results across all age groups showed indeed a wide range of individual differences. Some children were found to be able to pass the tasks completely, whereas other children of the same age group failed the tasks. Besides the increase in performance within the spatial-inhibition task for the 4-year-olds discussed above, I found for the majority of the IC tasks the typical effect of decreasing performance with ongoing trials. This pattern is expected (see also Gerstadt et al., 1994), because IC can be considered as a cognitive resource, which requires strength to maintain and gets weaker through prolonged effort. The inhibition-of-concept tasks for 3- and 4-year-olds displayed this pattern of stronger performances on the first half of the trials compared to the performance of the second half of the trials, indicating that children were losing strength of inhibition over the course of the task. For the youngest age group there was no decrease of performance revealed. This could be grounded in the fact that on this task a large number of children performed exactly at chance level (24.7% of children scored at the chance level of 50%). Reaching chance level on this task could have been the result of just sorting all blocks in one of the buckets. However, only eight children showed this sorting behavior. The explanation for this finding is very speculative, even if one could assume that these children did not follow the rule seriously. Regardless of this speculation, the high amount of children who performed exactly at chance level might be responsible for the missing effect of a decrease in performance between the first and the second block of trials analyzed for the entire sample size.

Surprisingly, results of the false-belief-understanding tasks for 2- and 3-year-old children deviated more than expected from the results of previous studies. The false-belief unexpected-content task of the 2-year-olds, originally assessed by Buttelmann et al. (2014), was passed by 18-month-old infants, when comparing results of a false-belief condition to results of a true-belief condition. The children in our sample only received the false-belief condition and did not pass the task significantly, but instead performed at chance level. As the design of the task did not differ, the explanation might lie in children's relationship to the test materials. Similarly to the original version, for the target object a block was used and for the unexpected object a spoon was used. It is possible that 2-year-old children are more distracted by a spoon than 18-month-old infants. With increasing age, the independent use of a spoon increases and 2 -year -old children could be more focused on the spoon when instructed to bring an object for the experimenter, compared to 18-month-old infants. For passing the task, children have to read the other's mind, but at the same time have to direct their attention from the unexpected object to the target object. There is a chance that the spoon was attracting high attention from 2-year-olds and, consequently, challenged the children to a higher extent to pass the task than the younger children of the original study. Although random statements from the parents are highly subjective and not reliable for this study, it is worth mentioning that a certain number of parents reported that spoons are of special interest to their child "at the moment". For future investigations I advise the use of more neutral objects as study material to reduce chances of bias. Furthermore, it is important to mention the false-belief unexpected-transfer task for the 2-year-olds, originally assessed by Buttelmann et al. (2009) and passed by 18-month-old infants. In our sample, children failed the task significantly, which might be explained by a slightly modified task design described in the method section. The version used for this study contained the priming of one of the boxes. At the beginning of the task, children saw the experimenter pulling on the lid of one of the boxes three times in a row (box A), remaining unsuccessful with opening box A. Right afterwards children were instructed to open box A for the experimenter, hence, the experimenter was able to place the toy into box A. In this way, children already had contact with box A and assisted the experimenter before the actual test phase. Importantly, box A is the incorrect box in the later test phase, since the toy was switched from box A to box B in the absence of the experimenter and the children had to infer her false belief. This modification, the priming of box A, is likely responsible for the decrease in children's success. Finally, it is important to address the weak results of the false-belief unexpected-transfer task for the 3-year-olds, the task for which was adopted from Southgate et al. (2010)

and originally passed by 17-month-olds. In the present study, 3-year-olds failed the task significantly. These findings remain puzzling because the procedure of this study is largely identical to the original version. Additionally, I assessed children's own preference for an object after finishing the task and found no influence of children's own preference on their performances, which could have contributed for an explanation of the weak performance. However, the study material could still account for the present results. Since the attractiveness of the study material could influence children's performance to a negative direction (Buttelmann & Buttelmann, 2013, September), it could be possible that the novel objects we used were simply too boring for the children. The distribution of the results of all other tasks turned out as expected and are likely derived from the fact that mean ages differ between the original studies and the present investigations.

When analyzing the single task results with respect to the session-order, only a minor number of effects were found. Children from the 2-year-old sample performed better on the inhibition-of-concept task, when solving this on the second session compared to children, who solved the task on the first session. Since, firstly, the same experimenter was involved in the second session; secondly, children received the same length of time for getting familiar with the experimenter during the warm-up phase as on the first session; and thirdly, no other task for the sample of the 2-year-olds resulted in a statistically significant difference, an explanation for this finding is not clear. A vague possibility could be that children were more comfortable with the testing situation at the second session in general and, thus, were more relaxed and willing to follow the task instructions. Since high concentration is required in this particular task and the entertainment factor is relatively low compared to all other tasks, children who received this task at the second session might have had an advantage because they were already familiar with the situation to follow instructions by the experimenter. Other session-order effects were only found for the sample of 4-year-olds. Children who completed the goal-understanding task at the first session performed better compared with children who completed the task at the second session, and children who completed the delay-of-gratification task on the first session performed worse compared with children who completed the task on the second session. It is difficult to find an explanation for the first reported finding, whereas the latter finding might be explained in a similar manner to the effect for the 2-year-olds. If children developed a more relaxed mood within the second test session, they might have been less tense and better able to wait for the reward. As an alternative explanation, it could also be possible that children built up greater trust in the promised return of the experimenter, because they already knew the experimenter from the first session, hence, were more convinced that the experimenter would really return soon. Funder, Block, and Block (1983) already indicated a link between children's characteristics like a general calm and relaxed mood and their ability to delay gratification. Nevertheless, these are only attempts to interpret the session-order effects. Compared to the large amount of single tasks I analyzed, only a very small proportion showed statistically significant results concerning effects of session order. Analyzing the IC and ToM scores for effect of session order, statistically significant differences were not present. Thus, it can be assumed that the order of the sessions did not influence children's performance.

Gender Differences and Internal Consistency of Tasks

To complete the discussion concerning the single tasks, I briefly refer to gender differences and internal consistency. The only gender difference was detected for the sample of the 2-year-old children. Girls passed the false-belief unexpected-content task better than boys did. Results reported by previous research are mixed. On the one hand, gender effects indicated an advantage for girls on ToM tasks (Carlson & Moses, 2001; Cutting & Dunn, 1999), but on the other hand not all previous studies could reveal significant gender effects

(Hughes & Cutting, 1999; Mink et al., 2014). Since the present dissertation project provides representative data by large sample sizes of three different age groups, the idea that there might only be a marginal advantage for girls outperforming boys on ToM is supported, since this difference did not appear for the majority of tasks, nor was it detected for different age groups. Consequently, it cannot be called a clear-cut finding. Investigations exclusively with regards to gender differences were done by Charman, Ruffman, and Clements (2002). They provided two representational samples, assessed independently in different laboratories, and revealed only weak effects consistently indicating an advantage for girls. A view on previous findings concerning gender differences on IC tasks delivered the same inconsistency as described for ToM tasks. Some studies found girls outperforming boys on several tasks (Carlson & Moses, 2001; Kochanska, Murray, & Coy, 1997), whereas other authors only report marginally advantages for girls (Carlson & Wang, 2007), or no differences at all (Apperly & Carroll, 2009; Gerstadt et al., 1994). In summary, as the representative sample of three age groups did not indicate any difference between girls and boys for IC performances and only a single effect related to ToM performances, I conclude there are no broad gender differences for mental state understanding and inhibition in preschool years that should alert major concern for a different treatment of boys and girls in terms of the promotion of competences.

An incidental finding from the study is of interest for further investigation of children's cognitive development. The results of the analysis concerning internal consistency of the batteries encourage the presumption that the different mental states map various concepts, which might form a unitary construct only later with progressive maturation. This perspective would support the positive gain accounts of ToM (Apperly & Butterfill, 2009; Low, 2010), which expect task-specific continuity rather than an overall coherence. In the same sense, Thoermer et al. (2012) found in a longitudinal study task-specific developmental relations among false-belief understanding, but no relations between false-belief and desire understanding. Referring to the comparison of mean scores of the single ToM tasks, the question can be addressed as to whether the single mental states are of varying difficulty. For the 3- and 4-year-old sample, mean rank comparison revealed that children gained better results on goal understanding compared to all other mental states and that intention understanding received the second-highest ranking. Notably, children received the lowest mean scores for false-belief understanding. The fact that the same pattern was found for two batteries and samples would rather argue for differences in conceptual content than in task material and in design. The findings reflect the theory by Tomasello et al. (2005), who claim that goal- and intention-understanding emerges earlier than belief-understanding in the sense of developmental pathways. Other investigations detected differences in difficulty, mainly that the majority of children from 3 years onwards revealed worse belief-understanding than desire-understanding, however this pattern gained stability by 4 and 5 years of age (Wellman & Liu, 2004). Surprisingly, the present data showed no such pattern for the youngest sample, which remains puzzling. The question about the competences these tasks further measured could be posed. Although a reliable answer can hardly be provided, I would like to emphasize that competences, which are included to some extent (i.e., language abilities and IC), were assessed simultaneously and were not related to ToM for the two younger samples. If those abilities would have been mainly responsible for passing the ToM tasks, then children who performed weakly on ToM should have also performed weakly on IC and language understanding. This pattern was not present for the majority of samples. The present data delivers valuable insights for future research, since relationships between the different dimensions of mental state understanding have not yet been fully explored.

In conclusion, all batteries consisted exclusively of established tasks, which were selected by a theory-guided approach and all tasks either aimed to infer others' mental states or to inhibit a prepotent response. Results of inter-correlation analyses are often used to warrant the formation of composite scores, but notably do not allow for valid conclusions about whether the single components actually measure the same cognitive processes. Therefore, I made the conscious decision to use the total scores of the multidimensional batteries for final analyses to avoid a small incremental procedure.

Possible Limitations of the Current Study

Some critical evaluations about the task construction were already discussed. Besides these remarks, a few additional comments concerning other limitations will follow. One shortcoming of this study is that the socio-economic status of the children's families was not assessed. Even if the investigation of this relationship was not the major interest of the project, it would have been an advantage to control for maternal education and family income, since connections between socio-economic status and children's behavioral problems and socio-emotional development have been found (Bradley & Corwyn, 2002; Dodge, Pettit, & Bates, 1994). If one considers the agents and institutions of a child's microsystem (e.g., family, peers and daycare institutions), which influences the development of a child (Bronfenbrenner, 1979), it is not possible to control all cofactors, and control measures must be carefully selected. The decision to choose language as a control variable is substantiated in the fact that language repeatedly was revealed as a significant correlate to all variables of interest, which are behavioral problems, ToM and IC (Astington & Jenkins, 1999; Hinshaw, 1992; Wolfe & Bell, 2004). However, future research might expand the number of control variables that might affect the relationship between the concepts of interest. For instance, evidence for an existing link between the socio economic status and cognitive abilities was delivered (Bradley & Corwyn, 2002; Dunn & Cutting, 1999), although relationships have not yet been fully investigated. Furthermore, moderating or mediating factors should also be considered. I previously mentioned possible influences of children's prosocial attitude, but also parenting style could be involved in relationships between social-cognitive competences and behavioral problems (Ruffman, Slade, Devitt, & Crowe, 2006). In addition to all the benefits of control variables, it has to be recognized that the assessment of numerous variables are time-consuming and distressing for children.

Another important matter is the size of the samples, which might have hindered us from detecting possible relations. Although the number of children included in our sample is indeed comparable to that of other developmental studies (Capage & Watson, 2001; Mink et al., 2014) or even higher, the still relatively small sample sizes of the current study should be borne in mind when interpreting the outcomes. Gaps in the data-sets resulted for instance from questionnaires, which were not returned, or absence at the second test session. For the sample of 2-year-old children only 57% of the caregiver questionnaires were completed, because not all of the children attended daycare. Firstly, this reduced the number of evaluable data, and secondly, it might assume a heterogeneous social background within this sample, because some children already experience social interactions in larger groups and education at kindergarten and others might experience mainly the interaction with their mothers. Particularly the 3-year-old children were somewhat less cooperative than the other age groups, which was why this sample size was increased as a reaction to this in the current project. Since capacities were limited for this project, it was not possible to expand the number of participants to a larger extent. Especially for regression analyses, a large body of data would improve the statistical power and generalizability of the findings. Some results only reached statistical significance at trend level, and it might be expected that results would have been stronger, if the sample size had been larger.

Furthermore, the study only provides correlational data and therefore delivers evidence for a link between ToM, IC and behavioral problems, but causal effects cannot be drawn and require further investigations. In consideration of a developmental path it neither can be concluded that the same direct links persist in later childhood when cognitive, emotional and regulative competences improve, nor can it be excluded that earlier variables influenced the link which was found at 4 years of age. Thus, it is of future interest to use the current tasks for further longitudinal examinations.

2.4 Conclusion

This study is the first providing extensive batteries for assessing the mental state understanding and inhibitory skills, including different dimensions of ToM and IC, for 2-, 3- and 4-year-old children. In this way, it was possible to investigate the unique contributions of ToM and IC to behavioral characteristics within different stages of early development. A major achievement is represented by the provision of comparable tasks for children from 2 to 4 years of age, because the construction of tasks was matched to a high degree with respect to children's age-specific cognitive abilities. With the multi-task and multi-informant design, I showed that there are no robust links between ToM or IC and behavioral problems at 2 and 3 years of age. Accordingly, there is no indication that behavioral difficulties early in childhood could be explained by the capacity of mental state understanding or inhibitory skills. Importantly, the issue changes for older children. At the age of 4 years, IC and ToM are both correlated to different forms of children's behavioral problems. By the detection of IC as an important predictor for behavioral problems, I offer the basis for future training programs aiming at a reduction of behavioral problems in preschoolers. The detection of the positive link between ToM and behavioral problems allows for the assumption that high social-cognitive abilities may put children at risk for developing difficulties in behavior. It is noticeable that children who are better in recognizing others' mental states show a high sensitivity to perceived impressions. This relationship and its underlying mechanisms appear quite complex and certainly deserves researchers' attention in the future. The dissertation project delivers a substantial contribution to identifying possible factors, which may contribute to behavioral problems in preschool age. From my point of view, it highlights the necessity to not only monitor children's early competences, but also on children's personal sensitivities, to evaluate their development on a professional basis, and to provide them with emotional support to guarantee positive growth. Particularly in today's society, where diagnoses and treatment rates of psychological diseases have been reported to increase within recent years and thereby receive more attention (Aurich-Beerheide & Knieps, 2014; Torio, Encinosa, Berdahl, McCormick, & Simpson, 2015), the present results are of significance. Early prevention programs should be initiated, as late interventions against behavioral problems only showed limited success (McNeil, Eyberg, Hembree Eisenstadt, Newcomb, & Funderburk, 1991). Follow-up work is required to replicate the current findings and to investigate the development longitudinally. An exciting topic for subsequent research is whether the observed relationships at 4 years of age continue over the course of development and whether there are early risk factors of the younger samples, which are predictive in nature and show only late manifesting effects.

References

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/ 2-3 and 1992 Profile*. Burlington, VT: Univeritsy of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles.* Burlington, VT: University of Vermont Department of Psychiatry.

Ahadi, S. A., Rothbart, M. K., & Ye, R. (1993). Children's temperament in the US and China: similarities and differences. *European Journal of Personality, 7,* 359–378. doi:10.1002/per.2410070506

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). Washington, D.C: American Psychiatric Association.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116,* 953–970. doi:10.1037/a0016923

Apperly, I. A., & Carroll, D. J. (2009). How do symbols affect 3- to 4-year-olds' executive function? Evidence from a reverse-contingency task. *Developmental Science, 12,* 1070–1082. doi:10.1111/j.1467-7687.2009.00856.x

Astington, J. W. (2001). The future of theory-of-mind research: Understanding motivational states, the role of language, and real-world consequences. *Child Development, 72,* 685–687. doi:10.1111/1467-8624.00305

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35,* 1311–1320. doi:10.1037//0012-1649.35.5.1311

Aurich-Beerheide, P., & Knieps, F. (2014). *BKK Gesundheitsreport 2014 (BKK Health report 2014): Gesundheit in Regionen - Zahlen, Daten, Fakte (Health in the regions - figures, data, facts)*. Berlin: MWV Medizinisch Wissenschaftl. Verl.-Ges.

Badenes, L. V., Clemente Estevan, R. A., & Garcia Bacete, F. J. (2000). Theory of mind and peer rejection at school. *Social Development, 9,* 271–283. doi:10.1111/1467-9507.00125

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14,* 110–118. doi:10.1016/j.tics.2009.12.006

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121,* 65–94. doi:10.1037/0033-2909.121.1.65

Bateman, A., & Fonagy, P. (2008). 8-year follow-up of patients treated for borderline personality disorder: Mentalization-based treatment versus treatment as usual. *American Journal of Psychiatry, 165,* 631–638. doi:10.1176/appi.ajp.2007.07040636

Bates, J. E., & Pettit, G. S. (2007). Temperament, parenting, and socialization. In J. E. Grusec & P. D. Hastings (Eds.), *Handbook of socialization. Theory and research* (pp. 153–177). New York: Guilford Press.

Bayley, N. (1993). *Bayley scales of infant development* (2nd ed.). San Antonio, TX: The Psychological Corporation, Harcourt Brace & Company.

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology, 41,* 328–337. doi:10.1037/0012-1649.41.2.328

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *The Quarterly Journal of Experimental Psychology A, 53,* 153–164. doi:10.1080/027249800390718

Bennett, M., & Galpert, L. (1993). Children's understanding of multiple desires. *International Journal of Behavioral Development, 16,* 15–33. doi:10.1177/016502549301600102

Berlin, L., Bohlin, G., & Rydell, A.-M. (2004). Relations between inhibition, executive functioning, and ADHD symptoms: A longitudinal study from age 5 to 8 ½ Years. *Child Neuropsychology (Neuropsychology, Development and Cognition: Section C), 9,* 255–266. doi:10.1076/chin.9.4.255.23519

Bodrova, E., & Leong, D. (2007). *Tools of the mind: The Vygotskian approach to early childhood education* (2nd ed). Upper Saddle River, N.J: Pearson/Merrill Prentice Hall.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53,* 371–399. doi:10.1146/annurev.psych.53.100901.135233

Brady, E. U., & Kendall, P. C. (1992). Comorbidity of anxiety and depression in children and adolescents. *Psychological Bulletin, 111,* 244–255. doi:10.1037/0033-2909.111.2.244

Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Bretherton, I., Fritz, J., Zahn-Waxler, C., & Ridgeway, D. (1986). Learning to talk about emotions: A functional perspective. *Child Development, 57,* 529–548.

Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication: When do infants acquire a "theory of mind"? In M. E. Lamb & L. R. Sherrod (Eds.), *Infant social cognition* (pp. 333–373). Hillsdale, NJ: Erlbaum.

Briggs-Gowan, M. J., Carter, A. S., Skuban, E. M., & Horwitz, S. M. (2001). Prevalence of social-emotional and behavioral problems in a community sample of 1- and 2-year-old children. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 811–819. doi:10.1097/00004583-200107000-00016

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, Mass: Harvard University Press.

Burgess, K. B., Wojslawowicz, J. C., Rubin, K. H., Rose-Krasnor, L., & Booth-LaForce, C. (2006). Social information processing and coping strategies of shy/withdrawn and aggressive children: Does friendship matter? *Child Development, 77,* 371–383. doi:10.1111/j.1467-8624.2006.00876.x

Buss, A. H., & Plomin, R. (1984). *Temperament: Early developing personality traits*. Hillsdale, N.J: L. Erlbaum Associates.

Buttelmann, D., Call, J., & Tomasello, M. (2009). Do great apes use emotional expressions to infer desires? *Developmental Science, 12,* 688–698. doi:10.1111/j.1467-7687.2008.00802.x

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2007). Enculturated chimpanzees imitate rationally. *Developmental Science, 10*(4), 31–38. doi:10.1111/j.1467-7687.2007.00630.x

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2008). Rational tool use and tool choice in human infants and great apes. *Child Development, 79,* 609–626. doi:10.1111/j.1467-8624.2008.01146.x

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition, 112,* 337–342. doi:10.1016/j.cognition.2009.05.006

Buttelmann, D., Over, H., Carpenter, M., & Tomasello, M. (2014). Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology, 119,* 120–126. doi:10.1016/j.jecp.2013.10.002

Buttelmann, F., & Buttelmann, D. (2013, September). *Einfluss der Attraktivität des Stimulusmaterials auf die Messung der Fähigkeit zum Verständnis mentaler Prozesse anderer bei 4-Jährigen.* In F. Buttelmann (Chair), Einflussfaktoren auf das Verständnis für die Mentalen Prozesse Anderer. Symposium conducted at the meeting of 21. Fachgruppentagung Entwicklungspsychologie der DGPs (EPSY), Saarbrücken, Deutschland.

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry, 36,* 113–149. doi:10.1111/j.1469-7610.1995.tb01657.x

Campbell, S. B., Pierce, E. W., March, C. L., Ewing, L. J., & Szumowski, E. K. (1994). Hard-to-manage preschool boys: Symptomatic behavior across contexts and time. *Child Development, 65,* 836–851. doi:10.2307/1131422

Capage, L., & Watson, A. C. (2001). Individual differences in theory of mind, aggressive behavior, and social skills in young children. *Early Education & Development, 12,* 613–628. doi:10.1207/s15566935eed1204_7

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: stability and prediction from ages 2 to 3. *Developmental Psychology, 40,* 1105–1122. doi:10.1037/0012-1649.40.6.1105

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032–1053. doi:10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69,* 672. doi:10.2307/1132197

Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development, 22,* 489–510. doi:10.1016/j.cogdev.2007.08.002

Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science, 8,* F13-F20. doi:10.1111/j.1467-7687.2004.00385.x

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language, 28,* 141–172. doi:10.1111/mila.12014

Caspi, A., Henry, B., McGee, R. O., Moffitt, T. E., & Silva, P. A. (1995). Temperamental origins of child and adolescent behavior problems: From age three to age fifteen. *Child Development, 66*(1), 55. doi:10.2307/1131190

Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development? *Social Development, 11,* 1–10. doi:10.1111/1467-9507.00183

Chess, S., & Thomas, A. (1996). *Temperament: Theory and practice*. *Brunner/Mazel basic principles into practice series: v. 12*. New York: Brunner/Mazel.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*(4), 377–395. doi:10.1016/0885-2014(94)90012-4

Cole, K., & Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. *British Journal of Developmental Psychology, 18,* 279–295. doi:10.1348/026151000165698

Compas, B. E., Connor-Smith, J. K., Saltzman, H., Thomsen, A. H., & Wadsworth, M. E. (2001). Coping with stress during childhood and adolescence: Problems, progress, and potential in theory and research. *Psychological Bulletin, 127,* 87–127. doi:10.1037//0033-2909.127.1.87

Coplan, R. J., & Armer, M. (2007). A "multitude" of solitude: A closer look at social withdrawal and nonsocial play in early childhood. *Child Development Perspectives, 1,* 26–32. doi:10.1111/j.1750-8606.2007.00006.x

Cullinan, D. (2004). Classification and definition of emotional and behavioral disorders. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 32–53). New York: Guilford Press.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development, 70,* 853–865. doi:10.1111/1467-8624.00061

Dettling, A. C., Gunnar, M. R., & Donzella, B. (1999). Cortisol levels of young children in full-day childcare centers: Relations with age and temperament. *Psychoneuroendocrinology, 24,* 519–536. doi:10.1016/S0306-4530(99)00009-8

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development,* 1–18. doi:10.1111/cdev.12237

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). Oxford, New York: Oxford University Press.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). The early years: Preschool program improves cognitive control. *Science, 318,* 1387–1388. doi:10.1126/science.1151148

Diamond, A., Prevor, M. B., Callender, G., & Druin, D. P. (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. *Monographs of the Society for Research in Child Development, 62.* doi:10.2307/1166208

Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65,* 649–665. doi:10.1111/j.1467-8624.1994.tb00774.x

Doherty, M. J. (2009). *Theory of mind: How children understand others' thoughts and feelings*. Hove, New York: Psychology Press.

Dowsett, S. M., & Livesey, D. J. (2000). The development of inhibitory control in preschool children: Effects of "executive skills" training. *Developmental Psychobiology, 36,* 161–174. doi:10.1002/(SICI)1098-2302(200003)36:2<161::AID-DEV7>3.0.CO;2-0

Dunn, J. (1995). Children as psychologists: The later correlates of individual differences in understanding of emotions and other minds. *Cognition & Emotion, 9,* 187–201. doi:10.1080/02699939508409008

Dunn, J., & Cutting, A. L. (1999). Understanding others, and individual differences in friendship interactions in young children. *Social Development, 8,* 201–219. doi:10.1111/1467-9507.00091

Durkheim, E. (1950). *The rules of sociological method*. Glencoe, IL: The Free Press.

Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry, 47,* 313–337. doi:10.1111/j.1469-7610.2006.01618.x

Eisenberg, N., Gershoff, E. T., Fabes, R. A., Shepard, S. A., Cumberland, A. J., Losoya, S. H., . . . Murphy, B. C. (2001). Mother's emotional expressivity and children's behavior problems and social competence: Mediation through children's regulation. *Developmental Psychology, 37,* 475–490. doi:10.1037//0012-1649.37.4.475

Ermisch, J. (2008). Origins of social immobility and inequality: Parenting and early child development. *National Institute Economic Review, 205,* 62–71. doi:10.1177/0027950108096589

Espy, K. A., Kaufmann, P. M., McDiarmid, M. D., & Glisky, M. L. (1999). Executive functioning in preschool children: Performance on A-not-B and other delayed response format tasks. *Brain and Cognition, 41,* 178–199. doi:10.1006/brcg.1999.1117

Espy, K. A., Sheffield, T. D., Wiebe, S. A., Clark, C. A., & Moehr, M. J. (2011). Executive control and dimensions of problem behaviors in preschool children. *Journal of Child Psychology and Psychiatry, 52,* 33–46. doi:10.1111/j.1469-7610.2010.02265.x

Faust, M. E., Balota, D. A., Duchek, J. M., Gernsbacher, M. A., & Smith, S. (1997). Inhibitory Control during sentence comprehension in individuals with dementia of the alzheimer type. *Brain and Language, 57,* 225–253. doi:10.1006/brln.1997.1747

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the level 1-level 2 distinction. *Developmental Psychology, 17,* 99–103. doi:10.1037//0012-1649.17.1.99

Franić, S., Middeldorp, C. M., Dolan, C. V., Ligthart, L., & Boomsma, D. I. (2010). Childhood and adolescent anxiety and depression: Beyond heritability. *Journal of the American Academy of Child & Adolescent Psychiatry, 49,* 820–829. doi:10.1016/j.jaac.2010.05.013

Frye, D. (1999). Development of intention: The relation of executive function to theory of mind. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention. Social understanding and self-control* (pp. 119–132). Mahwah, N.J: Lawrence Erlbaum Associates.

Frye, D., Zelazo, P. D., & Burack, J. A. (1998). Cognitive complexity and control: I. Theory of mind in typical and atypical development. *Current Directions in Psychological Science, 7,* 116–121. doi:10.1111/1467-8721.ep10774754

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10,* 483–527. doi:10.1016/0885-2014(95)90024-1

Funder, D. C., Block, J. H., & Block, J. (1983). Delay of gratification: Some longitudinal personality correlates. *Journal of Personality and Social Psychology, 44,* 1198–1213. doi:10.1037/0022-3514.44.6.1198

Gavazzi, I. G., & Ornaghi, V. (2011). Emotional state talk and emotion understanding: a training study with preschool children. *Journal of Child Language, 38*(05), 1124–1139. doi:10.1017/S0305000910000772

Gerardi-Caulton, G. (2000). Sensitivity to spatial conflict and the development of self-regulation in children 24-36 months of age. *Developmental Science, 3,* 397–404. doi:10.1111/1467-7687.00134

Gergely, G., Bekkering, H., & Király, I. (2002). Developmental psychology: Rational imitation in preverbal infants. *Nature, 415,* 755. doi:10.1038/415755a

Gernsbacher, M. A., & Faust, M. (1991). Chapter 5 The Role of Suppression in Sentence Comprehension. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 97–128). Amsterdam: Noth Holland.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2–7 years old on a stroop-like day-night test. *Cognition, 53,* 129–153. doi:10.1016/0010-0277(94)90068-X

Giancola, P. R., Mezzich, A. C., & Tarter, R. E. (1998). Executive cognitive functioning, temperament, and antisocial behavior in conduct-disordered adolescent females. *Journal of Abnormal Psychology, 107,* 629–641. doi:10.1037/0021-843X.107.4.629

Gini, G. (2006). Social cognition and moral cognition in bullying: What's wrong? *Aggressive Behavior, 32,* 528–539. doi:10.1002/ab.20153

Golden, M., Montare, A., & Bridger, W. (1977). Verbal control of delay behavior in two-year-old boys as a function of social class. *Child Development, 48,* 1107–1111. doi:10.2307/1128370

Goldsmith, H. H., Lemery, K. S., Aksan, N., & Buss, K. (2000). Temperament substrates of personality development. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 1–32). Mahwah, N.J: L. Erlbaum Associates.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry, 38,* 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59,* 26–37. doi:10.2307/1130386

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(02), 67. doi:10.1017/S1366728998000133

Gregory, A. M., & Eley, T. C. (2007). Genetic influences on anxiety in children: What we've learned and where we're heading. *Clinical Child and Family Psychology Review, 10,* 199–212. doi:10.1007/s10567-007-0022-8

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system*. Circle Pines, MN: American Guidance Service.

Grimm, H. (2001). *(Language development test for 3- to 5-year-old children) Sprachentwicklungstest für drei- bis fünfjährige Kinder (SETK 3-5)*. Göttingen: Hogrefe.

Grimm, H., Aktas, M., & Frevert, S. (2000). *(Language development test for 2-year-old children) Sprachentwicklungstest für zweijährige Kinder (SETK-2)*. Göttingen: Hogrefe.

Guajardo, N. R., & Watson, A. C. (2002). Narrative discourse and theory of mind development. *The Journal of Genetic Psychology, 163,* 305–325. doi:10.1080/00221320209598686

Hagekull, B. (1994). Infant temperament and early childhood functioning: Possible relations to the five-factor model. In C. F. Halverson, G. A. Kohnstamm, & R. Martin (Eds.), *The Developing structure of temperament and personality from infancy to adulthood* (pp. 227–240). Hillsdale, N.J: L. Erlbaum Associates.

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Developmental Science, 6,* 346–359. doi:10.1111/1467-7687.00289

Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as I do: 7-month-old infants selectively reproduce others' goals. *Developmental Science, 11,* 487–494. doi:10.1111/j.1467-7687.2008.00694.x

Hansen, M. (2010). If you know something, say something: Young children's problem with false beliefs. *Frontiers in Psychology,* 1–7. doi:10.3389/fpsyg.2010.00023

Hare, B., & Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences, 9,* 439–444. doi:10.1016/j.tics.2005.07.003

Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review, 102,* 458–489. doi:10.1037//0033-295X.102.3.458

Harvey, R. J., Fletcher, J., & French, D. J. (2001). Social reasoning. *Clinical Psychology Review, 21,* 447–469. doi:10.1016/S0272-7358(99)00068-9

Heckhausen, H., & Gollwitzer, P. M. (1987). Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and Emotion, 11*(2), 101–120.

Helland, W. A., Lundervold, A. J., Heimann, M., & Posserud, M.-B. (2014). Stable associations between behavioral problems and language impairments across childhood – The importance of pragmatic language problems. *Research in Developmental Disabilities, 35,* 943–951. doi:10.1016/j.ridd.2014.02.016

Higgins, E. T. (1989). Continuities and Discontinuities in Self-Regulatory and Self-Evaluative Processes: A Developmental Theory Relating Self and Affect. *Journal of Personality*, (2), 407–444. doi:10.1111/j.1467-6494.1989.tb00488.x

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111,* 127–155. doi:10.1037//0033-2909.111.1.127

Hoffmann, H. (1847). *Der Struwwelpeter*. Frankfurt am Main: Rütten & Loening.

Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57,* 567–582. doi:10.2307/1130337

Hughes, C. (1998). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34,* 1326–1339. doi:10.1037/0012-1649.34.6.1326

Hughes, C., & Cutting, A. L. (1999). Nature, nurture, and individual differences in early understanding of mind. *Psychological Science, 10,* 429–432. doi:10.1111/1467-9280.00181

Hughes, C., & Dunn, J. (1997). "Pretend you didn't know": Preschoolers' talk about mental states in pretend play. *Cognitive Development, 12,* 477–497. doi:10.1016/S0885-2014(97)90019-8

Hughes, C., & Ensor, R. (2006). Behavioural problems in 2-year-olds: Links with individual differences in theory of mind, executive function and harsh parenting. *Journal of Child Psychology and Psychiatry, 47,* 488–497. doi:10.1111/j.1469-7610.2005.01519.x

Hughes, C., & Ensor, R. (2007). Positive and protective: Effects of early theory of mind on problem behaviors in at-risk preschoolers. *Journal of Child Psychology and Psychiatry, 48,* 1025–1032. doi:10.1111/j.1469-7610.2007.01806.x

Hughes, C., & Ensor, R. (2008). Does executive function matter for preschoolers' problem behaviors? *Journal of Abnormal Child Psychology, 36,* 1–14. doi:10.1007/s10802-007-9107-6

Hughes, C., White, A., Sharpen, J., & Dunn, J. (2000). Antisocial, angry, and unsympathetic: "Hard-to-manage" preschoolers' peer problems and possible cognitive influences. *Journal of Child Psychology and Psychiatry, 41,* 169–179. doi:10.1111/1469-7610.00558

Jellesma, F., Rieffe, C., Terwogt, M., & Westenberg, M. (2009). Do I feel sadness, fear or both? Comparing self-reported alexithymia and emotional task-performance in children with many or few somatic complaints. *Psychology & Health, 24,* 881–893. doi:10.1080/08870440801998970

Kloo, D., & Perner, J. (2003). Training transfer between card sorting and false belief understanding: Helping children apply conflicting Descriptions. *Child Development, 74,* 1823–1839. doi:10.1046/j.1467-8624.2003.00640.x

Kloo, D., Perner, J., & Giritzer, T. (2010). Object-based set-shifting in preschoolers: Relations to theory of mind. In B. W. Sokol (Ed.), *Self and social regulation. Social* *interaction and the development of social understanding and executive functions* (pp. 193–217). Oxford, New York: Oxford University Press.

Kochanska, G., Murray, K., & Coy, K. C. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. *Child Development, 68,* 263–277. doi:10.1111/j.1467-8624.1997.tb01939.x

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67,* 490–507. doi:10.2307/1131828

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330,* 1830–1834. doi:10.1126/science.1190792

Lane, J. D., Wellman, H. M., Olson, S. L., Miller, A. L., Wang, L., & Tardif, T. (2013). Relations between temperament and theory of mind development in the united states and china: Biological and behavioral correlates of preschoolers' false-belief understanding. *Developmental Psychology, 49,* 825–836. doi:10.1037/a0028825

Lengua, L. J. (2003). Associations among emotionality, self-regulation, adjustment problems, and positive adjustment in middle childhood. *Journal of Applied Developmental Psychology, 24,* 595–618. doi:10.1016/j.appdev.2003.08.002

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind.". *Psychological Review, 94,* 412–426. doi:10.1037/0033-295X.94.4.412

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in ‘theory of mind’. *Trends in Cognitive Sciences, 8,* 528–533. doi:10.1016/j.tics.2004.10.001

Lewis, E. E., Dozier, M., Ackerman, J., & Sepulveda-Kozakowski, S. (2007). The effect of placement instability on adopted children's inhibitory control abilities and oppositional behavior. *Developmental Psychology, 43,* 1415–1427. doi:10.1037/0012-1649.43.6.1415

Linck, J. A., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition, 15,* 651–662. doi:10.1017/S136672891100054X

Lockl, K., Schwarz, S., & Schneider, W. (2004). Sprache und Theory of Mind. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 36,* 207–220. doi:10.1026/0049-8637.36.4.207

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development, 74,* 1130–1144. doi:10.1111/1467-8624.00597

Losoya, S. H., Eisenberg, N., & Fabes, R. A. (1998). Developmental issues in the study of coping. *International Journal of Behavioral Development, 22,* 287–313. doi:10.1080/016502598384388

Low, J. (2010). Preschoolers' implicit and explicit false-belief understanding: Relations with complex syntactical Mastery. *Child Development, 81,* 597–615. doi:10.1111/j.1467-8624.2009.01418.x

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology, 30,* 1–13. doi:10.1111/j.2044-835X.2011.02074.x

Lundy, S. M., Silva, G. E., Kaemingk, K. L., Goodwin, J. L., & Quan, S. F. (2010). Cognitive functioning and academic performance in elementary school children with anxious/depressed and withdrawn symptoms. *The Open Pediatric Medicine Journal, 4,* 1–9. doi:10.2174/1874309901004010001

Luo, Y., & Baillargeon, R. (2010). Toward a mentalistic account of early psychological reasoning. *Current Directions in Psychological Science, 19*(5), 301–307. doi:10.1177/0963721410386679

Maggio, V., Grañana, N. E., Richaudeau, A., Torres, S., Giannotti, A., & Suburo, A. M. (2014). Behavior problems in children with specific language impairment. *Journal of Child Neurology, 29,* 194–202. doi:10.1177/0883073813509886

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology, 33,* 101–121. doi:10.1006/jesp.1996.1314

Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development, 70,* 1297–1313. doi:10.1111/1467-8624.00095

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11.* doi:10.1017/S1366728907003227

Mash, E. J., & Johnston, C. (1983). Parental perceptions of child behavior problems, parenting self-esteem, and mothers' reported stress in younger and older hyperactive and normal children. *Journal of Consulting and Clinical Psychology, 51,* 86–99. doi:10.1037/0022-006X.51.1.86

McCormick, M. C., Gortmaker, S. L., & Sobol, A. M. (1990). Very low birth weight children: Behavior problems and school difficulty in a national sample. *The Journal of Pediatrics, 117,* 687–693. doi:10.1016/S0022-3476(05)83322-0

McGee, R., Partitdge, F., Williams, S., & Silva, P. A. (1991). A twelve-year follow-up of preschool hyperactive children. *Journal of the American Academy of Child & Adolescent Psychiatry, 30,* 224–232. doi:10.1097/00004583-199103000-00010

McGuigan, N., Makinson, J., & Whiten, A. (2011). From over-imitation to super-copying: Adults imitate causally irrelevant aspects of tool use with higher fidelity than young children. *British Journal of Psychology, 102,* 1–18. doi:10.1348/000712610X493115

McNeil, C. B., Eyberg, S., Hembree Eisenstadt, T., Newcomb, K., & Funderburk, B. (1991). Parent-child interaction therapy with behavior problem children: Generalization of treatment effects to the school setting. *Journal of Clinical Child Psychology, 20,* 140–151. doi:10.1207/s15374424jccp2002_5

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31,* 838–850. doi:10.1037//0012-1649.31.5.838

Menyuk, P. (2000). (Major aspects of lexical and semantic development)Wichtige Aspekte der lexikalischen und semantischen Entwicklung. In H. Grimm (Ed.), *Enzyklopädie der Psychologie. Themenbereich C: Theorie und Forschung. Serie III Sprache, Band 3: Sprachentwicklung* (pp. 172–192). Göttingen: Verlag für Psychologie C.J. Hogrefe.

Menyuk, P., Liebergott, J. W., & Schultz, M. C. (1995). *Early language development in full-term and premature infants*. Hillsdale, N.J: L. Erlbaum.

Miller, S. E., & Marcovitch, S. (2012). How theory of mind and executive function co-develop. *Review of Philosophy and Psychology, 3,* 597–625. doi:10.1007/s13164-012-0117-0

Miner, J. L., & Clarke-Stewart, K. A. (2008). Trajectories of externalizing behavior from age 2 to age 9: Relations with gender, temperament, ethnicity, parenting, and rater. *Developmental Psychology, 44,* 771–786. doi:10.1037/0012-1649.44.3.771

Mink, D., Henning, A., & Aschersleben, G. (2014). Infant shy temperament predicts preschoolers theory of mind. *Infant Behavior and Development, 37,* 66–75. doi:10.1016/j.infbeh.2013.12.001

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology, 54,* 687–696. doi:10.1037//0022-3514.54.4.687

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21,* 8–14. doi:10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100. doi:10.1006/cogp.1999.0734

Moffitt, T. E. (1990). Juvenile delinquency and attention deficit disorder: Boys' developmental trajectories from age 3 to age 15. *Child Development, 61,* 893–910. doi:10.1111/j.1467-8624.1990.tb02830.x

Moffitt, T. E. (2005). Genetic and environmental influences on antisocial behaviors: Evidence from behavioral–genetic research. *Advances in Genetics, 55,* 41–104. doi:10.1016/S0065-2660(05)55003-X

Morris, A. S., Silk, J. S., Steinberg, L., Sessa, F. M., Avenevoli, S., & Essex, M. J. (2002). Temperamental vulnerability and negative parenting as interacting predictors of child adjustment. *Journal of Marriage and Family, 64,* 461–471. doi:10.1111/j.1741-3737.2002.00461.x

Morris, N., Keane, S., Calkins, S., Shanahan, L., & O'Brien, M. (2014). Differential components of reactivity and attentional control predicting externalizing behavior. *Journal of Applied Developmental Psychology, 35,* 121–127. doi:10.1016/j.appdev.2014.02.002

Moses, L. J., Coon, J. A., & Wusinich, N. (2000). Young children's understanding of desire formation. *Developmental Psychology, 36,* 77–90. doi:10.1037/0012-1649.36.1.77

Müller, U., Liebermann-Finestone, D. P., Carpendale, J. I., Hammond, S. I., & Bibok, M. B. (2012). Knowing minds, controlling actions: The developmental relations between theory of mind and executive function from 2 to 4years of age. *Journal of Experimental Child Psychology, 111,* 331–348. doi:10.1016/j.jecp.2011.08.014

Nielsen, M., & Blank, C. (2011). Imitation in young children: When who gets copied is more important than what gets copied. *Developmental Psychology, 47,* 1050–1053. doi:10.1037/a0023866

Olson, S. L., Schilling, E. M., & Bates, J. E. (1999). Measurement of impulsivity: Construct coherence, longitudinal stability, and relationship with externalizing problems in middle childhood and adolescence. *Journal of Abnormal Child Psychology, 27*(2), 151–165. doi:10.1023/A:1021915615677

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science (New York, N.Y.), 308*(5719), 255–258. doi:10.1126/science.1107621

Oosterlaan, J., & Sergeant, J. A. (1996). Inhibition in ADHD, aggressive, and anxious children: A biologically based model of child psychopathology. *Journal of Abnormal Child Psychology, 24*(1), 19–36. doi:10.1007/BF01448371

Ormel, J., Oldehinkel, A. J., Ferdinand, R. F., Hartman, C. A., de Winter, A. F., Veenstra, R., . . . Verhulst, F. C. (2005). Internalizing and externalizing problems in adolescence: general and dimension-specific effects of familial loadings and preadolescent temperament traits. *Psychological Medicine, 35,* 1825. doi:10.1017/S0033291705005829

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32,* 1081–1105. doi:10.1111/j.1469-7610.1991.tb00351.x

Patterson, G. R., & MacCoby, E. E. (1980). Mothers: The unacknowledged victims. *Monographs of the Society for Research in Child Development, 45,* 1. doi:10.2307/1165841

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development, 73,* 752–767. doi:10.1111/1467-8624.00436

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, (2), 125–137. doi:10.1111/j.2044-835X.1987.tb01048.x

Perner, J., & Ruffman, T. (2005). Psychology. Infants' insight into the mind: How deep? *Science (New York, N.Y.), 308*(5719), 214–216. doi:10.1126/science.1111656

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology, 113,* 535–551. doi:10.1016/j.jecp.2012.06.013

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1,* 515. doi:10.1017/S0140525X00076512

Prior, M., Smart, D., Sanson, A., & Oberklaid, F. (2000). Does shy-inhibited temperament in childhood lead to anxiety problems in adolescence? *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(4), 461–468. doi:10.1097/00004583-200004000-00015

Raaijmakers, M. A. J., Smidts, D. P., Sergeant, J. A., Maassen, G. H., Posthumus, J. A., van Engeland, H., & Matthys, W. (2008). Executive functions in preschool children with aggressive behavior: Impairments in inhibitory control. *Journal of Abnormal Child Psychology, 36,* 1097–1107. doi:10.1007/s10802-008-9235-7

Renouf, A., Brendgen, M., Parent, S., Vitaro, F., David Zelazo, P., Boivin, M., . . . Séguin, J. R. (2010). Relations between theory of mind and indirect and physical aggression in kindergarten: Evidence of the moderating role of prosocial behaviors. *Social Development, 19,* 535–555. doi:10.1111/j.1467-9507.2009.00552.x

Repacholi, B., Slaughter, V., Pritchard, M., & & Gibbs, V. (2003). Theory of mind, machiavellianism, and social functioning in childhood. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind. Implications for typical and atypical development* (pp. 67–97). New York: Psychology Press.

Repacholi, B. M. (1998). Infants' use of attentional cues to identify the referent of another person's emotional expression. *Developmental Psychology, 34,* 1017–1025. doi:10.1037/0012-1649.34.5.1017

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33,* 12–21. doi:10.1037//0012-1649.33.1.12

Riggs, N. R., Blair, C. B., & Greenberg, M. T. (2004). Concurrent and 2-year longitudinal relations between executive function and the behavior of 1st and 2nd grade children. *Child Neuropsychology (Neuropsychology, Development and Cognition: Section C), 9,* 267–276. doi:10.1076/chin.9.4.267.23513

Riggs, N. R., Greenberg, M. T., Kusché, C. A., & Pentz, M. A. (2006). The mediational role of neurocognition in the behavioral outcomes of a social-emotional prevention program in elementary school students: Effects of the PATHS curriculum. *Prevention Science, 7,* 91–102. doi:10.1007/s11121-005-0022-1

Rothbart, M. K. (2011). *Becoming who we are: Temperament and personality in development*. New York: Guilford Press.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40*(1), 21–39.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The children's behavior questionnaire. *Child Development, 72,* 1394–1408. doi:10.1111/1467-8624.00355

Rothbart, M. K., & Bates, J. E. (1998). Temperament. In Series Ed. (Damon, W.) & Eisenberg, N. (Vol. Ed.) (Eds.), *Handbook of child psychology* (5th ed., Vol. 3, pp. 105–176). New York [u.a.]: Wiley.

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology / ed.-in-chief William Damon and Richard M. Lerner: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 99–176). New York, NY [u.a.]: Wiley.

Rubenstein, J. L., & Howes, C. (1979). Caregiving and infant behavior in day care and in homes. *Developmental Psychology, 15,* 1–24. doi:10.1037//0012-1649.15.1.1

Rubin, K. H., Burgess, K. B., Dwyer, K. M., & Hastings, P. D. (2003). Predicting preschoolers' externalizing behaviors from toddler temperament, conflict, and maternal negativity. *Developmental Psychology, 39,* 164–176. doi:10.1037//0012-1649.39.1.164

Rubio-Fernandez, P., & Geurts, B. (2012). How to pass the false-belief task before your fourth birthday. *Psychological Science, 24,* 27–33. doi:10.1177/0956797612447819

Ruffman, T., Slade, L., Devitt, K., & Crowe, E. (2006). What mothers say and what they do: The relation between parenting, theory of mind, language and conflict/cooperation. *British Journal of Developmental Psychology, 24,* 105–124. doi:10.1348/026151005X82848

Russell, A., Hart, C., Robinson, C., & Olsen, S. (2003). Children's sociable and aggressive behaviour with peers: A comparison of the US and Australia, and contributions of temperament and parenting styles. *International Journal of Behavioral Development, 27,* 74–86. doi:10.1080/01650250244000038

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9,* 331–349. doi:10.1111/j.2044-835X.1991.tb00881.x

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind. A comparison of chinese and U.S. preschoolers. *Psychological Science, 17,* 74–81. doi:10.1111/j.1467-9280.2005.01667.x

Schofield, W. N. (1976). Do children find movements which cross the body midline difficult? *Quarterly Journal of Experimental Psychology, 28,* 571–582. doi:10.1080/14640747608400584

Schult, C. A. (2002). Children's understanding of the distinction between intentions and desires. *Child Development, 73,* 1727–1747. doi:10.1111/1467-8624.t01-1-00502

Searle, J. R. (1983). *Intentionality: an essay in the philosophy of mind*. New York: Cambridge.

Shiner, R., & Caspi, A. (2003). Personality differences in childhood and adolescence: measurement, development, and consequences. *Journal of Child Psychology and Psychiatry, 44,* 2–32. doi:10.1111/1469-7610.00101

Simpson, A., Riggs, K. J., Beck, S. R., Gorniak, S. L., Wu, Y., Abbott, D., & Diamond, A. (2012). Refining the understanding of inhibitory processes: how response prepotency is created and overcome. *Developmental Science, 15,* 62–73. doi:10.1111/j.1467-7687.2011.01105.x

Simpson, A., Riggs, K. J., & Simon, M. (2004). What makes the windows task difficult for young children: Rule inference or rule use? *Journal of Experimental Child Psychology, 87,* 155–170. doi:10.1016/j.jecp.2003.11.002

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23,* 117–141. doi:10.1348/026151004X21332

Slaughter, V., Dennis, M. J., & Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *British Journal of Developmental Psychology, 20,* 545–564. doi:10.1348/026151002760390945

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science, 13,* 907–912. doi:10.1111/j.1467-7687.2009.00946.x

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind and Language, 17,* 3–23. doi:10.1111/1468-0017.00186

Stevenson, J., & Richman, N. (1978). Behavior, language, and development in three-year-old children. *Journal of Autism and Childhood Schizophrenia, 8,* 299–313. doi:10.1007/BF01539633

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18,* 580–586. doi:10.1111/j.1467-9280.2007.01943.x

Sutton, J., Smith, P. K., & Swettenham, J. (2001). Bullying and 'Theory of Mind': A critique of the 'Social Skills Deficit' view of anti-social Behaviour. *Social Development, 8,* 117–127.

Suway, J. G., Degnan, K. A., Sussman, A. L., & Fox, N. A. (2012). The relations among theory of mind, behavioral inhibition, and peer interactions in early childhood. *Social Development, 21,* 331–342. doi:10.1111/j.1467-9507.2011.00634.x

Symons, D. K., Fossum, K.-L. M., & Collins, T. K. (2006). A longitudinal study of belief and desire state discourse during mother?Child play and later false belief understanding. *Social Development, 15,* 676–692. doi:10.1111/j.1467-9507.2006.00364.x

Thoermer, C., Sodian, B., Vuori, M., Perst, H., & Kristen, S. (2012). Continuity from an implicit to an explicit understanding of false belief from infancy to preschool age. *British Journal of Developmental Psychology, 30,* 172–187. doi:10.1111/j.2044-835X.2011.02067.x

Thüringer Ministerium für Bildung, Wissenschaft und Kultur. (2008). *(Thuringian curicula for children until 10 years of age)Thüringer Bildungsplan für Kinder bis 10 Jahre* (Stand: August 2008). Weimar, Berlin: Verl. Das Netz.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28.* doi:10.1017/S0140525X05000129

Torio, C. M., Encinosa, W., Berdahl, T., McCormick, M. C., & Simpson, L. A. (2015). Annual report on health care for children and youth in the United States: National estimates of cost, utilization and expenditures for children with mental health conditions. *Academic Pediatrics, 15,* 19–35. doi:10.1016/j.acap.2014.07.007

Uzgiris, I. C. (1981). Two functions of imitation during infancy. *International Journal of Behavioral Development, 4*(1), 1–12. doi:10.1177/016502548100400101

Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science, 20,* 174–180. doi:10.1177/0963721411408772

Vygotskiĭ, L. S., & Cole, M. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

Walker, S. (2005). Gender differences in the relationship between young children's peer-related social competence and individual differences in theory of mind. *The Journal of Genetic Psychology, 166,* 297–312. doi:10.3200/GNTP.166.3.297-312

Watamura, S. E., Donzella, B., Alwin, J., & Gunnar, M. R. (2003). Morning-to-afternoon increases in cortisol concentrations for infants and toddlers at child care: Age differences and behavioral correlates. *Child Development, 74,* 1006–1020. doi:10.1111/1467-8624.00583

Webber, J., & Plotts, C. A. (2008). *Emotional and behavioral disorders: Theory and practice* (5th ed). Boston: Pearson Education, Inc.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684. doi:10.1111/1467-8624.00304

Wellman, H. M., Lane, J. D., LaBounty, J., & Olson, S. L. (2011). Observant, nonaggressive temperament predicts theory-of-mind development. *Developmental Science, 14,* 319–326. doi:10.1111/j.1467-7687.2010.00977.x

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75,* 523–541. doi:10.1111/j.1467-8624.2004.00691.x

Wilson, D. S., Near, D., & Miller, R. R. (1996). Machiavellianism: A synthesis of the evolutionary and psychological literatures. *Psychological Bulletin, 119*(2), 285–299. doi:10.1037/0033-2909.119.2.285

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition,* 108–128.

Wolfe, C. D., & Bell, M. A. (2004). Working memory and inhibitory control in early childhood: Contributions from physiology, temperament, and language. *Developmental Psychobiology, 44,* 68–83. doi:10.1002/dev.10152

Woodward, A. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69,* 1–34. doi:10.1016/S0010-0277(98)00058-4

Yiwen, W., Chongde, L., & Wenxin, Z. (2004). A study on the relation of children's aggression to their theory-of-mind. *Psychological Science (China)*, (27), 540–544.

Zahn-Waxler, C., Klimes-Dougan, B., & Slattery, M. J. (2000). Internalizing problems of childhood and adolescence: Prospects, pitfalls, and progress in understanding the development of anxiety and depression. *Development and Psychopathology, 12,* 443–466. doi:10.1017/S0954579400003102

Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature protocols, 1*(1), 297–301. doi:10.1038/nprot.2006.46

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68,* 93–119. doi:10.1111/j.0037-976X.2003.00266.x

Zentner, M. (2011). *(The integrative child temperament inventory) Inventar zur integrativen Erfassung des Kind-Temperaments (IKT)*. Bern: Hans Huber.

Zentner, M., & Bates, J. E. (2008). Child temperament: An integrative review of concepts, research programs, and measures. *International Journal of Developmental Science, 2,* 7–37. doi:10.3233/DEV-2008-21203

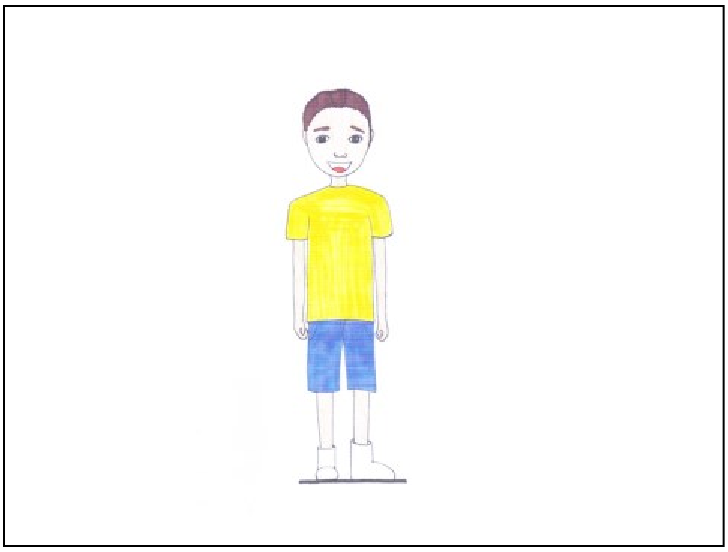
Zmyj, N., Daum, M., & Aschersleben, G. (2009). The development of rational imitation in 9- and 12-month-old infants. *Infancy, 14*(1), 131–141. doi:10.1080/15250000802569884

Appendix

Appendix A - Material of the Intention-Understanding Task for 4-Year-Old Children

1. Picture Story “Tom”

1. Card

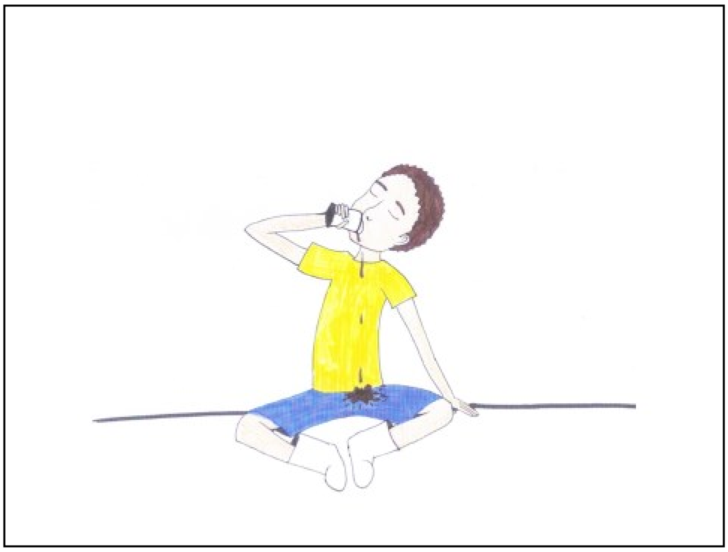


2. Card

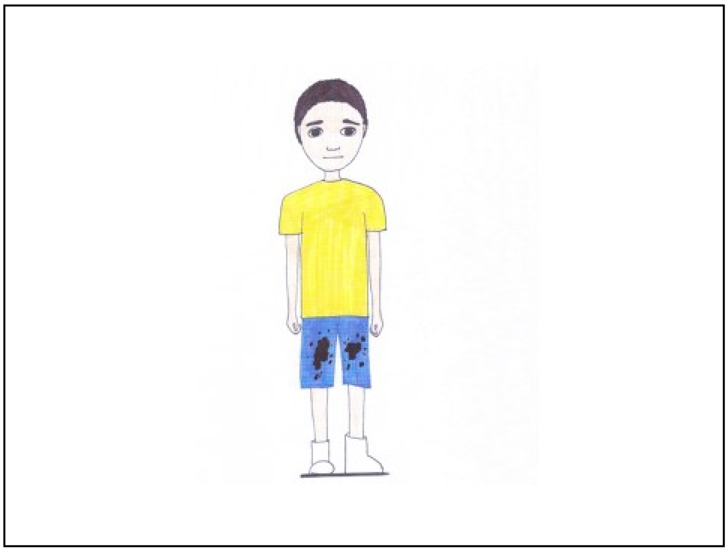
3. Card
Story Type A

3. Card
Story Type B

3. Card
Story Type C



4. Card



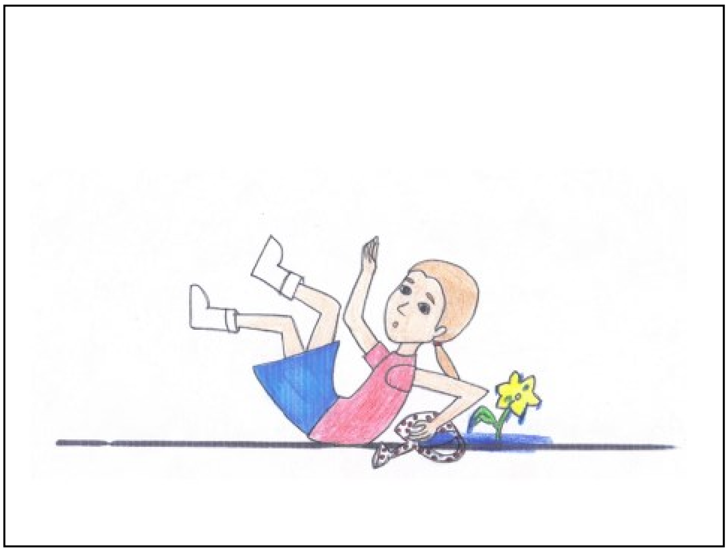
5. Cards shown for the forced-choice question. Pair of cards alternate according to story type.

2. Picture Story “Maria”

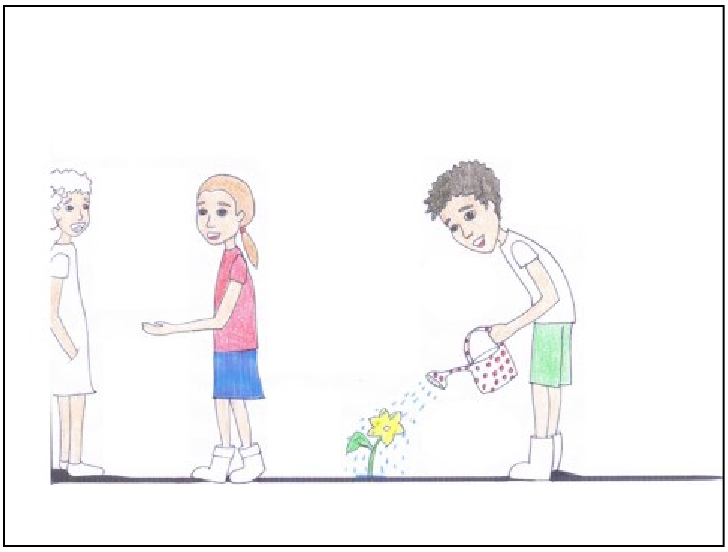
1. Card

2. Card

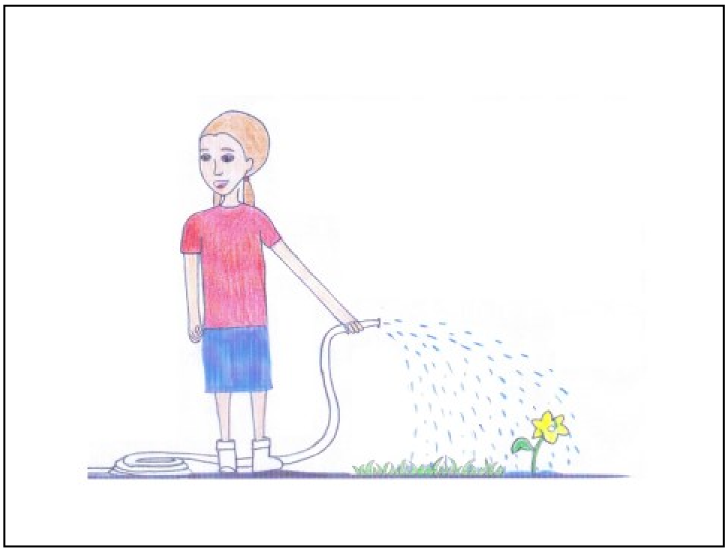
3. Card
Story Type A



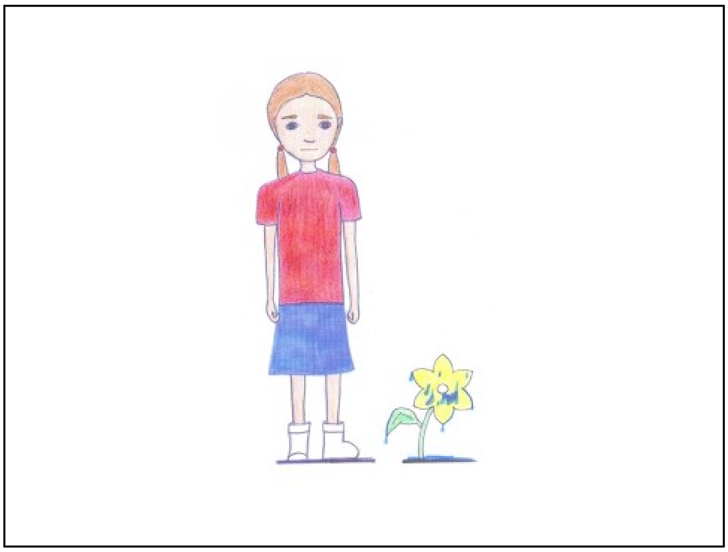
3. Card
Story Type B



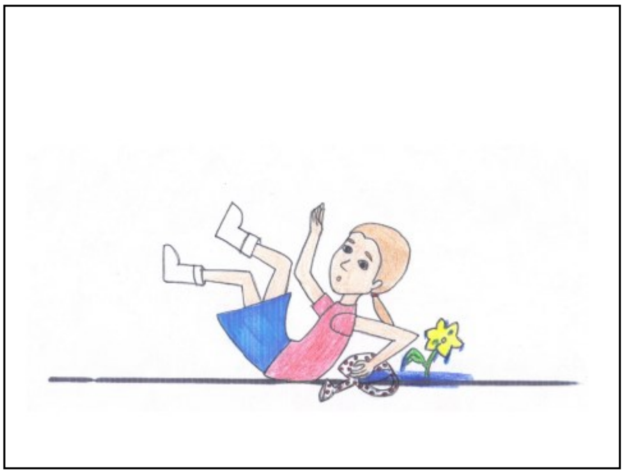
3. Card
Story Type C



4. Card

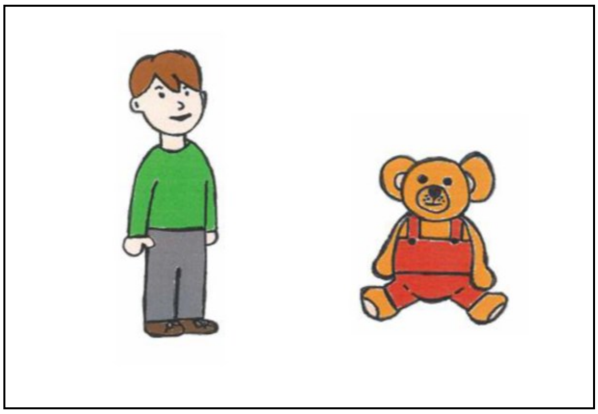


5. Cards shown for the forced-choice question. Pair of cards alternate according to story type.

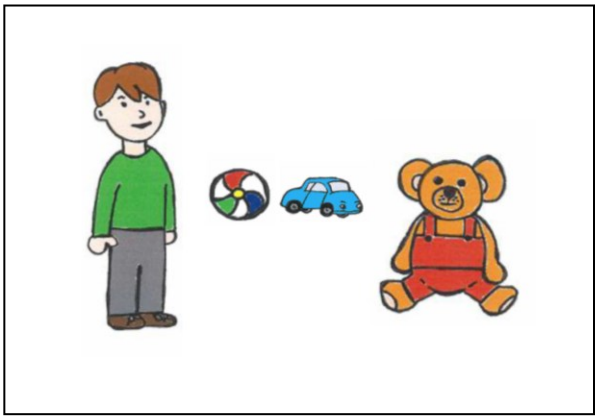


Appendix B – Material of the Preference-Understanding Task for 4-Year-Old Children

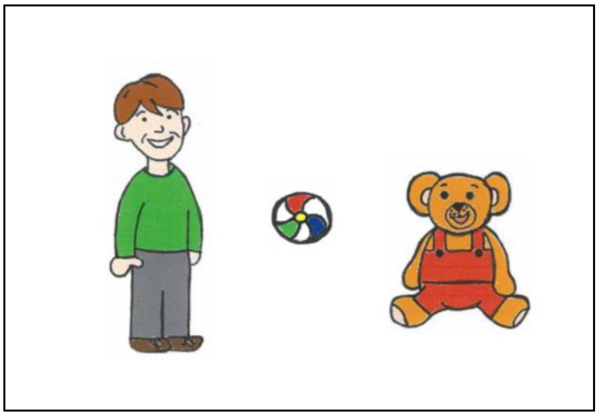
1. Card



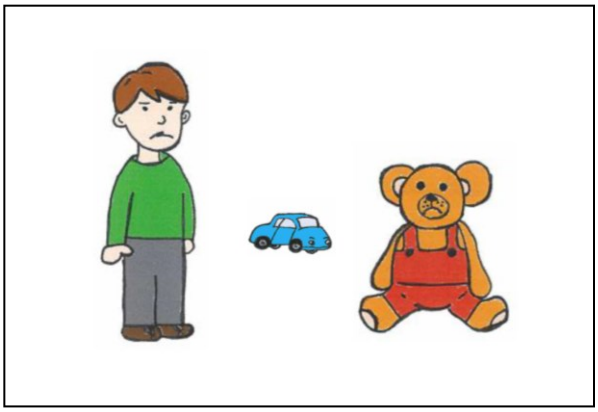
2. Card



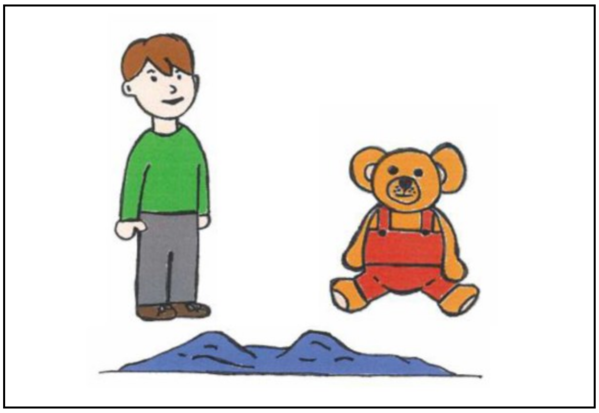
3. Card



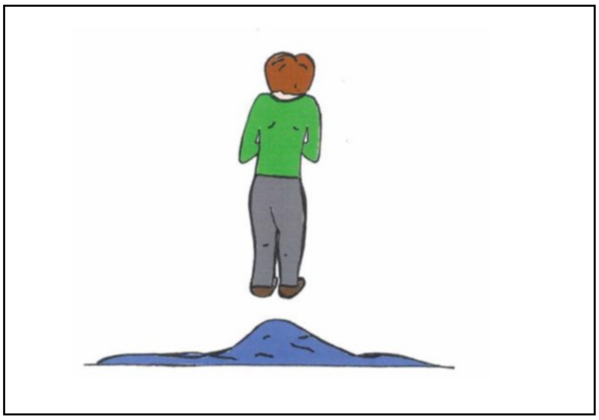
4. Card



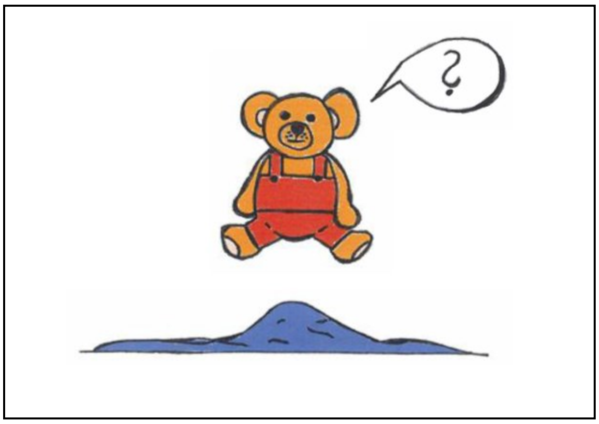
5. Card



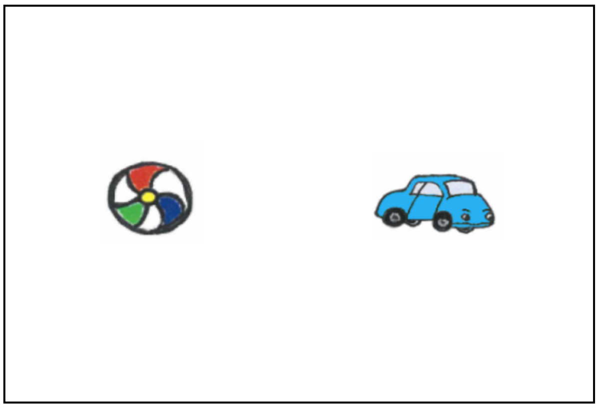
6. Card



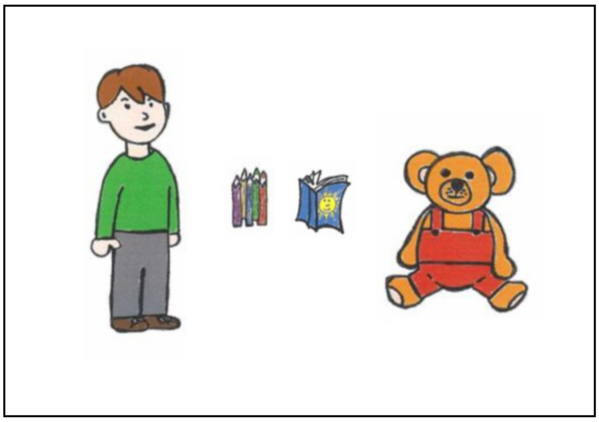
7. Card

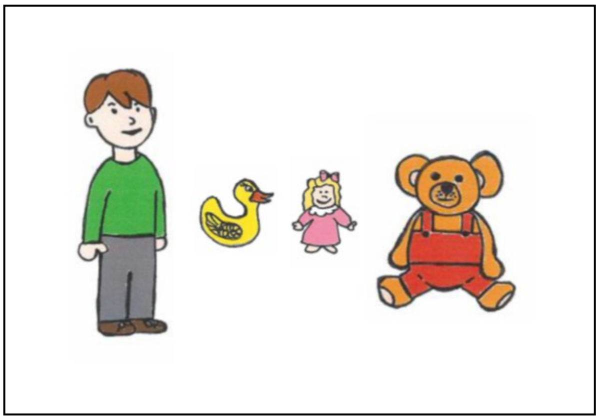


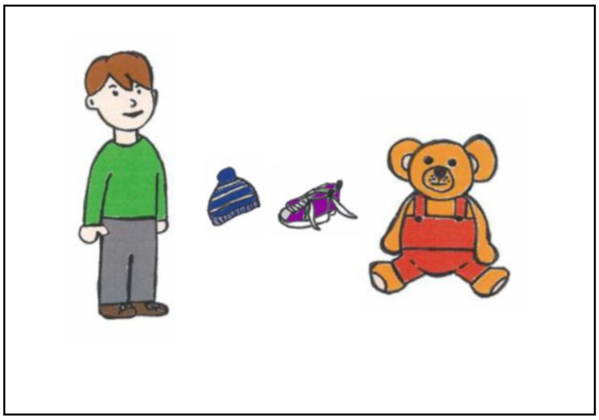
8. Card for the forced-choice question.

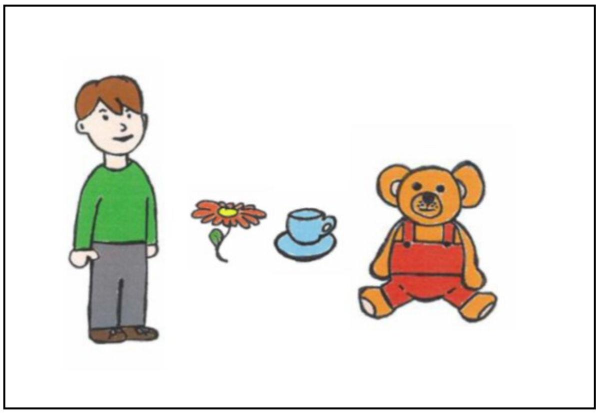


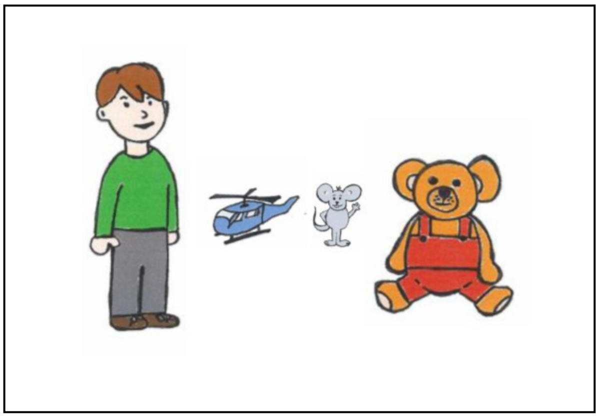
Other pairs of objects.











Appendix C - Analyses of ToM and IC Batteries

Friedman tests were run to compare mean performances on separate tasks within the batteries. For 2-year-old children no differences between the mean scores of the tasks included in the ToM battery were revealed ($\chi^2(3) = 5.693$, $p = .151$), but there was a statistically significant difference between the mean scores of the tasks included in the IC battery ($\chi^2(2) = 35.678$, $p \leq .001$). For mean ranks see Table 8. Further analyses using Wilcoxon tests showed that children scored better on the delay-of-gratification task compared to the inhibition-of-concept task ($N = 71$, $Z = -3.744$, $p \leq .001$, $r = .444$) and the spatial-inhibition task ($N = 71$, $Z = -5.329$, $p \leq .001$, $r = .633$).

Table 8.

Summary of the Mean Ranks of Tasks Included in the IC Battery at 4 Years of Age

IC Task	Mean Rank
Inhibition of Concept	1.90
Spatial Inhibition	1.55
Delay	2.55

Analyses for the sample of 3-year-old children revealed statistically significant differences between the mean scores of the tasks included in the ToM battery ($\chi^2(3) = 95.892$, $p \leq .001$) and the IC battery ($\chi^2(2) = 29.470$, $p \leq .001$). For mean ranks see Table 9. Further analyses using Wilcoxon tests showed for the ToM battery that children showed lower mean performances on the false-belief tasks compared to all other tasks (preference understanding: $N = 81$, $Z = -5.107$, $p \leq .001$, $r = .567$; intention understanding: $N = 78$, $Z = -6.255$, $p \leq .001$, $r = .708$; goal understanding: $N = 74$, $Z = -7.009$, $p \leq .001$, $r = .815$;), showed lower mean performances on the preference-task compared to the intention-understanding task ($N = 78$, $Z = -4.700$, $p \leq .001$, $r = .532$) and goal-understanding task ($N = 72$, $Z = -6.794$, $p \leq .001$, $r = .801$), and showed lower mean performances on the intention-understanding task compared to the goal-understanding task ($N = 71$, $Z = -4.756$, $p \leq .001$, $r = .564$). Within the IC battery, children showed lower mean performances on the spatial-inhibition task compared to the inhibition-of-concept task (N = 67, Z = -4.909, p ≤ .001, r = .560) and the delay task (N = 57, Z = -5.236, p ≤ .001, r = .694).

Table 9.

Summary of the Mean Ranks of Tasks Included in the ToM and IC Batteries at 3 Years of Age

ToM Task	**Mean Rank**	**IC Task**	**Mean Rank**
Goal	3.54	Inhibition of Concept	2.16
Intention	2.84	Spatial Inhibition	1.44
Preference	2.04	Delay	2.39
False Belief Score	1.57		

Analyses for the sample of 4-year-old children revealed statistically significant differences between the mean scores of the tasks included in the ToM battery ($\chi^2(3) = 61.622$, $p \leq .001$) and in the IC battery ($\chi^2(2) = 8.757$, $p = .013$). For mean ranks see Table 10. Further analyses using Wilcoxon tests showed for the ToM battery that children showed lower mean performances on the false-belief tasks compared to intention-understanding task ($N = 79$, $Z = -2.231$, $p = .026$, $r = .251$) and the goal-understanding task ($N = 78$, $Z = -6.572$, $p \leq .001$, $r = .744$). Furthermore, children showed better mean performances on the goal-understanding task compared to the preference-understanding task ($N = 66$, $Z = -6.794$, $p \leq .001$, $r = .836$) and the intention-understanding task ($N = 78$, $Z = -6.869$, $p \leq .001$, $r = .778$). For the IC battery Wilcoxon tests revealed no statistical significant differences.

Table 10.

Summary of the Mean Ranks of Tasks Included in the ToM and IC Batteries at 4 Years of Age

ToM Task	Mean Rank	IC Task	Mean Rank
Goal	3.54	Inhibition of Concept	1.75
Intention	2.36	Spatial Inhibition	1.95
Preference	2.13	Delay	2.30
False Belief Score	1.98		

Additionally, analyses were run to examine inter-correlations between the four dimensions of ToM and the three tasks of IC. Analyses revealed no inter-correlations between the dimensions within the batteries of 2-year-old children (ToM battery α = -.15, IC battery α = -.113), 3-year-old children (ToM battery α = -.26, IC battery α = .12) and 4 -year-old children (ToM battery α = .15, IC battery α = .40).

Tasks of the ToM Batteries

Goal-understanding task – 2-year-old children. Only the scores of the house condition were used, since 38 out of 82 children refused to participate at all in the no-house condition, or only performed half or less than half of the trials. Overall, children did not pass the task above chance level (Wilcoxon test, $N = 66$, $Z = -.430$, $p = .667$). For the means of performance, see Figure 6.

Goal-understanding task – 3-year-old children. Overall, children performed with total-scores above chance level (Wilcoxon test, $N = 75$, $Z = -7.383$, $p < .001$, $r = .852$). No difference between conditions (ear, dot, no-dot) and no order effects were present. For the means of performance, see Figure 6.

Goal-understanding task – 4-year-old children. Children performed with total scores above chance level (Wilcoxon tests, $N = 78$, $Z = -7.703$, $p < .001$, $r = 0.872$). A difference concerning conditions was not found. For the means of performance, see Figure 6.

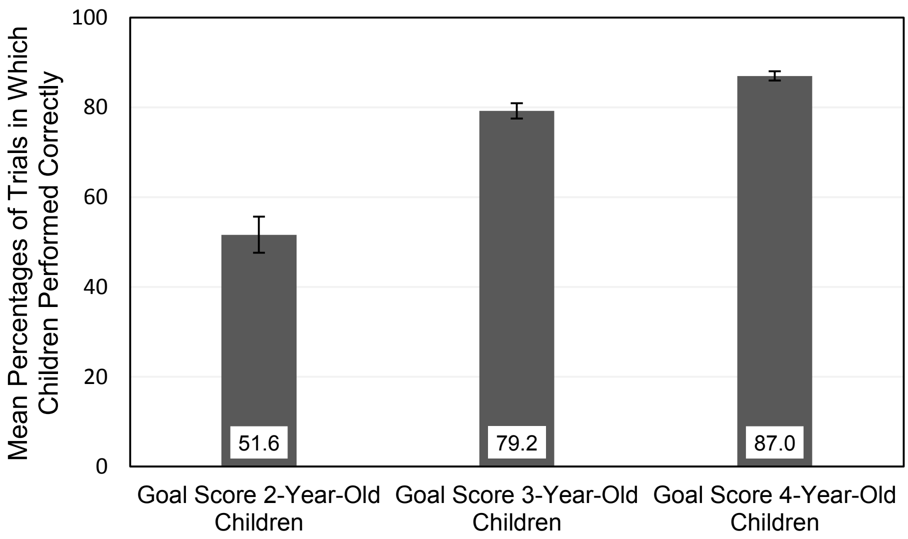


Figure 6. Mean percentages of trials in which children performed correctly on the goal understanding tasks.

Intention-understanding task – 2-year-old children. Overall, children did not pass the task above chance level (Wilcoxon test, $N = 80$, $Z = -1.622$, $p = .105$). No difference in scoring between the first block of trials and the second block of trials was found (Wilcoxon test, $N = 77$, $Z = -.093$, $p = .926$). No order effects were present. For the means of performance, see Figure 8.

Intention-understanding task – 3-year-old children. Overall, children passed the task above chance level (Wilcoxon test, $N = 81$, $Z = -5.163$, $p < .001$, $r = .574$). They performed better in the match condition than in the mismatch condition (Wilcoxon test, $N = 81$, $Z = -2.858$, $p = .004$, $r = .318$). For results see Figure 7. Children performed in the match condition above chance level (Wilcoxon test, $N = 82$, $Z = -3.377$, $p = .001$, $r = .373$). Children performed in the mismatch condition at chance level (Wilcoxon test, $N = 82$, $Z = -.693$, $p = .488$). No order effects were present. For the means of performance, see Figure 8.

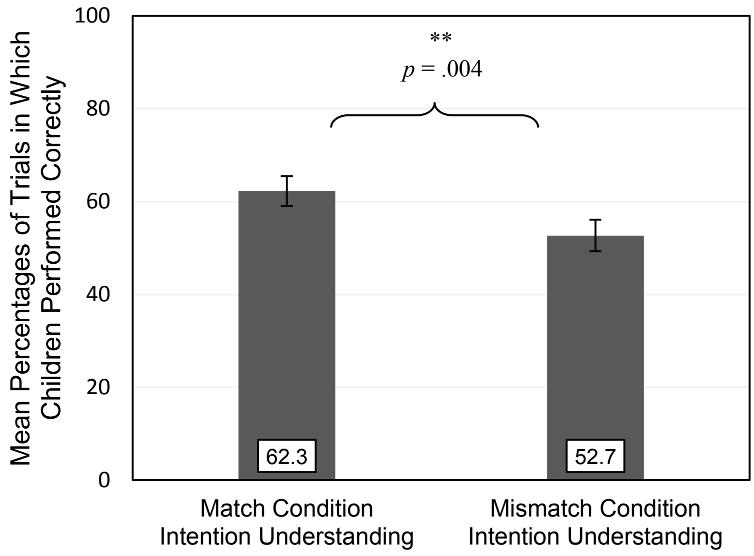


Figure 7. Mean percentages of trials in which children performed correctly on the intention understanding tasks comparing the match and mismatch condition.

Intention-understanding task – 4-year-old children. Children had difficulties in answering the open test questions. To avoid a bias due to children's communications skills, only answers to the forced-choice questions were analyzed. Children did not show differences between the Tom- and Maria-picture stories in the three story types A, B and C (McNemar tests, χ^2_A (N = 79) = 2.065, p = .151; χ^2_B (N = 79) = .346, p = .556; χ^2_C (N = 78), p = .064). In all story types, children did not perform the forced-choice questions above chance level (Wilcoxon tests, N_A = 79, Z_A = -.866, p = .386; N_B = 79, Z_B = -1.236, p = .216; N_C = 79, Z_C = .000, p = 1.000). For the means of performance, see Figure 8.

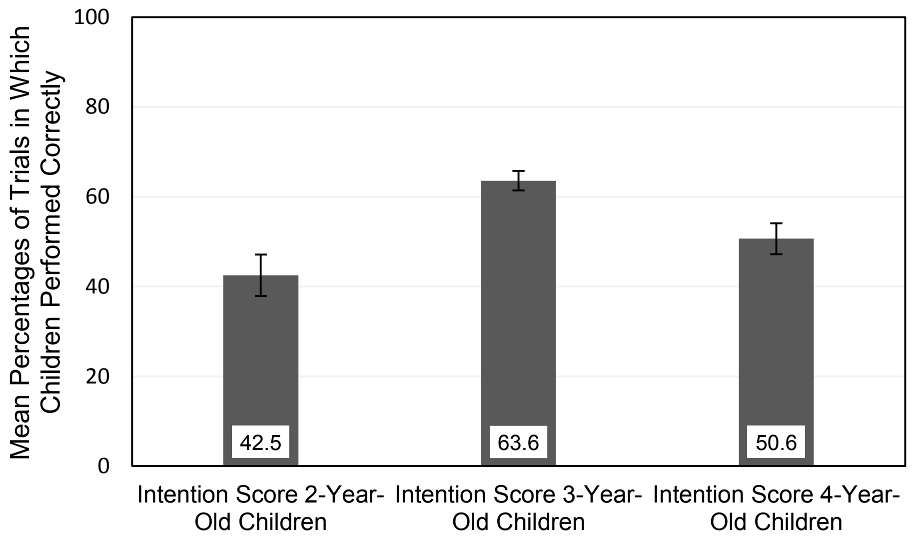


Figure 8. Mean percentages of trials in which children performed correctly on the intention understanding tasks

Preference-understanding task – 2-year-old children. Overall, children performed at chance level (Wilcoxon test, $N = 79$, $Z = -1.306$, $p = .192$). No difference in scoring between the match and mismatch conditions was found (Wilcoxon test, $N = 78$, $Z = -1.344$, $p = .179$). No order effects were present. For the means of performance, see Figure 10.

Preference-understanding task – 3-year-old children. Overall, children chose a cup at chance level (Wilcoxon test, $N = 84$, $Z = -.373$, $p = .709$). No difference in scoring between the first block of test trials and the last block of test trials was found (Wilcoxon test, $N = 81$, $Z = -.859$, $p = .390$). For the means of performance, see Figure 10.

Preference-understanding task – 4-year-old children. Overall children did not pass the task above chance level (Wilcoxon test, $N = 68$, $Z = 1.622$, $p = .105$). Children were better at passing the task if they had the same preference as the protagonist compared to when they had a different preference (Wilcoxon test, $N = 58$, $Z = -3.452$, $p = .001$, $r = .453$), see Figure 9. For the means of performance, see Figure 10. Children who had the same preference as the protagonist passed the task significantly above chance level, whereas children who had a different preference compared to the protagonist passed the task significantly below chance level (Wilcoxon tests, $N = 58$, $Z = -3.452$, $p = .001$, $r = .453$).

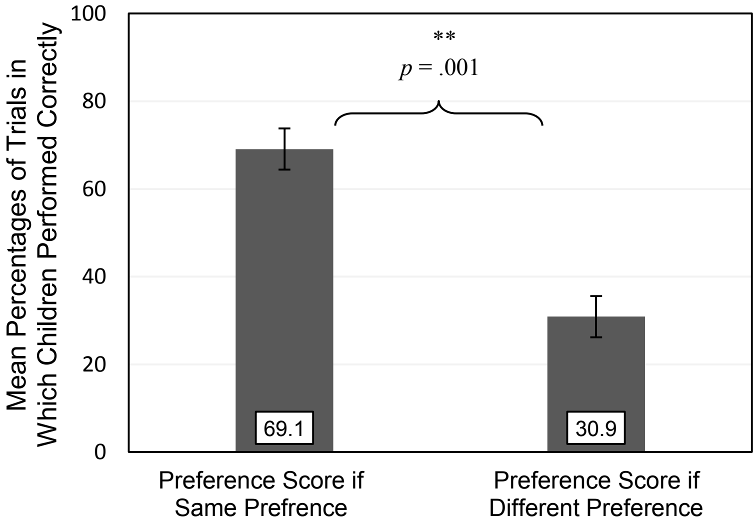


Figure 9. Mean percentages of trials in which children performed correctly on the preference understanding tasks comparing the conditions if they either had the same preference as the protagonist, or if they had a different preference to the protagonist.

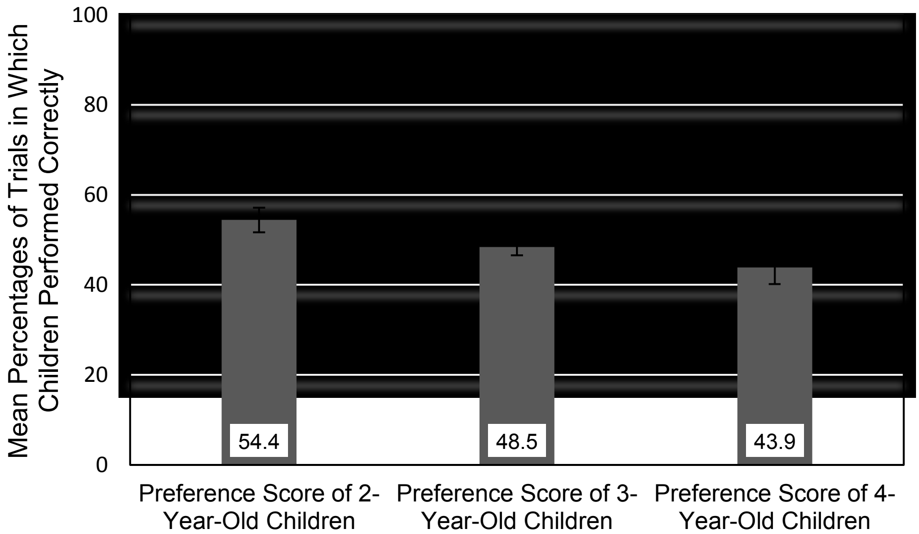


Figure 10. Mean percentages of trials in which children performed correctly on the preference understanding tasks.

False-belief-understanding task – 2-year-old children. Children did not pass the false-belief unexpected-content task (binominal test, $N = 57$, $p = .791$). They showed no difference in performance between their first touch and the actually given object (McNemar test, $\chi^2(57)$

= 35.581, p =.687). For the means of performance, see Figure 13. In the false-belief unexpected-transfer task, children significantly failed (binominal test, $N = 74$, $p = .007$). They also showed no difference in performance between their first touch and the actually opened box (McNemar test, $\chi^2(74) = 74.000$, $p = 1.000$). For the means of performance, see Figure 14. For both tasks, no order effects were found.

False-belief-understanding task – 3-year-old children. Children failed the false belief unexpected-content task regarding the self-question (binominal test, $N = 75$, $p = .005$) and the other-question (binominal test, $N = 69$, $p < .001$). Children answered the self-question better than the other-question (McNemar test, $\chi^2(68) = 13.505$, $p = .004$, $r = 1.638$), see Figure 11. For the means of performance on the other-question, see Figure 13.

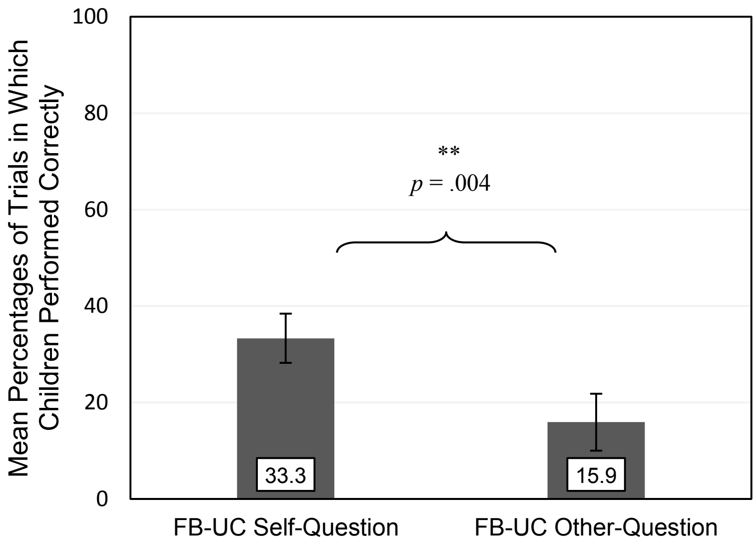


Figure 11. Mean percentages of trials in which 3-year-old children performed correctly on the false belief unexpected content tasks comparing the self-question and the other-question.

Children failed the unexpected-transfer task regarding the touch score (binominal test, N = 84, p = .021) and the actually given object score (binominal test, N = 84, p = .038). They showed no difference in performance between their first touch and the actually opened box (McNemar test, $\chi^2(84) = 56.829$, p = 1.000). Analyses showed that children did not choose the object they themselves preferred (binominal test, N = 81, p = .119). Their preference for one object did not promote passing or failing the task (McNemar test, $\chi^2(80) = .860$, p = .479). For both tasks, no order effects were found. For the means of performance, see Figure 14.

False-belief-understanding task – 4-year-old children. In the false-belief unexpected content-task, children significantly passed the self-question (binominal test, N = 69, p < .001), but not the other-question (binominal test, N = 71, p = .342). Children performed significantly better on the self-question than on the other-question (McNemar test, $\chi^2(69) = 12.971$, p < .001, r = .434), see Figure 12.

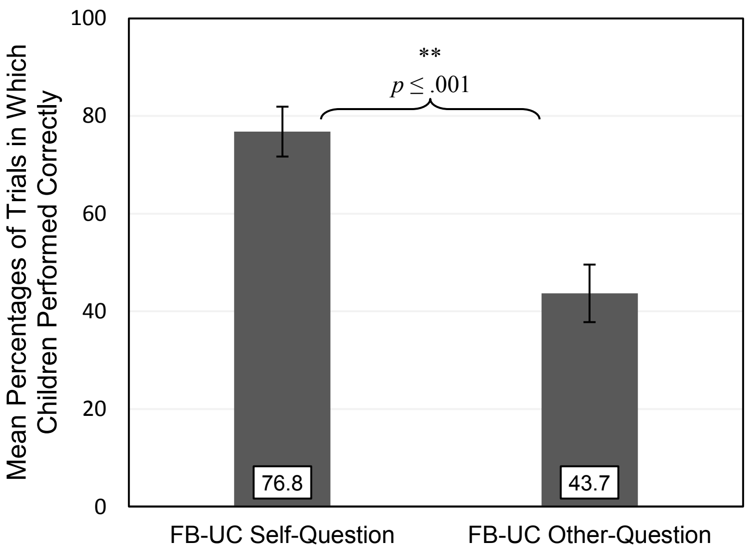


Figure 12. Mean percentages of trials in which 4-year-old children performed correctly on the false belief unexpected content tasks comparing the self-question and the other-question.

Further analyses revealed an order effect: Children were more likely to pass the self-question if it was asked before the other-question. This might be explained by either limited memory capacities or by confusion after they had to think carefully about the other person's mind. For the mean performance on the other-question, see Figure 13. For the false-belief unexpected-transfer task, the total score revealed that children correctly answered the test question below chance level (Wilcoxon test, $N = 75$, $Z = -2.288$, $p = .022$, $r = .264$). For the means of performance, see Figure 14.

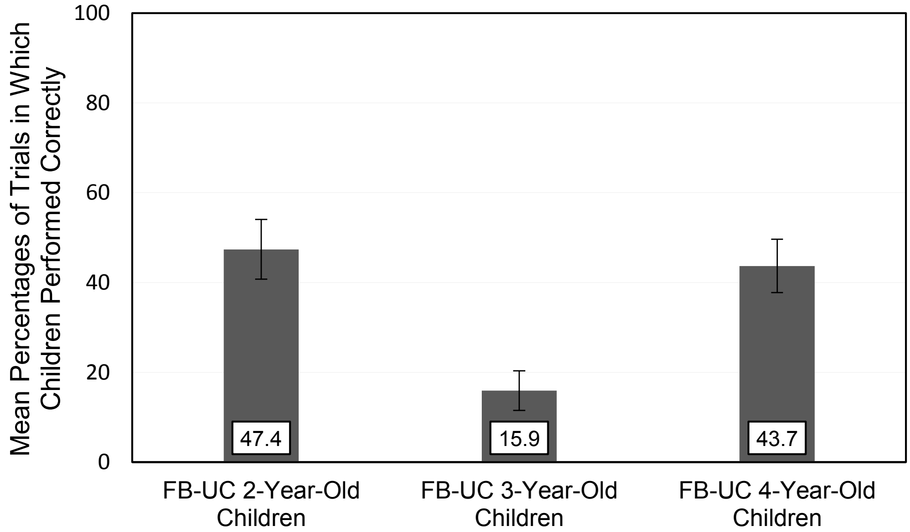


Figure 13. Mean percentages of trials in which children performed correctly on the false belief unexpected content tasks (FB-UC = false-belief unexpected-content score)

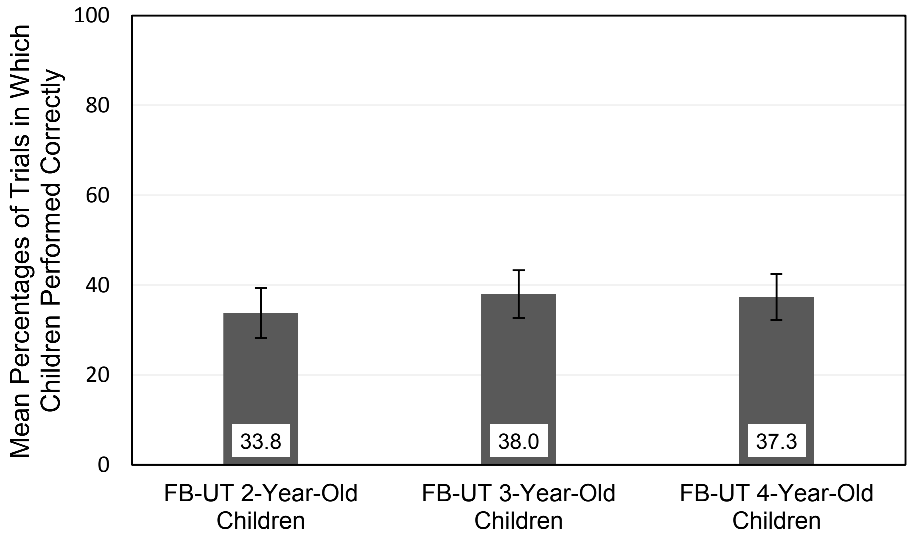


Figure 14. Mean percentages of trials in which children performed correctly on the false belief unexpected transfer tasks (FB-UT = false-belief unexpected-transfer)

Tasks of the IC Batteries

Inhibition-of-concept task – 2-year-old children. Children passed the training phase by performing at above chance level (Wilcoxon test, $N = 79$, $Z = -6.314$, $p < .001$, $r = .713$), whereas performance in the test phase was not above chance level (Wilcoxon test, $N = 77$, $Z = -.069$, $p = .945$). When comparing performances on the first block of trials with that on the second block of trials no difference was found (Wilcoxon test, $N = 72$, $Z = -.433$, $p = .665$). No effect was found concerning the position of the buckets. For the means of performance, see Figure 15.

Inhibition-of-concept task – 3-year-old children. Overall, children performed at chance level (Wilcoxon test, $N = 76$, $Z = -.735$, $p = .462$). There was a decrease in successful performance when comparing the performance in the first block of trials (74.02% of trials correct) with the performance in the second block of trials (60.94% of trials correct) (Wilcoxon test, $N = 64$, $Z = -3.268$, $p = .001$, $r = .409$). No effect was found concerning the position of the cards. For the means of performance, see Figure 15.

Inhibition-of-concept task – 4-year-old children. A difference between the conservative and non-conservative version was detected. Children scored higher in the non-conservative rating than in the conservative rating (Wilcoxon test, $N = 73$, $Z = -3.681$, $p < .001$, $r = 0.428$). Since this task was administered to investigate the ability to inhibit a specific concept independent of which word was used to refer to the concept, for further analysis the non-conservative scoring was used. Children performed better within the first half of trials (82.40% of trials correct) compared to the second half of trials (68.99% of trials correct) (Wilcoxon test, $N = 48$, $Z = -3.262$, $p = .001$, $r = 0.471$). For the means of performance, see Figure 15.

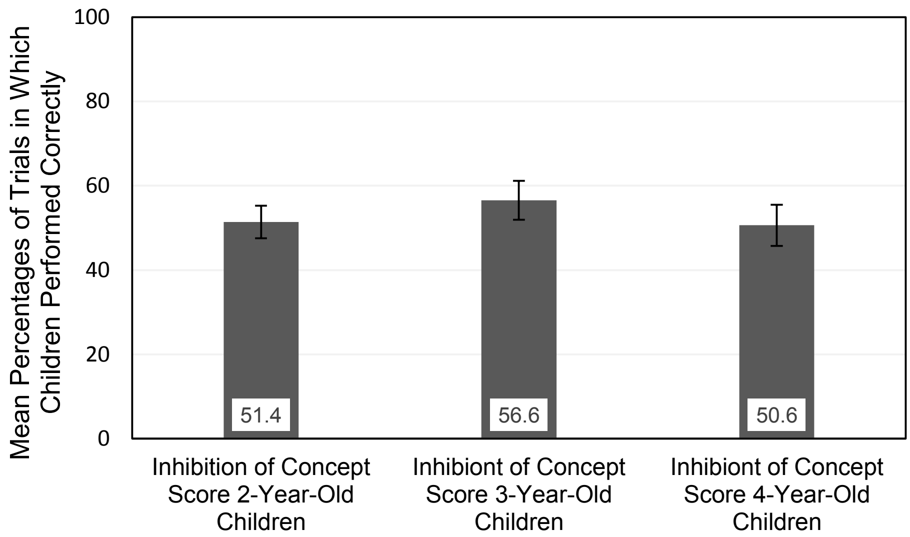


Figure 15. Mean percentages of trials in which children performed correctly on the inhibition-of-concept tasks.

Spatial-inhibition task – 2-year-old children. Overall, children failed the task (Wilcoxon test, $N = 77$, $Z = -2.670$, $p = .008$, $r = .304$). Comparison of performances in the first block of trials with performances in the second block of trials showed no difference (Wilcoxon test, $N = 74$, $Z = -.112$, $p = .910$). No order effect was found. For the means of performance, see Figure 16.

Spatial-inhibition task – 3-year-old children. Eleven children were excluded of the data set because of not understanding the rules of the task. Overall, children failed the task (Wilcoxon test, $N = 73$, $Z = -6.113$, $p < .001$, $r = .715$). Comparison of performance in the first block of trials with that in the second block of trials showed no difference (Wilcoxon test, $N = 72$, $Z = -.898$, $p = .369$). For the means of performance, see Figure 16.

Spatial-inhibition task – 4-year-old children. Nine children were excluded of the data set because of not understanding the rules of the task. Overall, children did not perform above chance level (Wilcoxon test, $N = 66$, $Z = -.831$, $p = .406$). When dividing the task into two blocks, children scored better in the second block of the trials (47.97% of trials correct) than in the first block (62.50% of trials correct) (Wilcoxon test, $N = 66$, $Z = -4.540$, $p < .001$, $r = 0.559$). For the means of performance, see Figure 16.

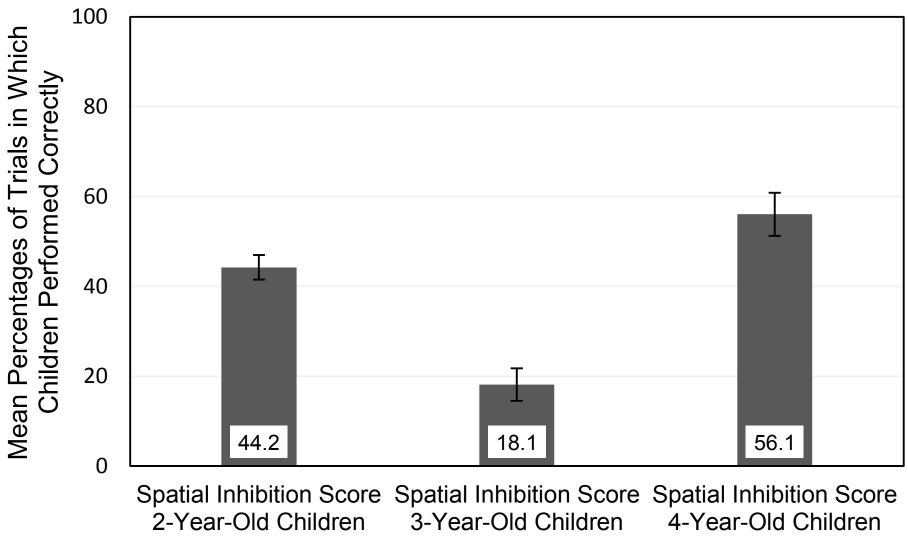


Figure 16. Mean percentages of trials in which children performed correctly on the spatial inhibition tasks.

Delay-of-gratification task – 2-year-old children. To analyze if the duration of trials influenced children's waiting performance, I analyzed whether children were more likely to wait during short trials (including trials from 5 to 20 s) than during long trials (including 35 to 50 s). Children performed better in the short trials than in the long trials (Wilcoxon test, $N = 77$, $Z = -3.415$, $p = .001$, $r = .389$). For the means of performance of waiting time, see Figure 17.

Delay-of-gratification task – 3-year-old children. As the procedure was slightly changed after the first few testing sessions, I checked for differences between the two versions of the task. The first version of the task was performed with only one piece of sweets (one gummy bear/one piece of chocolate), whereas the second version was performed with several pieces of sweets (a small bag of gummy bears/a bar of chocolate scattered over the plate). Mann-Whitney's *U*-tests revealed one statistically significant difference: Children waited longer before touching the sweets when they were facing only one piece of sweets ($U = 217.5$, $N_{first} = 10$, $N_{second} = 68$, $Z = -1.971$, $p = .049$, $r = 0.228$). Therefore, only children who had received the second version of the task were included in the final analyses. The total number of participants excluded from the data set was 22. This number also includes the children who did not attend the second session. For the means of performance of waiting time, see Figure 17.

Delay-of-gratification task – 4-year-old children. For the means of performance of waiting time, see Figure 17.

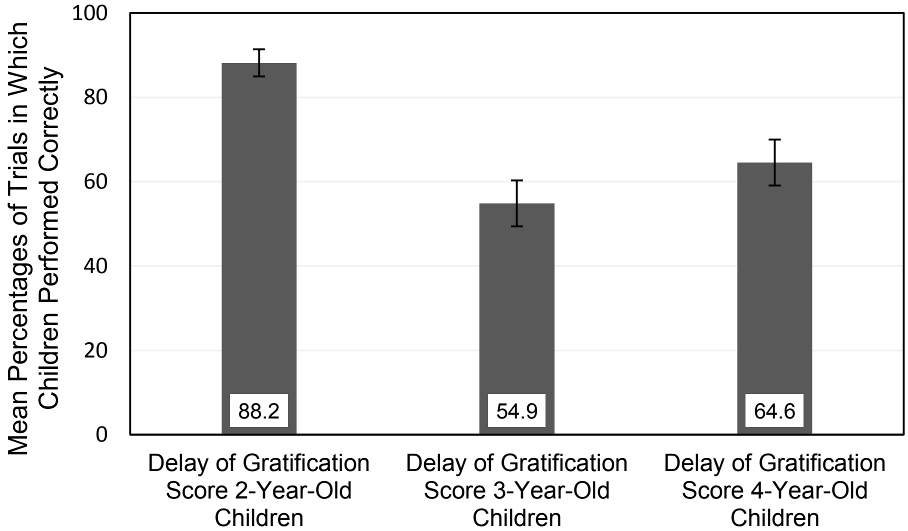


Figure 17. Mean percentages of trials in which children performed correctly on the delay of gratification tasks.

Ehrenwörtliche Erklärung

„Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Ich habe nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit oder Teile davon wurden bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde als Dissertation vorgelegt. Ferner erkläre ich, dass ich nicht bereits eine gleichartige Doktorprüfung an einer Hochschule endgültig nicht bestanden habe.“