Early Social Cognition: Development and Influences

Frühe soziale Kognition: Entwicklung und Einflussfaktoren

Dissertation

zur Erlangung des akademischen Grades eines

Doktors der Philosophie

der Philosophischen Fakultät III

der Universität des Saarlandes

vorgelegt von

Daniela Mink

aus Völklingen

Saarbrücken, 2016

Dekan:

Prof. Dr. Roland Brünken, Universität des Saarlandes

Berichterstatterinnen:

Prof. Dr. Gisa Aschersleben, Universität des Saarlandes

Prof. Dr. Julia Karbach, Goethe Universität Frankfurt

Tag der Disputation: 02. November 2015

This manuscript contains text passages, figures and tables out of scientific papers of which I am the first author that are already published or have been submitted for publication in peer reviewed journals. These parts are not explicitly identified. Part of the material is already published in the following papers:

Mink, D., Henning, A., & Aschersleben, G. (subm). Social-cognitive development from infancy to toddlerhood.

Mink, D., Henning, A., & Aschersleben, G. (2014). Infant shy temperament predicts preschoolers Theory of Mind. *Infant Behavior and Development*, *37*(1), 66-75.

doi: 10.1016/j.infbeh.2013.12.001

Mink, D., Henning, A., & Aschersleben, G. (2013). Visual habituation tasks: The role of infant temperament. *Infant Behavior and Development*, *36*(3), 377-390.

doi: 10.1016/j.infbeh.2013.03.005

In the following, I will use the term *we* when referring to these studies and results that I conducted as co-author with Gisa Aschersleben and Anne Henning.

Danksagung

Ich möchte an dieser Stelle all den Menschen danken, die mich während meiner Promotionszeit begleitet und bestärkt haben und zum Gelingen dieser Arbeit beigetragen haben.

Als erstes möchte ich mich bei Gisa Aschersleben bedanken, die mich in den letzten Jahren fortwährend unterstützt und gefördert hat und deren Tür immer offen stand. Durch sie hatte ich bereits früh die Möglichkeit Teile unserer Forschung national wie international zu präsentieren und zu veröffentlichen und letztlich diese Arbeit zu verfassen. Ein besonderer Dank gilt Anne Henning für ihre Ratschläge und Unterstützung von Anfang bis Ende, selbst noch aus der Ferne. Danke für alles, auch für das Korrekturlesen dieser Arbeit und die nette Reisebegleitung auf vielen Konferenzen und Tagungen!

Auch den restlichen Mitarbeiterinnen am Lehrstuhl Entwicklungspsychologie gilt mein Dank. Hervorheben möchte ich Marion Klein, Christiane Mink und Kristin Welsch, die nacheinander das Kinderlabor und all die Studientermine hervorragend organisiert haben. Vielen Dank an die zahlreichen Hiwinen, Hiwis und Diplomandinnen, die mich bei der Durchführung und Auswertung der einzelnen Studienaufgaben tatkräftig und unermüdlich unterstützt haben. Ganz herzlichen Dank an meine Mitstreiterinnen Katarzyna Hoffmann und Corina Möller für die gegenseitige Unterstützung und Motivation sowie Ablenkung, wenn sie nötig war. Danke für die schöne Zeit im Büro!

Besonderer Dank gilt ebenfalls den über 160 Kindern und deren Eltern, ohne deren beständige Teilnahme dieses Längsschnittprojekt nicht hätte realisiert werden können. Danke an die Deutsche Forschungsgemeinschaft für die Finanzierung des Projektes (As79/6-1 “Mütterlicher Interaktionsstil und sozial-kognitive Entwicklung in den ersten beiden Lebensjahren“).

Abschließend möchte ich mich herzlich bei meiner Familie und meinen Eltern bedanken, die mich immer unterstützt und dazu ermuntert haben, meinen Weg zu gehen und meine Ziele zu erreichen.

Von Herzen danke ich meinem Partner und bestem Freund Sascha Just, der mich in all meinen Vorhaben bestärkt und mir stets zur Seite steht.

Abstract

This dissertation investigated the development of early social-cognitive development in the first three years of life. While for a long time research on Theory of Mind (ToM), i.e. the ability to attribute mental states as wishes, intentions, and beliefs to self and others, has focused on preschool age, in recent years also earlier social-cognitive development has gained attention. The assumption of continuity in social-cognitive development is one question we considered in the present work. Moreover, we investigated the influence of various factors on this development, specifically the impact of temperament. In addition, we further examined the impact of temperament on general study performance in infancy. In order to assess children's temperament we used the questionnaires developed by Rothbart and colleagues as they provide a variety of parent-report questionnaires to adequately assess temperamental aspects across childhood from infancy to later ages.

In study 1, we longitudinally investigated the social-cognitive development in infants and toddlers. We tested the relation between infants' joint attention skills at 12 months and four further social-cognitive skills that emerge around the age of 18 months, namely children's pretend play behavior, their ability to recognize themselves in a mirror, to imitate an intended action, and to reason about other people's desires. Only single abilities were related to each other. Some joint attention skills, declarative pointing and detecting the experimenter's goal in a teasing task, were related to toddlers' understanding of intention-based imitation. Also, initiating joint attention and performance on a blocking task were both related to pretend solitary play. Additionally, pretend play and mirror self-recognition were related. The results of study 1 cast light on the relationship of different social-cognitive abilities in early childhood. They extend earlier findings and support the idea of continuity in social-cognitive development. Specifically, they suggest that this continuity is not global but rather task- and age-specific.

In study 2, we longitudinally investigated the relation between infant temperament at 18 months and early ToM abilities at 3 years of age. In order to do so, we assessed temperament with the Early Childhood Behavior Questionnaire (ECBQ) and ToM by examining children's understanding of divergent desires and beliefs, and of knowledge access. Recent research (Lane et al., 2013; Wellman et al., 2011) indicated a developmental link between specific childhood temperament and Theory of Mind abilities. This idea based on the emotional reactivity hypothesis (Hare, 2007) that originally accounts for social-cognitive capacities in canines. The results obtained in study 2 are in line with such a social-emotional reactivity perspective postulating more sophisticated ToM abilities for children with less reactive more observant temperament. Children with shy temperament at 18 months and at 3 years were better in reasoning about others' mental states at age 3. Findings indicate that temperament is related to ToM earlier in development than previously found, and that this relation is thus not unique to false belief understanding.

In study 3, we longitudinally investigated the relation between infant temperament and dropout rate in two visual habituation tasks when infants were 6 and 12 months of age. At both age points, infant temperament was assessed with the Infant Behavior Questionnaire Revised (IBQ-R) and infants were presented with two habituation tasks that were similar in set-up and procedure but different in content. Consistent with previous work, dropout rates in the habituation tasks were very high and we investigated if this dropout was systematically influenced by infant's temperament. Overall, only few temperamental traits, especially the ability to attend to something for an extended time, had an impact on dropout rate. This suggests that the relatively high dropout rates reported in infant looking time studies are not systematically related to infant temperament. However, findings also suggest that temperament might have an impact on the likelihood of dropout when a habituation task is conducted at the end of a longer test session.

To summarize, the findings of the present work partially support the assumption of continuity in social-cognitive development. Yet, this assumption seems to do not apply generally on all ages and tasks used to investigate social-

cognitive abilities emerging in the first three years of life. Besides, this dissertation provides further evidence for an early influence of temperament on children's social-cognitive development. Nevertheless, temperament seems not to exert biasing influence on study performance, an important finding especially for infant studies using habituation tasks.

Zusammenfassung

Das Ziel der vorliegenden Dissertation war es, die Entwicklung der frühen sozialen Kognition in den ersten drei Lebensjahren zu untersuchen. Lange Zeit hat sich die Forschung zum Thema *Theory of Mind*, also die Fähigkeit sich selbst und anderen Personen mentale Zustände wie Wünsche, Intentionen und Glauben zuzuschreiben, auf das Kindergartenalter fokussiert. In den letzten Jahren wurde das Augenmerk auch auf die frühe sozial-kognitive Entwicklung im Kleinkindalter gerichtet. Eine Frage, der wir in der vorliegenden Arbeit nachgekommen sind, ist die Annahme von Kontinuität in der sozial-kognitiven Entwicklung. Darüber hinaus haben wir den Einfluss verschiedener Faktoren auf diese Entwicklung untersucht, vor allem den Einfluss von Temperament. Zusätzlich haben wir den Einfluss von Temperament auf die generelle Studienleistung von Kindern im Kleinkindalter untersucht. Um das kindliche Temperament zu untersuchen, haben wir Elternfragebögen eingesetzt, die von Rothbart und ihren Kollegen entwickelt wurden. Diese Forschergruppe stellt eine Vielzahl an Elternfragebögen bereit um adäquat die unterschiedlichen Aspekte des kindlichen Temperaments vom Kleinkind- bis ins Schulalter zu untersuchen.

In Studie 1 haben wir längsschnittlich die sozial-kognitive Entwicklung von Säuglingen und Kleinkindern untersucht. Dazu haben wir die Beziehung zwischen den Joint Attention Fähigkeiten im Alter von 12 Monaten und vier weiteren sozial-kognitiven Fähigkeiten, die sich etwa im Alter von 18 Monaten zeigen, untersucht. Dazu gehören das Als-ob-Spiel sowie die Fähigkeiten sich selbst im Spiegel zu erkennen, eine Handlung zu imitieren und Schlüsse zu ziehen über die Wünsche einer anderen Person. Beziehungen zwischen diesen Fähigkeiten konnten nur vereinzelt gefunden werden. Einige der Joint Attention Fähigkeiten, nämlich deklaratives Zeigen und das Entdecken der Absicht eines Versuchsleiters in einer das Kind neckenden Aufgabe, hingen zusammen mit der korrekten Imitation von Handlungen, in denen die Absicht nicht vollständig dargeboten wurde. Darüber hinaus hingen das Initiieren von Joint Attention und das Verständnis für das Ziel des Versuchsleiters in einer

das Kind behindernden Aufgabe beide zusammen mit dem Niveau des Als-ob-Spiels, wenn das Kind alleine spielte. Eine weitere Relation wurde zwischen dem Als-ob-Spiel des Kindes und seiner Fähigkeit, sich selbst im Spiegel zu erkennen gefunden. Die Ergebnisse von Studie 1 beleuchten die Beziehung verschiedener sozial-kognitiver Fähigkeiten in der frühen Kindheit. Sie bauen bisherige Befunde aus und unterstützen die Idee von Kontinuität in der sozial-kognitiven Entwicklung. Insbesondere lassen unsere Ergebnisse vermuten, dass diese Kontinuität nicht global, sondern eher aufgaben- und altersspezifisch ist.

In Studie 2 haben wir längsschnittlich die Beziehung zwischen kindlichem Temperament im Alter von 18 Monaten und den frühen Theory of Mind Fähigkeiten im Alter von 3 Jahren untersucht. Zu diesem Zweck haben wir das Temperament der Kinder mit dem Early Childhood Behavior Questionnaire (ECBQ) erhoben. Die ToM Fähigkeiten der Kinder haben wir mit den Aufgaben Abgrenzung des eigenen Wunsches und der eigenen Überzeugung sowie Zugang zu Wissen überprüft. Jüngste Forschungen (Lane et al., 2013; Wellman et al., 2011) deuten auf eine Beziehung zwischen spezifischen, kindlichen Temperamentseigenschaften und Theory of Mind Fähigkeiten hin. Diese Idee fußt auf der Emotionalitäts-Reaktivitäts-Hypothese (Hare, 2007) die ursprünglich aufgestellt wurde um die sozial-kognitiven Fähigkeiten von Hunden erklären zu können. Die Ergebnisse aus Studie 2 entsprechen einer solchen Perspektive, die annimmt, dass Kinder mit einem weniger reaktiven, eher beobachtenden Temperament über fortgeschrittene Theory of Mind Fähigkeiten verfügen. Kinder, die mit 18 Monaten und 3 Jahren als schüchtern eingeschätzt wurden, waren im Alter von 3 Jahren besser darin sich in andere Personen hineinzuversetzen. Diese Befunde deuten darauf hin, dass Temperament in Beziehung zu ToM steht, und zwar bereits früher in der Entwicklung als bisher angenommen. Die Beziehung zwischen Temperament und ToM ist daher nicht spezifisch für das Verständnis von false belief Aufgaben.

In Studie 3 haben wir längsschnittlich die Beziehung zwischen kindlichem Temperament und der Abbruchrate in zwei visuellen Habituationsaufgaben im Alter von 6 und 12 Monaten untersucht. Das

Temperament der Kinder haben wir zu beiden Messzeitpunkten mit dem Infant Behavior Questionnaire Revised (IBQ-R) erhoben. Zusätzlich haben wir den Säuglingen zwei Habituationsaufgaben präsentiert, die sich in Aufbau und Ablauf ähnelten, aber inhaltlich verschieden waren. Die Abbruchrate in unseren Habituationsaufgaben war – übereinstimmend mit vorheriger Forschung – sehr hoch und wir haben untersucht, ob diese Abbruchrate systematisch durch das Temperament der Kinder beeinflusst wurde. Insgesamt hatten nur wenige Temperamentseigenschaften einen Einfluss auf die Abbruchrate. Die Fähigkeit, sich für eine längere Zeit einer bestimmten Sache zu widmen, könnte einen Einfluss nehmen. Dies lässt vermuten, dass die relativ hohe Abbruchrate, die in Blickzeitstudien mit Säuglingen üblicherweise berichtet wird, nicht systematisch mit dem Temperament der Säuglinge zusammenhängt. Allerdings lassen unsere Ergebnisse auch vermuten, dass gewisse Temperamentseigenschaften die Wahrscheinlichkeit eines Abbruchs erhöhen, wenn z. B. die Habituationsaufgabe erst am Ende einer längeren Testreihe durchgeführt wird.

Zusammenfassend lässt sich sagen, dass die Ergebnisse der vorliegenden Arbeit die Annahme einer Kontinuität in der sozial-kognitiven Entwicklung zumindest teilweise unterstützen. Jedoch scheint diese Annahme nicht auf alle Alterszeitpunkte und Aufgaben zuzutreffen, die verwendet werden um die frühen, sich in den ersten 3 Lebensjahren entwickelnden, sozial-kognitiven Fähigkeiten zu untersuchen. Abgesehen davon liefert diese Dissertation Evidenz für einen frühen Einfluss von Temperament auf die sozial-kognitive Entwicklung von Kindern. Temperament scheint jedoch die Leistung bzw. das Durchhaltevermögen von Kindern in Studien nicht zu verzerren, was ein wichtiger Befund vor allem für Säuglingsstudien darstellt, da diese häufig Habituationsaufgaben anwenden.

Table of Contents

List of Abbreviations

AF	Attention Following
BoH	Back-of-Hand
BR	Behavior Request
Caus	Causality
CBCL	Child Behavior Checklist
CBQ	Children's Behavior Questionnaire
ECBQ	Early Childhood Behavior Questionnaire
ELFRA	Elternfragebögen für die Früherkennung von Risikokindern
GF	Gaze Following
IBQ	Infant Behavior Questionnaire
IBQ-R	Infant Behavior Questionnaire- Revised
JA	Joint Attention
OS	Object Spectacle
PF	Point Following
SETK	Sprachentwicklungstest für Kinder
ToM	Theory of Mind

List of Tables

List of Figures

Chapter 1: Introduction

1 Social Cognition

From childhood on, fairy stories are familiar to us all. When reading them to children, parents might not be aware of the complex demands they make in terms of understanding the protagonists' mental states such as intentions, beliefs and feelings (Hinchcliffe, 1996). Just the social cognitive content of a fairy tale leads to an understanding why the story characters behave in the described manner and only when appreciating the protagonists' mental states their behavior makes sense (ibid). The well-known fairy tale of Little Red Riding Hood gives an example of how the ability to impute mental states is crucial for children either to fully grasp the meaning of the story or to only understand it as a succession of behavioral events (Lillard, 1997). In first instance, Little Red Riding Hood does not know that the wolf has eaten her grandmother and she thinks that her grandmother is lying in the bed. Even when children understand that Little Red Riding Hood is unaware of the presence of the wolf, at the age of 5 and 6 years some of them still ascribe her feelings of being afraid (Bradmetz & Schneider, 1999). This shows that understanding all facets of mind is a very complex ability that does not develop in short time. Beginning with the example of understanding fairy tales I will now describe the role social-cognitive understanding plays in children's daily life.

Social cognition concerns our understanding of people and their doings. As Flavell (1985) summarizes it includes "thinking and knowledge about the self and others as individuals, about social relations between people, about social customs, groups, and institutions" (p.159). In general, social cognition concerns reasoning about social world as opposed to "physical and logical-mathematical" world (p.119). During social-cognitive development children acquire the ability to recognize that they and other people perceive, think, and feel. Then, they begin to recognize that other's perspectives may be different from their own and potentially inferable from the other's perceptual experiences. Also, children recognize that thoughts might be recursive, that is,

one thought could have another thought as object. So, they were eventually able to build complex trains of thought and represent them. Much everyday interaction and communication seems to presuppose this kind of knowledge (Flavell, 1985).

To effectively interact with other people already in childhood it is essential to understand the mental states of others. The ability to attribute mental states such as beliefs, intentions, and desires to oneself and other people and to understand that actions are causally related to these mental states is commonly defined as Theory of Mind (ToM). This term traces back to Premack and Woodruff (1978a). The authors postulate that attributing mental states works as a theory as such "states are not directly observable but need to be inferred like theoretical terms in science" (Perner, 1999) and as these, inferred mental states can be used to appropriately predict other's behavior. To test if chimpanzees are able to infer such mental states to others, too, Premack and Woodruff (1978a) conducted the following study: They showed an adult chimpanzee a variety of videotapes displaying a human actor facing a problem, as for example being locked in a cage and trying to get out. For each problem, the chimpanzee had to choose the solution from different photographs (e.g., a key to open the cage) what the chimpanzee actually consistently did. Thus, the chimpanzee seemed to be able to infer the actor's intentions to solve the given problems. At first, Premack and Woodruff (1978a) tentatively concluded that inferences about motivation might precede inferences about knowledge, both across species and developmental stages. In fact, rather than testing if chimpanzees are able to attribute mental states to others, the described task is suited for testing their problem solving abilities (Premack & Woodruff, 1978b). Fuelled by this classical work and the question if the ability to attribute mental states to others is specific for human cognition, ToM development has become a widely researched topic in developmental psychology (see Sodian & Thoermer, 2006, for a review). Between 3 and 5 years of age, children develop an explicit understanding of the causal relation between mental states and actions and therefore are able to correctly predict other's actions. Around the age of 4 years, they understand that a belief about a state of affairs might be true or false and might therefore lead to a successful or faulty action (Wimmer

& Perner, 1983). The ability to comprehend another person's false belief is usually tested with a false-belief task. One example for the so-called change-of-location paradigm is the Maxi story: Children are told that Maxi puts his chocolate into a cupboard A. Then, that in his absence, the mother transfers the chocolate from A into cupboard B. In test, children are asked where Maxi will look for his chocolate when he returns. When children are able to represent Maxi's false belief, i.e. "The chocolate is in A.", apart from what they themselves know to be true, i.e. "The chocolate was transferred into B.", they are able to correctly indicate where Maxi will look for the chocolate when he returns (Wimmer & Perner, 1983). Hence, children need to understand that another person's mental representation is different from their own, and to additionally take this knowledge into account when predicting or explaining her behavior. In their meta-analysis, Wellman and colleagues (Wellman, Cross, & Watson, 2001) showed that this developmental change is a robust phenomenon not specific to western culture. Additionally, they reported that reducing task difficulty did not increase the performance of younger children above chance.

1.1 Early social-cognitive development

Well before children pass this false-belief task, they seem to regard the subjectivity and directedness of mental states while interpreting human behavior. While for a long time ToM research has focused on the ability to comprehend another person's false belief in preschool age, in recent years earlier social-cognitive development has gained attention, too (see Sodian & Thoermer, 2006). In this section, I will give a chronological overview about some of the social-cognitive abilities that are assumed to emerge during ToM development from infancy to early childhood.

For example, there is evidence that infants as young as 6 months of age begin to understand human actions as goal-directed (Woodward, 1998, 1999). In her seminal looking time studies, Woodward habituated 6-month-olds to an event in which a hand grasped one of two objects that were located at different sides. During test phase, object locations were switched. In one test event,

reaching properties had changed while the goal maintained the same as in habituation phase, in the other test event it was vice versa. Infants showed a stronger novelty response to events in which the goal had changed (see also Hofer, Hohenberger, Hauf, & Aschersleben, 2008). This finding indicates that they are able to represent the actor's goal. Beyond, studies applying a modified Woodward paradigm by adding a salient action effect, as for example pushing the object backwards, (Jovanovic, Király, Elsner, Gergely, Prinz & Aschersleben, 2007; Király, Jovanovic, Prinz, Aschersleben & Gergely, 2003; for an overview see Aschersleben, 2006) showed that 6- to 10-month-olds are able to interpret even unfamiliar actions as goal-directed. The ability to understand human actions as goal-directed is potentially one of the first social-cognitive abilities in children's development. It is assumed to be related to preschoolers' ability to attribute mental states to others.

At the end of the first year of life, infants understand gaze and pointing gestures as goal-directed and they are able to share attention with another person in joint play (Tomasello, 1995). Tomasello and colleagues (Tomasello, Kruger, & Ratner, 1993) argue that understanding of others as intentional agents, whose attention and behavior can either be followed or directed to third entities, is essential for infants' early skills of joint attention. Carpenter and colleagues (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998) longitudinally observed infants' early skills of social cognition and communication from 9 to 15 months of age at monthly intervals. They found that the considered skills followed a common order of emergence, namely share attention, follow attention/behavior, direct attention/behavior. From simply looking to an adult's face to share her attention, infants have to take into account what an adult is attending to in a relatively distal space in order to follow or direct her attention (ibid.). In addition, they begin to use an adult's social cues to decide how to behave in emotional situations, i.e. social referencing, and they begin to manipulate objects in the same manners as adults do through imitative learning.

By 18 months of age, children differentiate between their own and another person's opposite desire (Repacholi & Gopnik, 1997). After observing

an experimenter expressing disgust while tasting one food (e.g. crackers) and expressing happiness while tasting the other food (e.g. broccoli), children were asked to give her something to eat out of the two food bowls on a tray. 18-month-olds were able to correctly infer the experimenter's desire for the food she had prior associated with a happy facial expression. They gave her the preferred food (e.g. broccoli) even when this choice differed from their own desire. Children did not act egocentrically and besides, Repacholi and Gopnik (1997) could rule out that children simply gave the experimenter the food they themselves did not want. Hence, children at 18 months are not only able to infer a desire to another person from anterior experience. They also understand that desires are related to the person's emotions and that desires might subjectively differ.

Also at the age of 18 months, children infer the goal of an action, even if they only observe a failed attempt instead of the intended action (Bellagamba & Tomasello, 1999; Meltzoff, 1995). For example, children saw an experimenter trying to pull a dumbbell apart but her hands slipped off and she failed. When handed the toy after the experimenter's demonstration, children rather produced the intended target action than imitated the observed failed attempt. Only by observing the experimenter trying but failing to perform a certain action on the object, children were able to understand her intention. In contrast, they did not tend to produce the intended action when the same failed attempt was demonstrated by a mechanical device. Hence, Meltzoff (1995) concludes that 18-month-olds' attributions of intentions are restricted to social agents.

At 18 months, children also engage in pretend play (Leslie, 1994; Piaget, 1962). According to Piaget (1962), pretend play – as well as deferred imitation and language – indicate the development of symbolic function that enables children to elaborate on mental content separately from reality (see Lillard et al., 2013). Thus, pretense seems to be "an early manifestation of the ability to understand mental states including one's own as well as another's" (Lewis & Ramsay, 2004, p. 1821). Lillard (1993) defines pretend play as "the projection of a supposed situation onto an actual one" (p. 349) with the purpose

of having fun. It can involve the imagination of an object where there is nothing at all or the substitution of an object as if it were another. Essential for pretense is that the pretending person and the potential playmates know that the pretended situation is different from reality and that a pretended identity is not bound by the features of an object (Harris & Kavanaugh, 1993). At the age of 18 months, children begin to appropriately respond to an experimenter's request based on the actual "make-believe stipulation" (Harris & Kavanaugh, 1993, p. 30). Children acted differently on a single prop depending on its temporary identity, for example, using a stick first as a spoon for stirring tea and then as a toothbrush. Hence, children at this age understand that a temporary pretend identity is appropriate only for a given context.

In addition, 18-month-olds recognize themselves in a mirror reflection (Lewis, Sullivan, Stanger, & Weiss, 1989). Pretending to wipe the children's noses mothers applied some rouge on them. While placed in front of a mirror, the children's behavior towards this mark was observed. Nose respectively spot touching serves as index for self-referential behavior (Lewis et al., 1989) and serves as a measure of a self-meta-representation (Lewis & Ramsay, 2004). Gallup (1991) postulates that self-recognizers are capable of becoming the object of their own attention. Through this introspection, self-recognizers have access to their own mental states and therefore have intuitive access to the mental states of others. To take into account that the ability to recognize oneself in the mirror does not suddenly emerge at the age of 18 months but undergoes a gradual transition, Bischof-Köhler (1994) distinguishes between non-recognizing children, children in transition and recognizing children. The group of children in transition shows ambiguous behavior that falls into two categories: those children who don't perceive the spot mainly show avoidance to prevent eye contact to the mirror image. Those who perceive the spot try to catch it in the mirror, search behind the mirror or treat the reflection as a partner. Children who already recognize themselves often experiment with the mirror e.g. monitoring while grimacing. This phase of transition indicates that the onset of mirror self-recognition underlies a broad variance. It can be assumed that this is true also for the other social-cognitive abilities but only few tasks allow the observation of progression.

Wellman and Liu (2004) assume that ToM also develops on a continuum. To adequately trace the social-cognitive development in early childhood, they constructed a ToM battery with increasing degree of difficulty assessing a range of different developmental attainments in children aged 3 to 6 years. The scale consists of six tasks with two of them testing children's false belief abilities. The tasks are described from least to most difficult: In the Divergent Desires task, the child needs to differentiate his or her own desire (e.g., preference for a cookie over a carrot) from another person's differing desire about the same food items (e.g., preference for a carrot) to correctly predict the other person's snack choice (e.g., the carrot and not the cookie). Whereas already 18-month-olds are able to infer subjective desires to others (Repacholi & Gopnik, 1997), in addition, 3-year-olds are able to predict and interpret other people's behavior based on the understanding of their desires (Sodian & Thoermer, 2006). In the Diverse Beliefs task, the child needs to differentiate his or her own belief about the location of an object (e.g., the cat is hiding in the bush) from another person's differing belief about the location of same object (e.g., the cat is hiding in the garage) to correctly predict the other person's action (e.g., look for the cat in the garage). The Knowledge Access task requires an understanding of the causal relation between seeing and knowing independent of the own visual access to an object. In this task, the child is first shown the content of a box (toy dog) and then asked to judge whether another person who did not have visual access to the content of the box, knows its content. In the Content False-Belief task (or unexpected-content task), another classic false-belief task beside the above-mentioned change-of-location task, the child needs to differentiate his or her own true belief from another person's false belief about the content of a box. The child is first shown an unexpected content of a candy box (toy pig instead of chocolate beans) and then asked to judge another person's belief who did not have visual access to the actual content. In the Explicit False-Belief task, the child is told the correct location of an object (e.g., the gloves are in the backpack) and that another person expects it in a different location (e.g., in the closet). To accurately predict that the other person will look for the object in the wrong location (closet) the child needs to understand that a false belief leads to faulty actions. Finally, in the Real Apparent Emotion task, a pretest controls for the

child's ability to differentiate sad, happy and neutral facial expressions. In the task, the child needs to understand that a displayed emotion in the face (e.g., happiness) might differ from the emotion a person really feels (e.g., sadness) in order to hide the true emotion from a counterpart (see also Henning, Spinath, & Aschersleben, 2011 for task descriptions). Several studies showed that the children's abilities increase with advancing age and that they can understand people's desires, intentions, and ignorance well before they understand false-beliefs (Kristen, Thoermer, Hofer, Aschersleben, & Sodian, 2006; Wellman & Liu, 2004). By 3 years of age, children refer to desires and true belief in predicting another person's action and begin to understand that visual access leads to knowing about the content of a box. Only at the age of 4 to 5 years, children are able to pass the false-belief tasks (e.g., Kristen et al., 2006).

Similarly to ToM, also successful lying develops between 3 and 5 years of age (Sodian & Thoermer, 2006). Lying in children is often studied with the temptation resistance paradigm (Talwar & Lee, 2008). Children were told not to peek at a toy while left alone. In their classical study, Lewis and colleagues (Lewis, Stanger, & Sullivan, 1989) examined 3-year-olds' capacity for deception. Indeed, the majority of children used verbal deception: they peeked but did not admit their transgression when asked. Around the age of 4 years, children begin to consider the mental state of their counterpart when lying (see Talwar & Lee, 2008, for a model of lying development). The ability to lie seems to be strongly related to and maybe fostered by ToM development (Evans & Lee, 2013; Talwar & Lee, 2008). Newton and colleagues (Newton, Reddy, & Bull, 2000), however, claim that young children's lies are adaptive social strategies, for example to avoid negative consequences, that rather lead to an understanding in mental states of others than being based on ToM understanding.

These studies show that a child's understanding of subjectivity is already "progressively broadening and developing" (Wellman & Liu, 2004, p. 536) in early childhood. In sum, already in their first 3 years of life and well before their ToM is fully developed, children know a lot about what is going on in another person's mind. It is assumed that the early social-cognitive

abilities are related to children's later ToM. In section 1.2, I will give an overview about the most prominent theories of social-cognitive development and then report empirical evidence that supports this assumption in section 1.3.

1.2 Theoretical aspects of social-cognitive development

There exist various theories to explain the mechanisms involved when reasoning about the mental states of others. In the developmental psychology literature, theory theory is supposed to best explain how ToM abilities are acquired (Mahy, Moses, & Pfeifer, 2014).

1.2.1 Theory Theory

Theory theorists postulate that people have a naïve folk psychology trying to construct everyday theories to explain their observations and experiences. There are three theory characteristics that also apply to children's understanding of mental states: (1) children's theories involve characteristic explanations for unobservable entities, (2) they include also incorrect predictions as well as predictions about behavior the children have never experienced themselves and (3) they lead to distinctive interpretations as the theories differ among children (Gopnik & Wellman, 1994). According to theory theory, children's intuitive theories undergo a conceptual change when they experience new evidence that is inconsistent with their current ToM. Children develop a more complex mental state concept by testing, modifying and reorganizing present knowledge about the causal structure of the world (Gopnik & Wellman, 1994, 2012). In the following part I present two approaches of this account.

1.2.1.1 Conceptual change of representations

Perner (1991) assumes that the development of children's ToM origins in the change and extension of their representational understanding. He defines three levels of representation that children pass in their first years of life: primary representation, secondary representation, and meta-representation. During the first year of life, children are tied to primary representations. They

are only able to conceive a currently real situation by a single updating model of the world. In the second year of life, secondary representations allow children to represent different situations (past vs. future, real vs. pretend) by constructing multiple, complex models. This enables children to break loose from present reality by representing past or hypothetical situations. Perner (1991) named children at this stage *situation theorists* as they now "understand representations as a special kind of *represented situation*" (p. 71). In the represented situation, children differentiate between the projection and reality without having a concept of representation. At around 4 years, children achieve the level of meta-representation, i.e. they are able to understand that representations represent representations. As *representation theorists* children are able to meta-represent a model about a model. They acquire a proper understanding of representation that helps them to form a ToM as they now understand the representational functions of mental states. This conceptual change enables them to understand that the mind can also misrepresent reality and only with that knowledge they are able to pass false-belief tasks.

1.2.1.2 Three steps of a belief-desire psychology

Bartsch and Wellman (1995) assume that people act in a way they believe will lead to what they desire. Thus, by considering both desires and (potentially false) beliefs, we can understand other people's behavior and their mind. From preschooler's everyday conversations about the mind containing words as *want* and *think/guess* to talk about desires and beliefs, Bartsch and Wellman (1995) construed how children's understanding of mind progresses. They describe three phases of development: early desire psychology, intermediate desire-belief psychology, and belief-desire psychology. In the second year of life, children possess a desire psychology. In this phase, they have no understanding of belief and they reason about own and other's people actions and feelings only in terms of desires. These desires are seen as related to real objects, actions and state of affairs in the world and therefore, early desire understanding is considered to be nonrepresentational. Children in this phase are not able to consider more than this one set of real contents. They have no conception of representational mental states such as represented

objects of desire but can attribute desires to objects that are actually not there. They can predict an action but only on the basis of what they believe about the world. At 3 years of age, children have a desire-belief psychology. They recognize that beliefs exist and attribute them to others but when reasoning and explaining behavior they primarily refer to the actor's desires without making recourse to beliefs. This expanded working model is only considered when desire psychology is insufficient to a consistent explanation. In belief-desire psychology, people's beliefs are central for understanding their mind and actions. As mentioned earlier, children now understand that "people engage in actions that they believe will achieve their desires" (p. 149). According to Bartsch and Wellman (1995), preschoolers' mental reasoning consists of the following constructs: Basic emotions and physiological states such as love and hunger that fuel one's desire and perceptional experiences that lead to one's belief and knowledge. Both, desires and beliefs lead to a corresponding action that leads again to predictable reactions depending on whether the action satisfies the desire or matches the belief.

Whereas both, Perner (1991) and Bartsch and Wellman (1995), claim a conceptual change in children's ToM development, "that is, developmental changes in performance on false-belief tasks reflect genuine changes in children's conceptions of persons" (Wellman et al., 2001, p. 671), they provide different explanations how children achieve a ToM. Perner (1991) assumes a cognitive change in children's representational abilities as basis for ToM. He argues that young children have a nonrepresentational understanding not only of the mind but also of typical physical representations as for example pictures or drawings. Children are only able to understand other's mental states as they acquire an understanding of representations in general. As they achieve a concept of representation at the age of 4 years, they are able to understand the representational function of mental states that leads to a ToM. Differently, Bartsch and Wellman (1995) view "the change to understanding representational states of mind […] as a development within children's theory of mind" (p. 194). Hence, they assume that already at the age of 3 years, children represent mental states in terms of beliefs. They postulate desire-belief psychology as intermediate stadium between nonrepresentational desire

psychology and later belief-desire psychology that eventually enables children to understand other's minds. As Perner (1991) views representation in a more complex way, he ascribes this ability only to children at the age of 4 years. In contrast, Bartsch and Wellman (1995) propose not to determine on this specific age but rather on developmental sequences (see Bartsch & Wellman, 1995, p. 194ff. for a comparison of the approaches).

1.2.2 Simulation Theory

Different from theory theory, the simulation theory denies that a theory-like system drives our understanding of the mind. Harris (1992) assumes that children acquire the ability to attribute mental states to self and others "by means of a simulation process" (p. 120). Children predict and anticipate other's behavior in a two-step process. First, they imagine having a particular desire or belief including the corresponding thoughts. Then, they transfer their own desires and beliefs to their counterpart by attributing the simulated desire or belief (Harris, 1991). Children's ToM abilities increment by adjusting their working model that leads to a more accurate simulation of another's person mental states. Hence, the difficulty of simulation depends on the degree to which the child's own working model has to be adjusted (Harris, 1992). To eventually being able to represent another person's false-belief, two default settings have to be adjusted: children have to ignore not only their own mental state but also the actual state of reality and instead have to feed relevant input to reach an accurate simulation.

In regard to ToM development, both theories differ crucially. Whereas the simulation theory predicts that children have privileged access to their own mental states while still having difficulties in understanding the mental states of others, the theory theory presumes that the understanding of self and others' mental states develop simultaneously when children reach the level of representation (Sodian & Thoermer, 2006). Empirical evidence supports the assumptions of theory theory, as children conceptualize own and others' mental states at about the same time (e.g., Gopnik & Meltzoff, 1994). Particularly in false-belief tasks, children were not able to reason about others'

mental states before accurately remembering their own false belief in this situation (Gopnik & Astington, 1988).

1.2.3 Modularity Theory

Modularity theory postulates that our understanding of the mind originates from an innate domain-specific cognitive module. Leslie and colleagues (Leslie, 1994; Leslie, Friedman, & German, 2004) claim a specialized ToM module that matures in children's second year of life. In order to pass false-belief tasks, children additionally need an inhibitory controlled selection process that develops in preschool age. Therefore, only when children were able to inhibit their own true-belief about a situation, they can pass false-belief tasks.

Modularity theory was developed and is mainly applied in research on autism as the majority of children with autism typically fail to pass false-belief tasks (Baron-Cohen, Leslie, & Frith, 1985). Many following studies supported the assumption "that autism involves a damaged theory-of-mind-module" (Leslie, 1992, p. 21). In this work, I will not further expand on modularity theory as for typically developing children, there is a lot of empirical evidence supporting theory theory (Mahy, Moses, & Pfeifer, 2014). I pursue Wellman's assumption of continuity in social-cognitive development (e.g., Wellman et al., 2008) and present empirical evidence for a relation between putative precursor abilities and later ToM.

1.3 Empirical evidence for continuity in social-cognitive development

Many early social-cognitive abilities as described in section 1.1 have been supposed to be related to an explicit ToM. For example, Woodward (1998) holds that the early understanding of goal-directed action is a first step "toward developing an understanding of the relationship between intentional agents and the objects they act on" (p. 31) without reasoning of the actor's intentions as mental states. Tomasello (1999) postulates that joint attention

skills were based on the understanding of self and others as intentional agents. This understanding also leads to an understanding of the voluntary attention and actions of agents when pursuing a goal. If the supposed relation exists in fact, then there have to be longitudinal relations from the single early abilities to later ToM as well as interrelations between these early social-cognitive abilities. By now, a variety of studies reported evidence that supports this relation to a ToM for early action-understanding (Aschersleben, Hofer, & Jovanovic, 2008; Wellman, Lopez-Duran, LaBounty, & Hamilton, 2008; Wellman, Phillips, Dunphy-Lelii, & Lalonde, 2004; Yamaguchi, Kuhlmeier, Wynn, & van Marle, 2009), joint attention (Charman, Baron-Cohen, Swettenham, Baird, Cox, & Drew, 2000; Sodian & Kristen-Antonow, 2015), intention-based imitation (Colonnesi, Rieffe, Koops, & Perucchini, 2008; Olineck & Poulin-Dubois, 2005), and pretend play (Youngblade & Dunn, 1995).

A very early precursor to an explicit ToM seems to be the attention to goal-directed action. Aschersleben and colleagues (Aschersleben et al., 2008) found that 6-month-old infants' decrement of attention in a visual habituation task using a modified Woodward-paradigm (Woodward, 1999) was positively related to their false belief understanding at 4 years of age. This link between early action understanding and later ToM persists during infancy and was also found in infants aged 10 to 14 months (Wellman et al., 2004, 2008; Yamaguchi et al., 2009).

Longitudinal evidence for a relation between joint attention abilities and later ToM development was found by Charman and colleagues (Charman et al., 2000). They showed that 20-months-olds' gaze switching behavior between an adult and an active mechanical toy as well as their looking to an adult in ambiguous goal detection tasks (*teasing* and *blocking*) were positively related to children's ToM abilities at 44 months. In these goal detection tasks, children were expected to look at the experimenter for disambiguation when he teases the child by offering and then withholding a toy (*teasing*) or when he blocks the child's view of a toy that she is manipulating (*blocking*). Recently, Sodian and Kristen-Antonow (2015) reported a positive correlation between infants'

declarative pointing to an object out of the experimenter's sight at 12 months and their comprehension of false belief at 50 months.

As to early intention understanding, a study by Colonnesi and colleagues (Colonnesi et al., 2008) revealed a relation between infants' intention understanding at 12 and 15 months and their later understanding of another's visual perspective as well as another's intention at 39 months of age. In addition, Olineck and Poulin-Dubois (2005) reported a longitudinal relation between intention understanding at 14 and 18 months and internal state language at 32 months.

Another potential precursor ability to a later ToM might be pretend play, however, findings are controversial. Whereas Youngblade and Dunn (1995) found a longitudinal relation between children's role enactment in social pretend play with mother or sibling at 33 months and their false belief understanding at 40 months, subsequent studies did not find a relation between pretend play and ToM (Charman et al., 2000; Nielsen & Dissanayake, 2004). It is noteworthy that these studies differed in the type of pretend play assessed. Perhaps only joint pretend play in a spontaneous play situation might be a precursor to ToM in contrast to solitary or prescribed pretend play, as only in joint pretend play children have to decode their partner's nonliteral actions (Harris & Kavanaugh, 1993).

Beside these findings reporting a relation of early social-cognitive abilities to later ToM and thus supporting the assumption of continuity in social-cognitive development (e.g., Wellman et al., 2008), there are only few studies investigating the interrelations between these early social-cognitive abilities. As all these early abilities are part of social cognition and all finally result in ToM, there should also be a relation between these social-cognitive abilities. By investigating how the single abilities were interrelated, the supposed continuity can be supported further on. So far, only very few interrelations were reported, which will be summarized in the following.

Olineck and Poulin-Dubois (2009) revealed a relation between infants' understanding of intentional actions at 10 months and their performance in

intention-based imitation tasks at 14 months. Other studies showed that children's declarative pointing relates to their understanding of other's intentions in intention-based imitation tasks at 15 months (Camaioni, Perucchini, Bellagamba, & Colonnesi, 2004; Colonnesi et al., 2008; Kristen, Sodian, Thoermer, & Perst, 2011). Also infants' performance on the blocking and teasing tasks was found to relate to their understanding of other's intentions (Charman et al., 2000). Recently, Sodian and Kristen-Antonow (2015) revealed a relation between children's declarative pointing at 12 months and their mirror self-recognition at 18 and 24 months of age. Finally, mirror self-recognition and children's pretend play correlate at the ages 15, 18 and 21 months (Lewis & Ramsay, 2004), and both, the self-recognition and pretend play emerge at around the same age (Baudonnière et al., 2002; Nielsen & Dissanayake, 2004). In sum, there are some hints for interrelations between some of these early social-cognitive abilities, which are supposed to be precursors of later ToM. It seems that especially those abilities relate that involve the understanding of another person's goal as well as those that require rather secondary representation, the ability to represent two different representations of the same object or situation (Perner, 1991).

The first study in this dissertation will be examining some of these relations more thoroughly. For this purpose, we investigated longitudinally as well as cross-sectionally the following early social-cognitive abilities within one study. At 12 months of age, we tested joint attention abilities including declarative pointing and the goal detection tasks teasing and blocking. At 18 months, we assessed children's pretend play behavior and their ability to recognize themselves in a mirror, to imitate an intended action, and to reason about other people's desires.

In the next section, I will give an overview of the various factors that are considered to influence social-cognitive development as some of them are also investigated in the studies presented in chapter 2. Although ToM development universally follows the same developmental trajectory (Wellman et al., 2001) some factors might foster social-cognitive development.

1.4 Influence on social-cognitive development

Various factors have been supposed to influence social-cognitive development such as attachment security (e.g., Meins, Fernyhough, Russell, & Clark-Carter, 1998), mother's interaction style (e.g., Symons & Clark, 2000), mother's use of mental state language (e.g., Meins et al., 2002), cultural practices (e.g., Liu, Wellman, Tardif, & Sabbagh, 2008) and education (e.g., Lecce & Hughes, 2010), executive control (e.g., Perner & Lang, 1999), language development (e.g., Milligan, Astington, & Dack, 2007), family background (e.g., Cutting & Dunn, 1999), and recently also child temperament (Wellman, Lane, LaBounty, & Olson, 2011). Below I will give a short overview on the influence of language, siblings, and parental education on ToM. As it is a poorly investigated research question, yet, I will elaborate explicitly on temperament in the next chapter.

1.4.1 Language development

The relation between children's language abilities and their ToM understanding is well investigated. A meta-analysis by Milligan and colleagues (Milligan et al., 2007), including 104 studies with English-speaking children below age 7, showed that both language ability in general as well as specific components of language ability (syntax, semantics, receptive vocabulary, memory for complements) were related to false belief understanding, with the lowest effect size for receptive vocabulary measures. Also for German speaking children, language skills are a significant predictor of ToM development (Lockl, Schwarz, & Schneider, 2004).

On the one hand, language might have an indirect effect on ToM as poor language abilities may limit children's task performance (DeVilliers & DeVilliers, 2000). Children have to possess sufficient language skills to comprehend the usually complex task demands. Otherwise, children's limited language abilities might strongly constrain their ToM performance. On the other hand, language might have a true effect on ToM abilities. The acquirement of different language skills might be leading to enhanced false belief understanding.

There exist various hypotheses on the contribution of different language aspects being crucial for ToM development (Sodian & Thoermer, 2006; Tomasello & Rakoczy, 2003). Perhaps, parents' use of linguistic symbols to indicate mental states such as think, know, and belief is important to focus children's attention to a mentalistic explanation of behavior (Bartsch & Wellman, 1995). The acquisition of these mental verbs seems to be important for the acquisition of the ability to infer mental states to self and others as these verbs refer to something that cannot be observed in others. Another possible explanation points out that the syntax in adults' talk about mental states makes their children sensitive for the complement structure of utterances about mental states (DeVilliers & DeVilliers, 2000). Typically, utterances about desires and beliefs are embedded in complement sentences, as e.g. "Maxi thinks that his chocolate is in cupboard A". The point is that the embedded part might be false while the main part is correct (DeVilliers & DeVilliers, 2000). That is, complementation allows representing counterfactual beliefs, for example what someone beliefs and what is actually the truth, and hence understanding of false beliefs. This assumption is supported by training studies (e.g., Lohmann & Tomasello, 2003) but there are also doubts as for example, German children understand desires earlier than belief although the linguistic complement structure is the same in both (Perner, Sprung, Zauner, & Haider, 2003). Also in English as well as Korean children sentential complementation was not per se related to false belief understanding (Farrar, Lee, Cho, Tamargo, & Seung, 2013). In general, it could also be the kind of discourse per se in which children recognize that people differ in what they know and think (Harris, 1996). Tomasello and Rakoczy (2003) conclude that different aspects of linguistic interaction are crucial for a better ToM understanding. When discourses include disagreements, misunderstandings and clarifications, children have the opportunity to learn that their own perspective on a situation may differ from another's understanding of the same situation. Even more important might be the opportunity to reflect their own perspective on a situation by parents evaluating the expressed thoughts of their children. In this way, normative perspectives are conveyed that lead to a differentiated understanding of individual beliefs and cultural norms and children have internalized "adult regulating speech" when they begin to pass false-belief

tasks. Another aspect seems to be the ability to understand and construct sentential complements, but as mentioned above there is little consent in literature.

To follow up the numerous studies that provided evidence for a relation between language and children's ToM abilities, we investigated the influence of children's language skills on early social-cognitive development in the studies reported in chapter 2.

1.4.2 Family background: Siblings and parental education

Among family background characteristics, often siblings and parental education are considered (see Pears & Moses, 2003, for a review). Dunn and colleagues (Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991) were the first to assume that various family characteristics may foster ToM development. Interactions with siblings seem to be relevant in that siblings share closer interests and feelings with the child as parents do and additionally, the child could profit by observing interactions between parents and sibling. As described above, family conversations and language abilities per se are another source for better ToM understanding. It is widely accepted that verbal ability depends on parental education (e.g., Cutting & Dunn, 1999; Dollaghan et al., 1999). Also, as particularly maternal education seems to be related to their children's intelligence, it might indirectly affect children's ToM abilities (Pears & Moses, 2003). Therefore, also maternal and paternal education level are relevant as influencing factors. These explanations show how social interactions contribute to and shape ToM understanding (Dunn et al., 1991) but study results are inconsistent for both.

Findings on the influence of siblings are controversial. A number of studies reported evidence for a positive influence of siblings on children's ToM development (Jenkins & Astington, 1996; McAlister & Peterson, 2006, 2007, 2013; Perner, Ruffman, & Leekam, 1994). In some studies, this positive influence on ToM development was found only for older siblings but not for younger ones (Farhadian, Gazanizad, & Shakerian, 2011; Lewis, Freeman, Kyriakidou, Maridaki Kassotaki, & Berridge, 1996; Ruffman, Perner, Naito,

Parkin, & Clements, 1998), or only for siblings that were between 12 months and 13 years of age that is, with whom the child may engage in sibling-based play (Cassidy, Fineberg, Brown, & Perkins, 2005; Peterson, 2000). Finally, several studies did not find any sibling effect at all (Carlson & Moses, 2001; Cole & Mitchell, 2000; Cutting & Dunn, 1999; Henning et al., 2011; Peterson & Slaughter, 2003).

Also controversial are the findings on the influence of parental education. Whereas some studies reported a strong relation between maternal education level and children's ToM development (Cutting & Dunn, 1999; Henning et al., 2011; Pears & Moses, 2003), others did not find such an effect (e.g., Ruffman, Perner, & Parkin, 1999). The same controversy is true for an influence of paternal education level on ToM (Cutting & Dunn, 1999; Dunn & Brown, 1994). Perhaps, comparison between studies is difficult as, for example, studies used different tasks to assess children's ToM abilities and also, they did not all control the influence of language (see Pears & Moses, 2003).

In study 2, one aim was to assess the influence of language, siblings, and parental education on early ToM development in children at the age of 3 years. Moreover, the focus of study 2 was on the influence of temperament on ToM development. Temperament has only come into focus recently. Therefore, I will present this factor more detailed in an extra chapter and take into account not only its influence on social-cognitive development but also on children's study performance.

2 Temperament

As it is known, children's social experiences and interactions are influenced by their temperament (Rothbart & Bates, 1998). Take, for example, stranger anxiety. Depending on the fear's intensity children will either avoid or approach to interact with a stranger. Similarly, the experiences in interactions differ for introvert vs. extrovert children as well as for shy vs. communicative children. As described in the preceding section, social interactions contribute to

and shape ToM understanding (Dunn et al., 1991), also outside the family. Therefore, it seems reasonable to assume that ToM development is also influenced by children's temperament (Wellman et al., 2011). In the next sections, I will first introduce temperament *per se* and then describe the few studies investigating the assumed relation between temperament and social-cognitive development, so far.

2.1 Definition, measurement and stability

Temperament is commonly defined as personality traits that emerge early in ontogeny, show stability throughout childhood into adulthood, and have a substantial genetic component (Buss & Plomin, 1984; see Henderson & Wachs, 2007, for a review). While there is a general consensus in the literature about this definition, there are different assumptions about the components of temperament. I will give an overview about three prominent approaches.

Thomas and Chess (1977) suppose a total of nine temperament categories based on empirical findings: rhythmicity of biological functions, activity level, approach to or withdrawal from new stimuli, adaptability, sensory threshold, predominant quality of mood, intensity of mood expression, distractibility, and attention span. Above, as these dimensions were developed for clinical purposes, they identified three temperamental patterns, namely easy, difficult and slow-to-warm-up temperament, to estimate the ontogenesis of behavior disorders. Buss and Plomin (1984) specify three independent temperamental dimensions of inherited personality traits: emotionality, activity and sociability. Emotionality refers to autonomous emotional arousal and is equivalent to the tendency to show distress. Activity refers to behavioral arousal and can well be measured by the rate and amplitude of body movements. Sociability refers to the preference for companionship instead of solitude and involves the number of social activities with others. A third approach offers theoretically derived temperament dimensions (Rothbart et al., 2001). According to Mary Rothbart (1986), temperament is defined "as constitutionally based individual differences in reactivity and self-regulation" (p. 356; Rothbart & Derryberry, 1981). Reactivity refers to differences in

infant's emotional, motor and attentional reactions in terms of, for example, threshold and intensity, whereas self-regulation refers to behavioral processes such as attention or approach and withdrawal that regulate this arousability of responses.

Whereas Thomas and Chess (1977) predominantly address behavioral style and Buss and Plomin (1984) are restricted to inherited traits, Rothbart (1986) is not that limited in focusing on reactivity and self-regulation. In our studies, we rely on Rothbart's approach. She and her colleagues provide a battery of measures that enable one to assess temperament not only in different situations and contexts but also at different ages (see Goldsmith et al., 1987, for a review of the presented approaches). Also, their instruments are based on empirically findings and three broad temperament dimensions are found across ages.

To measure individual differences in temperament from early infancy to childhood, Rothbart and colleagues developed a series of parent-report instruments, including the Infant Behavior Questionnaire (IBQ, Rothbart, 1981) for infants between 3 and 12 months of age. The original version of the IBQ assesses six domains of infant temperament, namely Activity Level, Soothability, Fear, Distress to Limitations, Smiling and Laughter, and Duration of Orienting (for a validation of the German version of the IBQ see Pauli-Pott, Mertesacker, & Beckmann, 2003). A decade ago, Gartstein and Rothbart (2003) revised the questionnaire (IBQ-R) to assess temperament in a more differentiated way by including an additional eight subscales: Approach, Vocal Reactivity, High Pleasure, Perceptual Sensitivity, Sadness, Falling Reactivity, Low Pleasure, and Cuddliness (for a validation of the German version of the IBQ-R see Vonderlin, Ropeter, & Pauen, 2012). Analyses of the underlying structure through factor analysis suggest that infant temperament is structured in three broad dimensions: Surgency/Extraversion, Negative Affectivity and Orienting/Regulation (Gartstein & Rothbart, 2003). For the temperament questionnaire developed for toddlers (ECBQ, Putnam, Gartstein, & Rothbart, 2006), the third dimension emerging was Effortful Control, which includes additional subscales to those related to the dimension Orienting/Regulation, to

account for developmental changes in self-regulation in the second year of life. These three dimensions are also to be found in the Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001) for children between 3 and 7 years of age.

Research employing these instruments showed both developmental changes in infant temperament as well as relative stability of the underlying factorial structure. In a longitudinal study, Rothbart, Derryberry, and Hershey (2000) showed that already newborns differed in temperamental aspects such as distress proneness, activity level, and visual orienting. Further individual differences emerged for frustration and positive affect in the first months of life, for fear by 6 months of age, and for self-regulatory control in late infancy. Other changes in temperament subscales seem to be related to changes in infants' emotional and self-regulatory development. For example, Activity Level as well as Visual Orienting lack stability in the first few months, potentially because activity level is initially related to negative affect and later in development to positive affect (Rothbart & Bates, 1998). A major development of the orienting system in the infant brain may explain early instability in visual orienting (Rothbart et al., 2000). In addition, in early infancy, attention seems to be more reactive than actively self-regulatory (Zentner & Bates, 2008), with self-directed attention focusing emerging only at the end of the first year of life. However, despite this developmental change, several studies provide evidence for stability of the three-factor temperament dimensions from infancy to toddlerhood (Casalin, Luyten, Vliegen & Meurs, 2012; Komsi, Räikkönen, Heinonen et al., 2008; Komsi, Räikkönen, Pesonen et al., 2006; Putnam, Rothbart & Gartstein, 2008) and also single components of positive emotionality like Smiling and Laughter tend to be quite stable (Zentner & Bates, 2008).

2.2 Influence of temperament on social-cognitive development

Compared to the vast amount of research on influencing factors on ToM development, relatively few studies have assessed the impact of temperament on social-cognitive development (Banerjee & Henderson, 2001;

Carlson & Moses, 2001; Lane et al., 2013; Walker, 2005; Wellman et al., 2011). Carlson and Moses (2001) found a positive relation between preschoolers' inhibitory control and their ToM performance. Inhibitory control was rated by their parents in the same-titled subscale of the CBQ (Rothbart et al., 2001) and measured in a task that requested the children to respond counter to a prepotent tendency. ToM was assessed in a battery containing false-belief tasks as well as a deception task and appearance-reality task, where children had to distinguish the appearance of an object from reality. Banerjee and Henderson (2001) found a negative relation between school children's self-reported social anxiety and their performance in various social-cognition tasks, particularly when the social anxiety was paired with shy-negative affect. In a study with 5-year-olds, Walker (2005) assessed children's temperament via teacher's ratings of their peer-related social skills. She obtained positive relations between false belief understanding and aggressiveness as well as negative relations between false belief understanding and shyness, but only in boys. Whereas these earlier studies do suggest some influence of child temperament, they provide little consent with regard to what specific temperamental characteristics might enhance ToM development.

Recent research suggests a developmental link between childhood temperament and ToM abilities that is specific to a less reactive more observant temperament (Lane et al., 2013; Wellman et al., 2011). Wellman and colleagues (Lane et al., 2013; Wellman et al., 2011) refer in their work to the emotional reactivity hypothesis (Hare, 2007) to explain how child temperament might influence ToM development. This hypothesis originally accounts for social-cognitive capacities in dogs. It holds that dogs that were selected for domestication due to their nonaggressive and non-fearful temperament regarding humans, developed human-like social-communicative skills during domestication in convergent evolution with humans, i.e. level of emotional reactivity has modulated social-cognitive performance. In cooperative-communicative situations, dogs show capacities similar to early social-cognitive capacities of children, whereas wild canines and even chimpanzees perform poorly in such situations (Hare & Tomasello, 2005; Povinelli & Eddy, 1996). Further evidence supporting the emotional reactivity hypothesis derives

from findings showing that even bonobos and chimpanzees, the closest relatives to humans, differ in their social-cognitive abilities (Hare, Melis, Woods, Hastings, & Wrangham, 2007; Herrmann, Hare, Call, & Tomasello, 2010; Okamoto-Barth, Call, & Tomasello, 2007). Like dogs, bonobos outperform chimpanzees in social-cognitive tasks. These two species also differ from each other in temperament with bonobos being less aggressive and shyer than chimpanzees (Hare et al. 2007; Herrmann et al., 2010).

Taking up this proposal by Hare and Tomasello (2005) that the initial difference in phylogeny might have regarded temperament, Wellman and colleagues (2011) assumed that also in child development an initial difference in temperament may lead to differences in interactive behavior and social experiences, which in turn may foster or interfere with the development of mental understanding. Supporting evidence is provided by their longitudinal study showing that 3-year-old children, who were rated by their parents as shy, nonaggressive and perceptually sensitive, showed more sophisticated ToM abilities in a battery of false-belief tasks 2 years later. Temperament was assessed via the CBQ (Rothbart et al. 2001) and additionally, the Child Behavior Checklist/Ages 2-3 (CBCL/2-3; Achenbach, 1992) was used to assess aggressive and withdrawn behavior. The authors argued that this "less reactive more observant temperament" (p. 321) facilitates social participation and social information processing. Similarly, Lane and colleagues (2013) found relations between less aggressive temperament and false belief understanding both in Chinese and US American preschoolers. Again, parents rated their children's temperament with the CBCL/2-3 (Achenbach, 1992) and an abbreviated version of the CBQ (Rothbart, 1989). False belief understanding was tested with an unexpected-content task and a change-of-location task (see chapter 1 and 1.1 for examples of the tasks). In addition, Lane and colleagues (2013) measured children's hypothalamic-pituitary-adrenocortical HPA-axis reactivity via salivary cortisol and reported that children with moderately high reactivity, which is related to social engagement and attentiveness (Blair, Peters, & Granger, 2004), exhibited more advanced ToM understanding. By including this physiological measure, they could clarify the relation between social

withdrawal and social-cognitive development, as only shy but not socially avoidant behavior fosters ToM development.

2.3 Influence of temperament on study performance

Beside the substantial influence of temperament on children's social-cognitive development, it can be expected that temperament is also relevant in a methodological way, as it may influence study results. It is important to note here that temperament-based dispositions and tendencies are not continually expressed but activated by specific situations. Test situations in a laboratory, especially habituation tasks in infancy, contain contextual factors that may activate these dispositions. Habituation in young infants has been widely used to examine the early perceptual and cognitive development (e.g., Colombo & Mitchell, 2009). The habituation-dishabituation technique is based on the circumstances that infants' looking times decrease when they are repeatedly presented with the same stimulus (habituation), and that their looking times increase again when they are presented with a new stimulus they recognize as different from the previous one (dishabituation). When assessing infant cognition, infants are commonly habituated to one stimulus and then in the subsequent test phase presented with two new stimuli. These are both similar to the habituation stimulus but differ from each other in that one is consistent with the knowledge the infant is assumed to have, whereas the other stimulus is thought to violate the infant's expectations and should thus result in an increase in looking time (see Slater, 1995, for an overview about habituation techniques). Not only has the infant to deal with an unfamiliar environment and interact with strangers, but also the habituation task as such demands the infant to remain calm and focused on the presented stimuli. It might thus very well be that those infants whose temperamental dispositions help them to easily adapt to these general requirements, show an overall better test performance than infants with a so-called difficult temperament (Thomas & Chess, 1977).

This assumption raises the question whether task performance differs not only as a function of differences in the cognitive abilities under investigation, but also as a function of differences in temperament (Vonderlin,

Pahnke, & Pauen, 2008). Despite its wide use in infancy research, concerns have arisen about factors that might affect the validity of the habituation technique, such as systematic dropout (e.g., Slaughter & Suddendorf, 2007), that in turn might be influenced by the infant's temperament. In fact, dropout rates in studies using visual habituation or violation-of-expectation paradigms are relatively high compared to infant studies employing other paradigms in which infants are required, e.g., to produce actions or to interact with the experimenter. In a review covering 101 published studies, Slaughter and Suddendorf (2007) reported dropout rates up to more than 60 % due to crying and fussiness. This raises the question whether dropout rates in habituation paradigms randomly vary or whether they are systematically related to specific infant characteristics. Slaughter and Suddendorf (2007) discuss infant temperament as one possible factor that might have an impact on the likelihood of dropout rates in visual habituation tasks. In this case, results obtained in habituation studies might not be generalizable to all infants at a similar age but instead reflect the abilities of a specific subgroup with distinct characteristics.

So far, only few studies have assessed the impact of infant temperament on number of infants completing or not completing a visual habituation task, and they provided inconsistent results. As opposed to completers, those infants whose heightened level of distress during testing leads to termination of the experiment ahead of time, are in the following termed non-completers. Whereas older studies suggest an impact of temperamental factors on task performance (Miceli, Whitman, Borkowski, Braungart-Rieker, & Mitchell, 1998; Treiber, 1984; Wachs & Smitherman, 1985), no differences in temperament ratings were found between completers and non-completers in newer studies (Slaughter & Suddendorf, 2007; Vonderlin et al., 2008). Miceli and colleagues (1998), for example, assessed the performance of 4-month-olds in a paired comparison task and found that non-completers were rated as more active and more prone to smiling and laughter than completers. Treiber (1984) showed that infants, who did not complete a habituation task when they were 4 months of age, were rated as more active, withdrawing and negative in mood at 11 months of age. Also, Wachs and Smitherman (1985) tested 11-, 18- and 28-week-olds in a habituation study and found that only female non-completers

received higher scores in fussy-difficult and unadaptable behavior. In a further study employing not a looking time task but an operant-conditioning task, female non-completers were rated as less attentive for extended time periods and as showing more distress to novelty than female completers (Fagen, Ohr, Singer & Fleckenstein, 1987). In contrast, in more recent work by Vonderlin and colleagues (Vonderlin et al., 2008), 7-month-olds' differences in temperament were related to strength of familiarization response, but not to dropout rate. Similarly, Slaughter and Suddendorf (2007) did not find systematic relationships between experimental outcome and dropout in the sample of studies included in their review.

Although especially the older studies point to some impact of infant temperament on dropout in looking time tasks, findings show little consent with regard to what specific temperamental characteristics might be crucial for completing these tasks. Furthermore, comparison between studies is difficult as studies substantially differ in the specific technique used, the ability or knowledge assessed in the experiment, infants' age at testing and the temperament questionnaire used. The aim of study 3 was therefore to extend this line of research by also taking into account content of task as well as long-term stability of individual differences in temperament and dropout.

3 Overview and goals of this dissertation

The studies of this dissertation were part of a broader longitudinal project assessing a potential continuity in social-cognitive development in the first four years of life. In this project, 164 children were tested at five different age points: 6 months, 12 months, 18 months, 3 years, and 4 years of age.[1] The main research question of the whole project was to analyze how mothers' interaction style influences the social-cognitive development of their children, especially early social-cognitive abilities that were supposed to be precursor abilities to a later Theory of mind. A second aim was to assess if there is continuity in social-cognitive development as this assumption is still subject of controversial discussion in developmental psychology literature (Aschersleben et al., 2008; Henning et al., 2011). Additionally, it is still an open question to what extend this development can be fostered or impaired, for example through maternal interaction style quality or child temperament, language and cognitive abilities.

The majority of the tasks that we conducted at the first four assessments (6 months to 3 years) within this project were part of the present dissertation with the exception of mother-child interaction and general cognitive abilities. Next, I will describe the goals of the three studies reported in this work:

First, I examined the question if there is continuity in social-cognitive development. So far, only few studies have considered the interrelation between early social-cognitive abilities and only two of these studies (Charman et al., 2000; Nielsen & Dissanayake, 2004) have investigated the relation of more than two of these early social-cognitive abilities. In study 1, we aimed to extend the previous findings on continuity by examining longitudinally within one study to what extend five of these social-cognitive abilities in the same group of children are associated. At 12 months, we surveyed infants' joint

[1] As some of the children did not take part in the project from the beginning but were included with 12 or 18 months and due to various dropout rates in the single tasks, number of children varies in the single studies.

attentional skills and at 18 months, we assessed children's pretend play behavior and their ability to recognize themselves in a mirror, to imitate an intended action, and to reason about other people's desires. Based on previous studies (Charman et al., 2000; Kristen et al., 2011; Sodian & Kristen-Antonow, 2015) we expected a relation between various joint attention abilities and children's mirror self-recognition as well as their understanding of other's intentions. Also based on previous work (Lewis & Ramsay, 2004), we predicted a relation between children's mirror self-recognition and their pretend play. In addition, we expected to reveal further interrelations by investigating the above-mentioned abilities within one longitudinal study.

Second, I investigated influences on social cognition, especially the influence of temperament on social-cognitive abilities. Recent research by Wellman and colleagues (Lane et al., 2013; Wellman et al., 2011) suggests a developmental link between childhood temperament and ToM abilities that is specific to a less reactive more observant temperament. Their idea based on the emotional reactivity hypothesis (Hare, 2007) that originally accounts for social-cognitive capacities in dogs. In their work (Lane et al., 2013; Wellman et al., 2011) they have transferred research results on primates and canines on humans to explain how temperament might influence ToM development in preschoolers. The aim of study 2 was to extend this new line of research. First, we wanted to assess the influence of temperament on ToM development in addition to the well-investigated factors language abilities, parental education and siblings. Second, we wanted to test the influence of temperament on ToM development in children at the age of three years, that is, before the emergence of false belief reasoning. Based on the work by Wellman and colleagues (Lane et al., 2013; Wellman et al., 2011) we focused on four temperament characteristics, namely Shyness, Fear, Perceptual Sensitivity and Anger/Frustration. We expected to find a positive relation between the first three temperament variables and children's ToM at 3 years of age and, in contrast, a negative relation between Anger/Frustration and ToM.

Third, I considered the influence of some temperamental characteristics on children's study performance. There exists evidence that temperament

might be hindering in test situations, especially in visual habituation tasks (Miceli, et al., 1998; Treiber, 1984; Wachs & Smitherman, 1985). The question has risen if a high dropout rate is influenced by a special temperament of those infants not completing the tasks. As we follow the same children over a variety of tasks in this longitudinal project, it is important to test whether dropout was systematically related to infant temperament. Hence, the aim of study 3 was to longitudinally examine the impact of infant temperament on the dropout rate in visual habituation experiments in infants at 6 and 12 months of age, as habituation tasks are a particularly restrictive kind of study tasks. Besides, we checked for stability in temperament in the first year of life. Based on previous studies (Carranza Carnicero, Péréz-Lopéz, Del Carmen González Salinas, & Martínez-Fuentes, 2000; Slaughter & Suddendorf, 2007), we expected a moderate number of dropouts in the current sample as well as a general stability of temperament across the two age points. Also, we predicted the temperamental domains Distress to Limitations and Duration of Orienting to have an impact on dropout as both tap relevant features of habituation experiments: The first one includes fussing and crying while frustrated or being motor constrained, the second one serves as an indicator for infant attention to a specific object or event.

Following this introductory chapter, chapter 2 presents the empirical research: study 1 reports on continuity in social-cognitive development, study 2 deals with the influence of temperament on social-cognitive development and study 3 explored the influence of temperament on study performance. Finally, chapter 3 summarizes the three studies and integrates the respective main findings into the literature reviewed in chapter 1. To conclude, limitations and implications for future research are discussed.

Chapter 2: Studies

Study 1 – Social-cognitive development from infancy to toddlerhood

The aim of the current study was to extend findings on continuity in social-cognitive development by examining within one study to what extend prominent early social-cognitive abilities are associated. To our knowledge, this is the first longitudinal study that assesses cross-sectional and longitudinal relations between five of these social-cognitive abilities in the same group of children.

To that aim, children's social-cognitive abilities were assessed when they were 12 and 18 months of age. At 12 months, we surveyed infants' joint attention skills (following another's gaze and point gestures, ambiguous goal detection tasks, production of imperative and declarative pointing, and understanding of behavioral request). At 18 months, we assessed children's pretend play behavior and their ability to recognize themselves in a mirror, to imitate an intended action, and to reason about other people's desires. Additionally, we controlled for children's receptive language abilities at 12 and 18 months. Based on previous studies (Camaioni et al, 2004; Kristen et al., 2011) we expected a relation between children's declarative pointing and their understanding of other's intentions as well as children's mirror self-recognition (Sodian & Kristen-Antonow, 2015). Furthermore, we expected a relation between children's performance in goal-detection and their understanding of other's intentions (Charman et al., 2000). Also based on previous work (Lewis & Ramsay, 2004), we predicted a relation between children's mirror self-recognition and their pretend play. Moreover, we expected to reveal further interrelations by investigating the above-mentioned abilities within one longitudinal study.

2 Method

2.1 Participants

A total of 139 children (61 females) from a medium-sized city in the southwest of Germany were included in the final sample and tested at 12 months (*M* = 365.81 days, *SD* = 11.48, *range* = 324 - 397 days) and 18 months of age (*M* = 552.66 days, *SD* = 12.71, *range* = 530 - 607 days). Nine of these children only took part in the study at the age of 18 months. Parental educational level was relatively high: 59% of mothers and 63.3% of fathers held a college or university degree, 20.1% of mothers and 16.5% of fathers completed secondary school at top track (Abitur), and 20.9% of mothers and 20.1% of fathers completed secondary school at lower or middle track.

Parents were recruited by telephone from a list of families who had earlier expressed interest in volunteering for research on child development. At each visit, they received a recompense for travel expenses and children were given a small gift and a certificate for participating.

2.2 Tasks and materials

Children were tested in a quiet laboratory room. Sessions lasted between 60 to 90 minutes. Beside the above-mentioned social-cognitive abilities, we assessed also general cognitive abilities and observed a mother-child interaction. At 12 months, joint attention abilities were tested always at the end of the session, and always by the same female experimenter. At 18 months, the social-cognitive tasks were administered by a total number of five trained experimenters in the following order: Pretend Play, Imitation, Divergent Desires and Mirror Self-recognition.

2.2.1 Social-cognitive tasks at 12 months: Joint Attention

The tasks assessing infants' joint attentional skills were based on the Early Social Communication Scales (Mundy et al., 2003) and the ambiguous goal detection tasks by Charman and colleagues (Charman et al., 2000). In these interactive tasks, the child and the experimenter were sitting at a table facing each other. The children sat on their mother's lap. Mothers were

instructed not to interfere in the session. The session was videotaped for later coding.

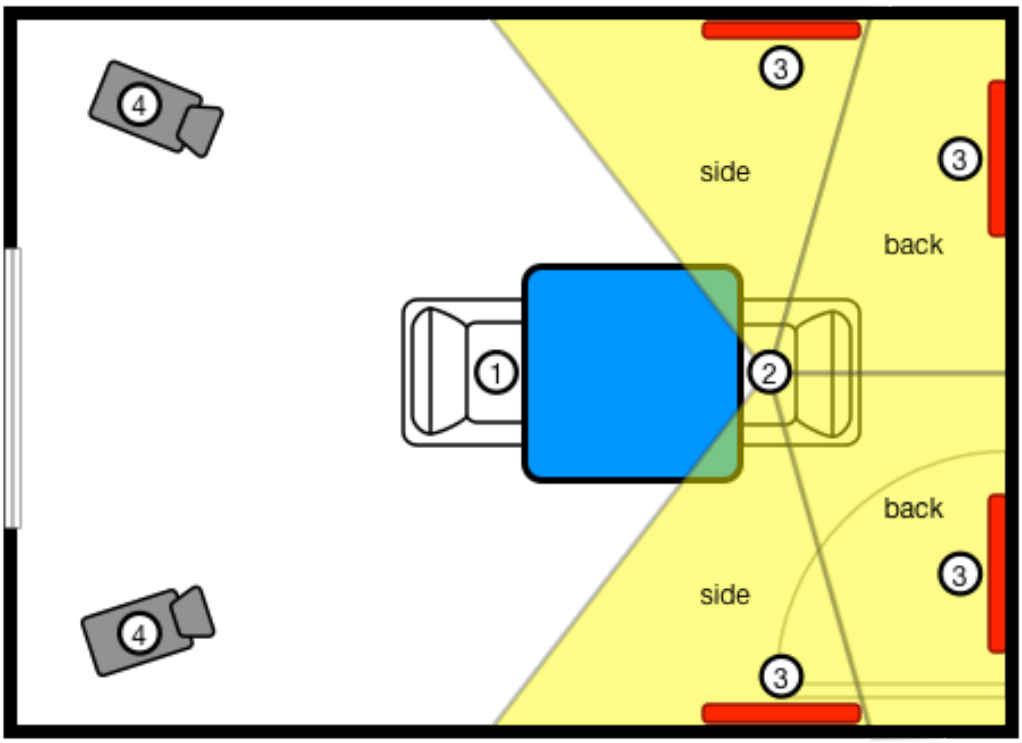


Figure 1: Setup for joint attention tasks.
1) Experimenter, 2) Child, 3) Target posters, 4) Cameras

Point following (PF) and Gaze following (GF) to distal objects. Four DIN A3 posters were fixed to the walls, two behind the child (ca. 40°angle, left and right of midline) and two on the child's side (90°angle, left and right, see Figure 1). Each infant received four trials, one per position, in one of two fixed orders: PF side (left), GF side (right), GF back (left), PF back (right) or GF side (left), PF side (right), PF back (left), GF back (right). After establishing eye-contact by calling the infant's name, the experimenter gasped excitedly, turned her head and torso and gazed (*Gaze following*) or gazed and pointed with an outstretched arm (*Point following*) at the target for 2 seconds, and looked back at the child (while keeping her arm outstretched in the PF trials) and shortly vocalized; she then gazed at the target and back to the child another two more times. The experimenter only vocalized when looking at the child. It was recorded for each trial whether the child scanned the right side without locating

the target (0), located the target (1), located the target and checked back with the experimenter (2), or did not follow the experimenter's attention (no AF). Aggregated attention following scores were calculated as a function of gesture (AF Gaze Score, AF Point Score) and of target position (AF Side Score, AF Back Score) according to whether children reached less than Level 1 in both trials (0), reached at least Level 1 in one trial (1), or reached at least Level 1 in both trials (2).

Object Spectacle (OS). To elicit joint attentional behavior, the experimenter activated one of three toys in front of the child and out of her reach: a jumping-jack puppet, a spinning top and a plastic Tigger that, when switched on, talked accompanied by a blinking nose. Toys were introduced in this order; following repetitions depended on the child's interest. The experimenter presented the toy for about 6 seconds, briefly activating it 3 times consecutively, and then paused with the toy out of the infant's reach while she continued gazing at the child. The occurrence of the following behaviors during the presentation and response period was coded: *Initiating Joint Attention* (gazing at the experimenter's face), *Showing* the experimenter a toy, *Declarative Pointing* (pointing at an object out of the child's reach with the index finger extended), and *Imperative Pointing* (pointing at an object out of the child's reach with all fingers extended). Up to five OS phases were conducted during the joint attention session.

Goal Detection Tasks. (1) *Teasing*. The experimenter offered the child a toy (mainly in response to an imperative point) and the moment the infant reached for it, the experimenter withdrew the toy for a period of 5 seconds. (2) *Blocking.* When the infant was manipulating a toy, the experimenter covered the infant's hands with her own for a period of 5 seconds, thereby blocking the infant's view of the toy. For both tasks it was recorded, whether the child looked up into the experimenter's eyes within the 5-second interval. This behavior was interpreted as an imperative gesture with instrumental function (see Charman et al., 1998).

Behavior Request (BR). Following an Object Spectacle phase, the experimenter handed the child the toy, let her manipulate it for about 10

seconds and then requested: “Give it to me!” The verbal request was accompanied by a palm-up hand gesture. Request bids were repeated up to 3 times. It was coded whether the infant handed the toy to the experimenter within 3 seconds after the request or showed clear signs of comprehension, e.g., saying “no”, shaking her head, or pulling back the toy. The occurrence of understanding this behavioral request in at least one of the two trials was included in the analyses.

The social-cognitive tasks were to be administered in a fixed order (OS 1, AF 1, OS 2, BR 1, AF 2, OS 3, Teasing, AF 3, OS 4, BR 2, AF 4, OS 5, and Blocking) that, however, was flexibly adjusted to the individual child’s state. A Behavior Request as well as Teasing and Blocking always immediately followed an Object Spectacle Phase.

Videos were coded offline for the occurrence of the joint attention behaviors. Ten of the 139 children were not tested in the joint attention tasks; further 2 children were excluded because of experimenter failure. For some children and tasks, data was missing due to fussiness (GF Back: 2, Teasing: 1, Blocking: 8, Behavior Request: 1) or experimenter failure (GF Back: 1, PF Back: 2, Teasing: 4, Blocking: 5, Behavior Request: 1). As OS 5 was missing for 42 children, only the occurrence of behaviors in the first four Object Spectacle Phases was included in the analyses. A second independent observer coded 30% of the videotapes of all children. Kappas for the occurrence of behaviors were: 1.0 (GF Side), .73 (PF Side), .77 (GF Back), .71 (PF Back), .78 (Initiating Joint Attention), .75 (Showing), .72 (Behavior Request), .72 (Declarative Pointing), .39 (Imperative Pointing, match n=27, no match n=5), .91 (Teasing), 1.0 (Blocking), all p-levels < .006.

2.2.2 Social tasks at 18 months

2.2.2.1 Pretend Play

Following Bornstein, Haynes, O'Reilly, and Painter (1996), the child’s pretend play behavior was observed in two consecutive 5-minute episodes in the following order: collaborative play with the mother and solitary play. A set of age appropriate toys were provided that included gender specific and gender

neutral categories and allowed for a variety of different play behaviors (e.g. puppets, teddy bears, cutlery, toy mobile phones, a photo camera, toy cars, or nesting barrels). Mothers were instructed to play with her child like they usually would at home. After 5 minutes had elapsed, mothers told the child to continue playing and then sat down at a table pretending to read a magazine. Mothers were instructed not to intervene in the child's solitary play.

Figure 2: Toys provided for pretend play.

The child's play behavior was coded offline from videotape. The interaction was divided into 10-second intervals and children's play behavior was categorized according to Bornstein and colleagues' (1996) eight level coding system. The highest level shown in each interval was coded. Levels 1-4 corresponded to non-symbolic play: *(1) Effect focused play* (production of an effect, e.g., pressing the keys of a phone), *(2) Inappropriate combination* of two different toys (e.g., putting a toy car in a pot), *(3) Appropriate combination* of toys (e.g., nesting barrels), or *(4) Transitional play* with the child showing signs of, but no confirmatory evidence for pretend play (e.g. putting the mobile phone to the ear without vocalization).

Levels 5-8 corresponded to evident pretend play: *(5) Self- directed*

pretence (e.g., drinking from a cup), *(6) Other directed pretence* (e.g., feeding a puppet with an empty spoon), *(7) Linking two or more pretence actions* in a sequence (e.g., stirring with a spoon in a pot and then eating from the empty spoon), or *(8) Substitution pretence*, with using one or more object substitutions in a play sequence (e.g., using a cup as a phone and talking into it). Two additional codes were used for intervals 1) without any toy play or 2) with toy play that did not fall into one of the 8 categories (e.g., manipulating a toy without producing an effect). For each play condition, the data of one child was missing due to refusal to play. The interactions of 25% of the children (n = 34, both play episodes per child) were coded by a second coder. Across all 8+2 levels, Kappas were .84 and .81 for collaborative and solitary play, respectively, p-levels < .001. As in Charman and colleagues (2000), for analyses, occurrence of symbolic play (levels 5-8) was used as dichotomous variable.

2.2.2.2 Imitation

Two of Meltzoff's (1995) six imitation tasks were administered: (a) pulling apart a dumbbell (*dumbbell task*), and (b) putting a loop on a prong (*prong and loop task*). The child on her mother's lap and the experimenter were sitting at a table at a 90° angle. In the baseline, the experimenter placed the object(s) on the table in front of the child and encouraged exploration ("Look, what is this?"). Children's spontaneous production of the target action was coded in the 20 seconds following the child first touching the object(s). In the demonstration phase, the experimenter tried but failed to perform the target action three times in a row without giving any nonverbal hints about her failure: (a) the experimenter's hand (right, left, and right hand) slipped of one end of the dumbbell, and (b) the loop was released slightly off the prong (right, left, above). In the test phase, the experimenter placed the material in front of the child with the words "Now it's your turn". Children's imitation of the intended action was coded in the 20 seconds following the child first touching the object(s).

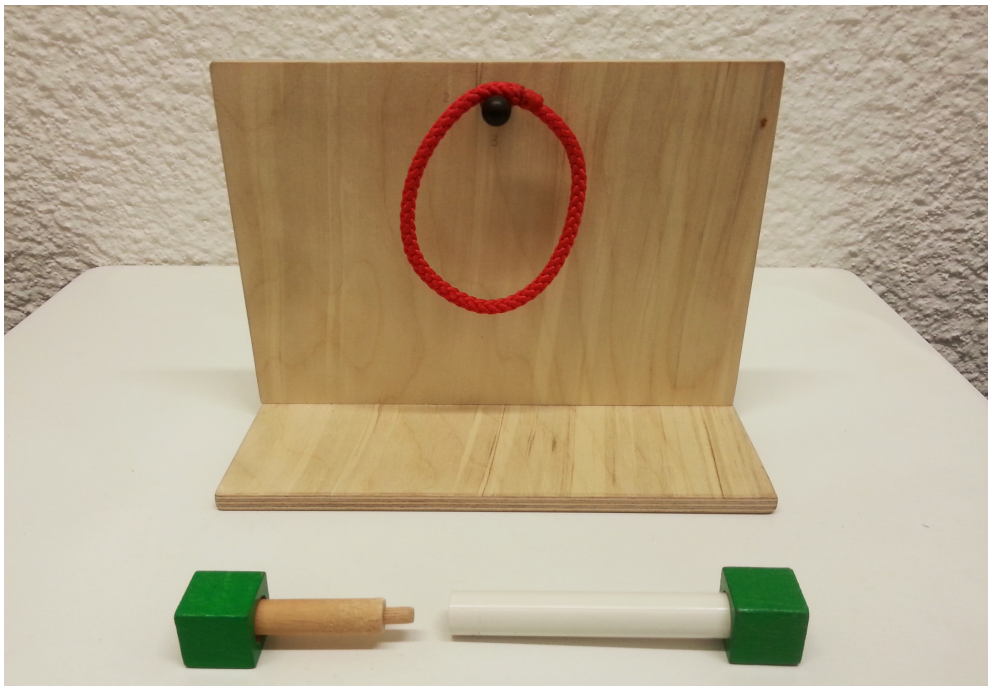


Figure 3: Material for imitation task.

Children's manipulation of the objects was coded from videotape. The child's action was scored 1 or 0 depending on whether she did or did not show the target action within the 20-second response interval in both baseline and test. Data of both tasks were excluded for 3 children because of fussiness (2) or experimenter failure (1). A second independent observer coded the performance of 27 % randomly selected children. For both tasks and both baseline and test, Kappa for the occurrence of the target action within the response interval was 1.

2.2.2.3 Divergent Desires

To test infants' understanding of divergent desires we conducted the mismatch group condition of Repacholi and Gopnik's classic study (1997). Two identical, opaque plastic bowls (diameter of 12 cm, height 6,5 cm) were placed on a wooden meal tray (44 x 32 cm), one filled with raw broccoli, the other filled with cookies. The bowls were placed in the left (broccoli) and right (cookies) corner on the infant's side of the tray. The child on her mother's lap and the experimenter sat at a table on opposite sides. In a 45-second baseline, the experimenter placed the tray in front of the child and asked her to taste the

food to reveal the child's preference. The experimenter then moved the tray out of the child's reach, picked up one bowl after the other and tasted the food in the following order: she first tested the child's preferred food and facially and vocally expressed disgust "Eeew, (name of the food), eew, (name of the food), eew!" Then she tasted the other food, now expressing pleasure "Hmm, (name of the food), hmm (name of the food) hmm!" These two sequences lasted for approximately 10 seconds each, with the verbal comments fitting with intonation to the expressed emotion. In test phase, the experimenter placed her hand with the palm facing up on the tray between both bowls, moved the tray towards the child again and asked "Can you give me some, please?" right before the child could reach the bowls. To avoid biasing the child's reaction, the experimenter did not look at any of the bowls but at the child's face. If the child did not respond, the experimenter repeated the request up to two times.

Figure 4: Divergent desire task.

It was coded from videotape whether the child's first action was to give the experimenter the food that the child (*0*) or the experimenter (*1, understanding of divergent desires*) preferred. Thirteen of the 139 children

were excluded because of experimenter failure (n = 9) and interfering behavior of parent (n = 4). Fifty-nine children showed other behavior such as only looking (n = 6) or gesturing (n = 12) at the correct bowl, irrelevant actions like eating (n = 36) or no action at all (n = 5), and were therefore not included in the analyses. A second independent observer coded the performance of 27% randomly selected children. Kappa for children's behavior was .86, $p < .001$.

2.2.2.4 Mirror Self-recognition

A mirror (height: 97 cm, width: 40 cm) covered by a black sheet was placed in front of the child in such a position that the child could only see herself in the mirror. The experimenter encouraged the child to remove the black sheet from the mirror and then ensured that the child took a look at herself. After a short while, the mother inconspicuously placed some red lipstick with her finger on the child's cheek while pretending to wipe the child's nose. Then, the child was encouraged to look into the mirror a second time. After waiting for the child's first response, the experimenter stated three questions based on Povinelli and Simon (1998): 1) she pointed at the child's mirror image and asked "Who is that?", 2) she pointed at the mirror image of the red dot and asked "What is that?", and (3) she asked "I think there is a spot, can you show it to me?" Each question was repeated once, if the child did not respond.

Children's behavior in front of the mirror was coded from videotape. Based on Bischof-Köhler (1994), the children were categorized as *(0) spot negative* when the child located the spot only in the mirror or only gazed at her mirror image, and *(1) spot positive* when the child touched the mark on his cheek or showed other behavior indicating self-recognition such as saying the personal pronoun or her name, exploring her mirror image or body parts only visible in the mirror. One child was not tested because of tiredness. Eleven of the remaining 138 children were excluded because of fussiness (n = 7) and experimenter failure (n = 4). Twenty children did not look sufficiently long into the mirror due to distress or disinterest and were thus excluded from analyses. A second independent observer coded the performance of 24% randomly selected children. Kappa for children's behavior was .88, $p < .001$.

Figure 5: Mirror Self-recognition: Rouge Test.

2.2.3 Receptive language ability

Parents completed the "Elternfragebögen für die Früherkennung von Risikokindern" (ELFRA-1, Grimm & Doil, 2006) at both age points, a parental questionnaire to identify children at risk for developmental language disorders. Although the questionnaire is generally used to detect a delay in language development judged at critical values, it can also be used to assess the actual state of the infants' language skills. As a measure of language ability for each age, a Word Comprehension score was computed by summarizing the scores of the subscales receptive vocabulary and reactions to sounds (a maximum of 171 can be reached). At 12 months of age, the ELFRA-1 Word Comprehension score ranged from 6 to 142 (M = 46.17, SD = 29.20, n = 127; 16% of children were at risk with a score < 17), at 18 months from 30 to 171 (M = 121.45, SD = 31.47, n = 134).

3 Results

3.1 Strategy of Data Analysis

The main objective of this study was to investigate the relations between the examined social-cognitive abilities. First, point-biserial or Pearson product-moment correlations and phi coefficients were used to assess whether the single joint attention abilities were interrelated. Second, we conducted point-biserial correlations and phi coefficients to examine the relation between all assessed social-cognitive abilities. One-tailed tests were used when a direction could be expected based on previous findings. In the following section, we will first report descriptive results and then report correlational data focusing on the main question regarding a relation between the single putative ToM precursor abilities. In addition to significant results ($p < .05$), marginally significant results ($.05 < p < .06$) and trends ($.06 < p < .10$) are reported as well.

3.2 Joint Attention

In at least one of the four object spectacle phases, 95,3% of the 12-month-olds (121 out of 127) initiated joint attention by looking at the experimenter, 48% of the children (61 out of 127) did so by showing her a toy, 12% of the children (15 out of 127) showed declarative pointing, and 86 % of the children (109 out of 127) showed imperative pointing. In the goal-detection tasks, 72,9% and 57% of the children (89 out of 122, 65 out of 114) looked at the experimenter's face in *Teasing* and *Blocking*, respectively. 51,2% of the children (64 out of 125) understood the behavioral request. In the attention following (AF) task, four different scores were analyzed. Table 1 shows the rate of gaze and point following behavior in each of the four AF trials as well as summarized performance scores as a function of gesture (Gaze and Point) or position of the target (Side and Back).

Table 1: *Attention Following measures*

Number (and percentages) of infants at 12 months (n = 127) showing gaze and point following in each of the four AF trials and as a function of gesture (Gaze and Point) or position of the target (Side and Back).

	no AF	Level 0	Level 1	Level 2
Gaze Side	24	47	30	26
Gaze Back	15	76	29	4
Point Side	2	34	30	61
Point Back	4	82	33	6

	less than Level 1 in both trials	at least Level 1 in one trial (1)	at least Level 1 in both trials (2)
AF Side	28 (22,0)	51 (40,2)	48 (37,8)
AF Back[a)]	68 (55,7)	38 (31,1)	16 (13,1)
AF Gaze[b)]	61 (49,2)	38 (30,6)	25 (20,2)
AF Point[c)]	32 (25,6)	57 (45,6)	36 (28,8)

Notes: AF = Attention Following. Gaze back is missing for 3 children and Point back for 2 children because of their fussiness. [a)] n=122. [b)] n=124. [c)] n=125.

3.3 Pretend Play

Table 2 displays descriptive information regarding levels of pretend play. In collaborative play with the mother 14 % of children showed non-symbolic play (levels 1-4) and 86 % of children showed symbolic play (levels 5-8). In solitary play, 34 % of children showed non-symbolic play and 66 % of children showed symbolic play. The occurrence of pretend play in children's solitary play and collaborative play was related, *phi* = .257, *p* = .003.

Table 2: *Pretend Play Levels*

Frequencies (and percentages) of children (n=138) who reached the play levels 1-8 in collaborative and solitary play

	Non-symbolic play				
Play scenario	1	2	3	4	**1-4**
Collaborative play	3 (2)	0 (0)	7 (5)	10 (7)	**20 (14)**
Solitary play	10 (7)	4 (3)	12 (9)	20 (15)	**46 (34)**
	Symbolic play				
Play scenario	5	6	7	8	**5-8**
Collaborative play	23 (17)	43 (31)	44 (32)	8 (6)	**118 (86)**
Solitary play	27 (20)	35 (25)	24 (17)	6 (4)	**92 (66)**

3.4 Imitation

In the *dumbbell* task, 12 of the remaining 136 children were excluded because of aversion to take the object (n = 6), experimenter failure (n = 5) and interfering behavior of parent (n = 1). In the baseline, already 66 children showed the target behavior and pulled the dumbbell apart, in the test further 18 children, in sum 84, did so, and 40 children failed to do so. In the *prong and loop* task, 13 children were excluded because of aversion to take the object (n = 6), experimenter failure (n = 2) and interfering behavior of parent (n = 5). In the baseline, thirty children showed the target behavior and put the loop on the prong, in the test further 51 children, in sum 81, did so, and 42 children failed to do so. Due to the high amount of children that showed the target behavior already in the baseline of the dumbbell task, only the prong and loop condition will be further analyzed.

3.5 Divergent Desires

Forty-two (33,3 %) of 126 children took into account the experimenter's divergent desire and gave the experimenter her preferred food. Forty-three (34,1 %) children failed the task and gave the experimenter his own

preferred food. Forty-one (32,5 %) children showed irrelevant behavior, primarily eating.

3.6 Mirror Self-recognition

Fifty (39,4 %) of 127 children located the spot in the mirror or only looked at their mirror image (spot negative behavior), 38 (29,9 %) children located the mark on their cheek, tried to touch or remove it (spot positive behavior), and 19 (15,0 %) children explored or played with their mirror image, said “me” or their own name (self positive behavior), the latter two behavior categories were rated as indexing mirror self-recognition. Further twenty (15,7 %) children were excluded from analysis because they were afraid of the mirror (n = 1) or showed disinterest as they rather wanted to play (n = 19).

3.7 Intercorrelations between putative precursor abilities

Point-biserial correlations and phi coefficients were used to assess whether the single precursor abilities were interrelated. One-tailed tests were used when a direction could be expected based on previous findings. Children’s declarative pointing at 12 months is expected to be related to their understanding of other’s intentions (Camaioni, Perucchini, Bellagamba & Colonnesi, 2004; Colonnesi et al., 2008; Kristen, Sodian, Thoermer, & Perst, 2011) and to their mirror self-recognition at 18 months of age (Sodian & Kristen-Antonow, 2015). Also, infants’ performance on the blocking and teasing tasks at 12 months should be related to their understanding of other’s intentions at 18 months of age (Charman et al., 2000). In addition, mirror self-recognition and children’s pretend play are expected to be related at 18 months of age (Lewis & Ramsay, 2004).

Table 3: *Intercorrelations of the joint attention abilities at 12 months*

	Imperative Pointing	AF Side	AF Back	AF Gaze	AF Point	Teasing	Blocking	Initiating JA	Show	Behavior Request
Declarative Pointing	-.131 (n = 127)	.149+ (n = 127)	.190* (n = 127)	.110 (n = 127)	.220* (n = 127)	-.109 (n = 122)	.055 (n = 114)	-.034 (n = 127)	.039 (n = 127)	.213* (n = 125)
Imperative Pointing		.055 (n = 127)	.103 (n = 127)	.047 (n = 127)	.105 (n = 127)	.059 (n = 122)	.013 (n = 114)	.016 (n = 127)	-.016 (n = 127)	-.061 (n = 125)
AF Side			.437*** (n = 127)	.749*** (n = 127)	.657*** (n = 127)	-.003 (n = 122)	-.006 (n = 114)	-.003 (n = 127)	.029 (n = 127)	.006 (n = 125)
AF Back				.655*** (n = 127)	.717*** (n = 127)	.135 (n = 122)	.037 (n = 114)	-.084 (n = 127)	.076 (n = 127)	.015 (n = 125)
AF Gaze					.343*** (n = 127)	.170+ (n = 122)	.009 (n = 114)	-.086 (n = 127)	-.015 (n = 127)	-.032 (n = 125)
AF Point						-.055 (n = 122)	.021 (n = 114)	.007 (n = 127)	.119 (n = 127)	.054 (n = 125)
Teasing							.050 (n = 109)	-.138 (n = 122)	.082 (n = 122)	-.055 (n = 120)
Blocking								.271** (n=114)	.043 (n=114)	.133 (n=113)
Initiating JA									.092 (n = 127)	.080 (n = 125)
Show										.009 (n = 125)

Notes: $+p < .10$. $*p < .05$. $**p < .01$. $***p < .001$. AF = Attention following. JA = Joint Attention. Number of children (n) in parentheses.

Table 3 shows the intercorrelations of the joint attention behaviors. All four AF scores were intercorrelated (*r*s between .343 and .749, *p*s < .001). In addition, declarative pointing correlated with behavioral request (*phi* = .213, *p* = .026) as well as with AF to the back or AF with Point (r_{pb}s .190 and .220, *p*s < .032). Also, there was a relation between initiating joint attention and reaction to blocking behavior (r_{pb} = .271, *p* = .005). The abilities in the goal-detection tasks teasing and blocking were not interrelated (*p*s > .59).

Tables 4 shows the cross-sectional correlations at 18 months of age. As supposed, mirror-self recognition was found to be correlated with the occurrence of pretence in both collaborative (*phi* = .215, $p_{one\text{-}tailed}$ = .013, *n* = 107) and solitary play (*phi* = .186, $p_{one\text{-}tailed}$ = .028, *n* = 107). Unexpectedly, understanding of divergent desires was found to be negatively correlated to mirror-self recognition (*phi* = -.254, *p* = .038, *n* = 67) and the occurrence of pretence in solitary play (*phi* = -.185, *p* = .088, n = 85).

Table 4: *Intercorrelations at 18 months*

	Divergent Desire	Mirror Self-recognition	Pretend Play (collaborative)	Pretend Play (solitary)
Imitation	-.063 (n = 79)	-.099 (n = 98)	.006 (n = 125)	-.116 (n = 124)
Divergent Desire		-.254* (n = 67)	-.036 (n = 85)	-.185+ (n = 85)
Mirror Self-recognition			.215* (n = 107)	.186* (n = 107)
Pretend Play (collaborative)				.257** (n = 138)

Notes. +*p* < .10. **p* < .05. ***p* < .01. ****p* < .001.

Table 5 shows the longitudinal correlations between the abilities at both ages. As expected, joint attention abilities were longitudinally related to intention-based imitation as declarative pointing (*phi* = .126, $p_{\text{one-tailed}}$ = .090, *n* = 113) and performance on the teasing task (*phi* = .157, $p_{\text{one-tailed}}$ = .050, *n* = 109) correlated with imitation.

Table 5: *Longitudinal correlations between abilities at 12 and 18 months*

	Divergent Desire	Imitation	Mirror Self-recognition	Pretend Play (collabo-rative)	Pretend Play (solitary)
Declarative Pointing	-.088 (n = 79)	.126 (+) (n = 113)	-.029 (n = 96)	-.052 (n = 127)	-.091 (n = 126)
Imperative Pointing	.081 (n = 79)	.018 (n = 113)	-.075 (n = 96)	.019 (n = 127)	.129 (n = 126)
AF Side	-.051 (n = 79)	-.071 (n = 113)	.059 (n = 96)	.146 (n = 127)	.022 (n = 126)
AF Back	-.066 (n = 79)	.089 (n = 113)	.014 (n = 96)	.055 (n = 127)	-.114 (n = 126)
AF Gaze	-.160 (n = 79)	.079 (n = 113)	.021 (n = 96)	.094 (n = 127)	.002 (n = 126)
AF Point	.053 (n = 79)	-.073 (n = 113)	.021 (n = 96)	.103 (n = 127)	-.089 (n = 126)
Teasing	-.259* (n = 75)	.157+ (n = 109)	-.118 (n = 93)	-.149(+) (n = 122)	-.064 (n = 121)
Blocking	-.244* (n = 73)	.050 (n = 101)	.076 (n = 85)	.135 (n = 114)	.157+ (n = 113)
Initiating JA	-.228 (n = 79)	-.035 (n = 114)	.104 (n = 97)	-.004 (n = 127)	.259** (n = 126)
Show	-.113 (n = 79)	-.026 (n = 114)	.075 (n = 97)	-.042 (n = 127)	.092 (n = 126)
Behavior Request	-.012 (n = 77)	-.055 (n = 111)	.065 (n = 94)	-.012 (n = 125)	-.197* (n = 124)

Notes: +$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.
AF = Attention following. JA = Joint Attention.

Also, joint attention abilities and pretend play were longitudinally interrelated as initiating joint attention (*phi* = .259, $p = .007$, $n = 126$) and performance on the blocking task (*phi* = .157, $p = .096$, $n = 113$) were related to the occurrence of pretence in solitary play. In addition, the understanding of

behavior request was negatively related to the occurrence of pretence in solitary play (*phi* = -.193, *p* = .030, *n* = 127).

Also longitudinally, analyses revealed unexpected correlations for understanding of divergent desires: it correlated negatively to the performance on the blocking task (*phi* = -.244, *p* = .037, n = 75) and the teasing task (*phi* = -.259, *p* = .025, n = 75).

To control if the performance in any of these precursor abilities was associated with children's receptive language abilities, we ran Spearman correlations and point-biserial correlations between these social tasks and the ELFRA word comprehension score at 12 and 18 months. Only at 12 months, significant correlations emerged between receptive language and the understanding of behavioral request (r_{pb} = .302, *p* = .001) and performance on the teasing task (r_{pb} = .184, *p* = .046). No other significant correlations emerged (*p* > .20).

4 Discussion

The aim of the present study was to longitudinally examine if and to what extend early social-cognitive abilities were interrelated. To this end, we tested infants' joint attention abilities at the age of 12 months and pretend play, intention-based imitation, reasoning about other people's desires and mirror self-recognition at the age of 18 months. The main findings showed a link between joint attention abilities at 12 months and intention-based imitation at 18 months as well as a link between pretend play and mirror self-recognition at 18 months.

First, the link between children's performance on the teasing task at 12 months and their understanding of intention-based imitation at 18 months corresponds to the findings by Charman and colleagues (Charman et al., 2000). The ability to detect that another person pursues a specific goal in the teasing task ("What are you doing?") and the understanding of another person's intention seem to be closely related as both enable the child to conceive the other person's inner intentions. In contrast, in our study there was no relation between performance on the blocking task and the understanding of intention-

based imitation – and there was also no relation between the children's performance in the blocking and teasing task. Only 57% of the children looked at the experimenter to uncover her goal in the blocking task (as compared to 73% in the teasing task). Maybe, the positioning of the object affected children's reactions in both tasks. Whereas the object was out of reach and near the experimenter's face in the teasing task, children still held it in their hands in the blocking task. Potentially, this way the object *per se* still received more attention in the blocking task than in the teasing task so that fewer children paid attention to the experimenter in the former. In addition, it seems that the ability to detect that another person has a specific goal in these tasks had just emerged at 12 months of age. This would explain why there was only a small relation between children's performance on the teasing task and their understanding of other's intentions in our study, but a robust relation between both goal detection tasks and the understanding of intention-based imitation when children were 15 and 20 months of age, respectively (Charman et al., 2000).

In addition, our study revealed a trend for a relation between declarative pointing at 12 months and the understanding of intention-based imitation at 18 months that is consistent with previous work (Camaioni et al., 2004; Kristen et al., 2011). The two authors assumed that both abilities "pave the way toward inferring information about objects or events, such as person's unseen action goals" (Kristen et al., 2011). In our study, only the data of one imitation task could be included in the analysis, whereas in the above-cited studies at least 4 tasks were assessed. This restriction in variance may explain why only a trend was revealed in the current work. Furthermore, the time interval between assessing declarative pointing and intention-based imitation was greater in our study, as the former studies both reported correlations to intention-based imitation with 15-months-olds and therefore the relation between both precursor abilities might have decreased until the age of 18 months.

Furthermore, our results are in accord with previous research reporting a link between pretend play and mirror self-recognition at 18 months (Lewis & Ramsay, 2004). According to Perner (1991), both abilities rely on the capacity

for secondary representation, that is, children are able to represent things not only how they are but how they could be (see also Nielsen & Dissanayake, 2004). In pretend play, children have a secondary representation beyond the real identity or feature of an object in order to pretend another thing; in mirror self-recognition, children have a secondary representation in order to match the mirror image with what they think they look like in reality.

Additionally, we found relations between some joint attention abilities and pretend solitary play. Initiating joint attention by gazing to the experimenter in the object spectacle phases as well as children's performance on the blocking task were both related to symbolic play, but only when children played alone. This result may seem counterintuitive, particularly because both joint attention abilities as well as collaborative symbolic play demand considering another's perspective. However, 86% of the children showed symbolic play when playing together with their parent, thus this restriction in variance may explain why relations were only found for solitary symbolic play.

Surprisingly, our results revealed some negative relations between the early social-cognitive abilities. Children's reasoning about other's divergent desire correlated negatively with performance on the blocking and teasing tasks at 12 months, occurrence of pretence in solitary play at 18 months as well as mirror-self recognition. Examining the results in the divergent desire task, one third of the children did not meet the experimenter's request to give her something to eat, although children were familiar with give-and-take routines through the preceding session. Out of the remaining children, only 49,4 % gave the experimenter the preferred food, thus understood her divergent desire. The low performance of the children in our sample is similar to the results reported by Carlson and colleagues (Carlson, Mandell, & Williams, 2004) showing that even at 24 months of age children performed at chance in the divergent desire task. Similarly, Poulin-Dubois and colleagues found that at the ages of 18, 24 and 30 months only half of the children passed this task (personal communication, June 2014). This might be the reason why the reported negative correlations occurred rather accidental and we desist to interpret them

content-related. Beyond, based on the assumption of a continuity in social-cognitive development, there is only a theoretical assumption that reasoning about other people's desires also belongs to the putative precursor abilities. To our knowledge, there is no empirical evidence for a relation between understanding of divergent desires and Theory of Mind. For example, Carlson and colleagues (Carlson et al., 2004) did not find a relation between children's performance in the Divergent Desire task conducted at two different age points (24 and 39 months) and their ToM at 39 months.

Our results revealed relatively few inter-relations between the single early social-cognitive abilities. One limitation of the present study was the fact that most of the variables were dichotomous rather than interval-scaled variables. Originally, tasks were designed for aggregating the data to a sum score but due to children's performance in baseline (dumbbell in intention-based imitation) or missing correlations (goal detection tasks) only one condition could be analyzed or data had to be analyzed separately.

Based on previous work, we had expected to reveal more inter-relations especially with attention following and understanding of divergent desires. Thus, the current results replicate previous findings only in part. Understanding of another person's intention was related to the joint attention abilities declarative pointing as well as to detecting another person's goal in the teasing task. Additionally, pretend play and mirror self-recognition were related.

In summary, the current work complements earlier findings and adds further evidence in support of continuity in social-cognitive development. Specifically, findings suggest that the assumed continuity in social-cognitive development is not global but task- and age-specific. First, inter-relations between the precursor abilities were rather task-specific than fundamental across the concepts. Second, it seems to have a great effect at what age point the different putative precursor abilities were assessed. Possibly, these abilities are part of an underlying "social cognitive representational ability" (Charman et al., 2000, p. 492) that emerges within a specific timeframe (e.g. second year of life), after which abilities develop along diverging pathways. In our case, alternatively, it might be that some associations were not found as children

were just at the beginning to develop the single abilities as we investigated relatively young children. To test this hypothesis, future studies should systematically assess the respective abilities not only with a single task at one age point but with a variety of complement tasks at different ages in continuous shorter intervals. Maybe in so doing, more (age- and task-) specific cross-sectional and longitudinal associations between the putative precursor abilities could be detected that could otherwise not be discovered by only assessing single abilities at single age points.

Study 2 – Temperament and early Theory of mind abilities

As outlined in the introduction, the relationship between temperament and ToM development has only been investigated in preschoolers or older children. Thus, the general aim of the present study was to extend this line of research by longitudinally investigating the impact of early temperamental characteristics on ToM development. Specifically, we wanted to assess whether individual differences in infant temperament is related to early ToM competence, that is, before the emergence of false belief reasoning. Based on previous work on factors influencing ToM development, we controlled for children's language abilities, parental education and presence as well as number of siblings.

To that aim, we tested children at 18 months and again at 3 years of age: At visit 1, we surveyed infants' temperament and language abilities via parental questionnaires, at visit 2, we surveyed children's temperament via parental questionnaire and tested their language as well as their early ToM abilities. Based on previous work (Lane et al., 2013; Wellman et al., 2011) and with regard to the emotional reactivity hypothesis, we focused in our analyses on the temperament characteristics Shyness, Fear, Perceptual Sensitivity and Anger/Frustration. We expected a positive relation between the first three temperament variables and ToM performance at 3 years of age and a negative relation between Anger/Frustration and ToM.

2 Method

2.1 Participants

A total of 88 children (40 females, 48 males) from a medium-sized city in the southwest of Germany were included in the final sample and tested at 18 months (M = 18.20 months, SD = 0.42, *range* = 17.41 - 19.94 months) and 3 years of age (M = 36.76 months, SD = 0.68, *range* = 36.07 – 39.69 months). Two additional children were excluded from analysis due to insufficient German language abilities. With regard to siblings, at 3 years of age, 32% of

the children were single children, 49% had one sibling, 16% had two siblings, and 3% had three siblings. According to parental report, 76% of the children were monolingual and 24% were bilingual (including 5.5% who were trilingual). Parental educational level was relatively high: 62.5% of mothers and 68% of fathers held a college or university degree, 20.5% of mothers and 18% of fathers completed secondary school at top track (Abitur), and 17% of mothers and 12% of fathers completed secondary school at lower or middle track; 2% of fathers were without school qualification.

Parents were recruited by telephone from a list of families who had earlier expressed interest in volunteering for research on child development. At each visit, they received a recompense for travel expenses and children were given a small gift and a certificate for participating.

2.2 Tasks and materials

2.2.1 Temperament

To assess infant temperament parents completed the German version of the Early Childhood Behavior Questionnaire (ECBQ, Putnam et al., 2006; Rink, 2006) at 18 months and the Children's Behavior Questionnaire (CBQ, Rothbart et al., 2001; Nikolaizig, 2007) at 3 years of age. For each item, parents were asked to rate on a 7-point Likert scale (1 = never, 7 = always), how frequent a specific infant behavior occurred during the preceding 1-2 weeks at 18 months of age and during the last 6 months at 3 years of age, respectively. The ECBQ includes 18 subscales and the CBQ includes 15 subscales. For both questionnaires, the same four subscales were focal to this study. As defined in the ECBQ and CBQ score sheets, *Shyness* characterizes a "slow or inhibited approach in social situations involving novelty or uncertainty" (e.g., "My child sometimes prefers to watch rather than join other children playing."). *Fear* characterizes a "negative affect, including unease, worry or nervousness related to anticipated pain or distress and/or potentially threatening situations" (e.g., "My child is afraid of the dark."). *Perceptual Sensitivity* characterizes a "detection of slight, low intensity stimuli from the external environment" (e.g., "My child seems to listen to even quiet sounds.").

As the ECBQ and CBQ do not include aggressiveness, we focused also on Anger/Frustration, a defining component of aggression (Crick, Bigbee, & Howes, 1996). *Anger/Frustration* characterizes a "negative affect related to interruption of ongoing tasks or goal blocking." (e.g., "My child becomes easily frustrated when tired."). To consider additionally temperament differences that assuming an emotional reactivity hypothesis should not be related to ToM, we included also Activity Level („level of gross motor activity including rate and extent of locomotion"), Attentional Focusing („tendency to maintain attentional focus upon task-related channels; resisting distraction"), and Inhibitory Control („capacity to plan and to suppress inappropriate approach responses under instructions or in novel or uncertain situations") in our analyses (see also Lane et al., 2013, Wellman et al., 2011).

2.2.2 Language

When children were 18 months old, parents completed the "Elternfragebögen für die Früherkennung von Risikokindern" (ELFRA-1, Parental questionnaire to identify children at risk for developmental language disorders, Grimm & Doil, 2006). Although the questionnaire is generally used to detect a delay in language development judged at critical values, it can also be used to assess the actual state of the infants' language skills. The ELFRA-1 assesses both the active and passive vocabulary of 12-18 month old children with 4 developmental scales: Word Production, Word Comprehension, Gestures and Fine Motor Skills. In the analyses, only the first two scales were included as the latter two scales serve as prognostic measures for language impairment. When children were 3 years old, we assessed their language abilities with the "Sprachentwicklungstest für Kinder" (SETK) [Language Development Test for Children]. This German language development test for 3- to 5-year-olds (Grimm, 2001) comprises six subscales. Each subscale assesses a specific aspect of children's syntactic and morphological competences and yields subscale-specific standardized scores (*T*-values) as well as percentile ranks. Four subscales are administered to 3-year-old children: the Sentence Comprehension scale (sentences that differ in syntactic complexity), the Encoding of Semantic Relations scale (description of

interrelations between objects), the Phonological Memory scale (repetition of fantasy names), and the Morphological Rules scale (pluralization of familiar words).

2.2.3 Theory of Mind scale

The German version of the ToM scale (Hofer & Aschersleben, 2007) was employed to assess children's ToM development. These five tasks were closely modeled to the original scale of Wellman and Liu (2004), assessing a range of different developmental attainments (diverse desires, diverse beliefs, knowledge access, contents false belief, and real–apparent emotion). These tasks have been shown to produce a coherent Guttman scale for typical preschoolers in the United States and China (Wellman, Fang, Liu, Zhu, & Liu, 2006; Wellman & Liu, 2004) as well as in Germany when the German version of this scale (Hofer & Aschersleben, 2007) was administered to a sample of 107 German 3- to 5-year-olds (Kristen et al., 2006). Following the manual, for each task, laminated colored cards were used to illustrate a story about a protagonist (a toy figurine) and children were questioned about the protagonist's mental state or action.

Children's ToM abilities at 3 years were assessed employing the German version of the ToM scale (Hofer & Aschersleben, 2007). As children in the current study were tested in the first weeks after their third birthday and given that earlier studies (Henning et al., 2011; Kristen et al., 2006) reported floor effects in the last two of the five tasks for this age group, only the first three tasks were administered. In the Divergent Desires task, the child needs to differentiate his or her own desire (e.g., preference for a cookie over a carrot) from another person's differing desire about the same food items (e.g., preference for a carrot) to correctly predict the other person's snack choice (e.g., the carrot and not the cookie). In the Diverse Beliefs task, the child needs to differentiate his or her own belief about the location of an object (e.g., the cat is hiding in the bush) from another person's differing belief about the location of same object (e.g., the cat is hiding in the garage) to correctly predict the other person's action (e.g., look for the cat in the garage). The Knowledge Access task requires an understanding of the causal relation between seeing

and knowing. In this task, the child is first shown the content of a box (toy figurine) and then asked to judge whether another person who did not have visual access to the content of the box, knows the content of the box. All tasks include a focal test question. The Knowledge Access task also includes a memory control question to ensure that children remembered that the protagonist had never looked into the box. Following Wellman and Liu (2004), tasks were administered in a fixed order with increasing degree of difficulty: diverse desires, diverse beliefs, knowledge access.

2.3 Procedure

Children were tested in a laboratory room. At 18 months of age, the ELFRA-1 questionnaires were mailed to parents together with the ECBQ, 1-2 weeks prior to the respective test session. At test scheduling, the primary caregiver was advised to fill out the questionnaires and to hand them back to the experimenter at testing. At 3 years of age, procedure was the same for the CBQ. Administration of the SETK 3-5 and the ToM scale took place in a quiet test room with the experimenter and the child sitting at a table. The SETK 3-5 was administered prior to the ToM scale.

2.4 Scoring

2.4.1 Temperament.

The average score for each of the subscales was calculated for parents' ratings at each visit. At 3 years of age, the temperament data from one child was excluded because the parents did not complete the CBQ.

2.4.2 Language development.

As a measure of language ability at 18 months of age, a Word Production score was computed by summarizing the scores of the subscales productive vocabulary and speech sounds. Also a Word Comprehension score was computed by summarizing the scores of the subscales receptive vocabulary and reactions to sounds. The data from one child was missing because parents did not hand back the ELFRA-1.

As a measure of language ability at 3 years of age, a total mean *T*-score (*M* = 50.00, *SD* = 10.00) of the four SETK subscales was used for analyses. The SETK data for all subscales of two children were excluded, because the testing had to be aborted already during or after the administration of the first subscale. The data on phonological memory of 16 children were excluded due to children's refusal to participate (n = 12) or uneasiness (n = 4). The data on morphological rules of 8 children were excluded due to children's refusal to participate.

2.4.3 Theory of Mind scale.

Children needed to answer the focal test question, as well as the control question in task 3 ("Has the protagonist ever looked into the box?"), to count as passing this task. All children that failed this control question (n = 27) also failed the test question. For each of the tasks, children received a 0 (fail) or a 1 (pass). The resulting ToM total score ranged between 0 (no task solved) and 3 (all tasks solved). Please note, that no false-belief task was administered here.

3 Results

In the following sections, we will first describe the main results obtained for each measure separately and then focus on the main question regarding a relation between infant temperament and early ToM capacities at age 3. In addition to significant results ($p < .05$), marginally significant results ($.05 < p < .06$) and trends ($.06 < p < .10$) are reported as well. Effect sizes are reported as *f* with the ranges small: $.10 \leq f < .25$, medium: $.25 \leq f < .40$, and large: $f \geq .40$ (Cohen, 1988).

Preliminary analyses

As 24% of the children were bilingual, their SETK and ToM performance was compared to that of monolingual children. Bilingual children (*M* = 43.89, *SE* = 13.2) had a significantly lower SETK Total score than monolingual children (*M* = 51.42, *SE* = 7.09), $t(1,84) = 3.31$, $p = .001$, but ToM scale scores did not differ between these groups ($p > .70$).

3.1 Temperament

Table 6 displays descriptive information regarding parental ratings of infant and child temperament at 18 months and 3 years of age. Additionally, it displays Pearson product-moment correlations (or Spearman Rho if scales were not normally distributed) that were computed to assess the stability of infant temperament ratings and paired *t*-tests (or Wilcoxon tests) to assess developmental change. Scale results indicated moderate to high normative stability in parental temperament ratings from 18 months to 3 years (*r*s between .226 and .562, *p*s < .02, for Fear, r_{sp} = .186, *p* = .086; see Table 6). In addition, scores significantly increased from 18 months to 3 years for Attention Focusing, Fear, Frustration, Inhibitory Control, and Perceptual Sensitivity (*p*s < .001; see Table 6). No significant change was observed for Activity Level and Shyness (*p*s > .33).

Table 6: *Parental Temperament ratings at ages 18 months and 3 years*

Means (M) and standard deviations (SD) of parental temperament ratings at ages 18 months (n=88) and 3 years (n=87) and their longitudinal correlations

ECBQ/CBQ subscales		*M (SD)*			
	Number of items	18 months	3 years	*r*	*t/Z*
Activity Level [a)]	12 / 13	4.92 (.84)	5.00 (.73)	.357***	-.969
Attention Focusing [a)]	12 / 14	4.08 (.92)	4.58 (0.62)	.301**	-4.534***
Fear [a)]	11 / 12	2.05 (.62)	3.42 (.93)	.186^{+}	-7.658***
Frustration / Anger	12 / 13	3.42 (.92)	4.26 (.92)	.562***	-9.088***
Inhibitory Control	12 / 13	3.46 (.95)	4.57 (.84)	.540***	-11.960***
Perceptual Sensitivity	12 / 12	4.69 (.96)	5.51 (.66)	.226*	-7.185***
Shyness [a)]	12 / 13	3.12 (.96)	3.27 (1.15)	.446***	-.817

Note. [a)] Scales are not normally distributed at 18 months why non-parametric comparisons were used. Sample sizes for correlations ranged between n = 69 and n = 87. $^{+}p < 0.10$. $*p < 0.05$. $**p < 0.01$. $***p < 0.001$. All *p*s two-tailed.

Gender differences were found in only few of the temperament characteristics. At 18 months of age, ANOVAS revealed differences in Inhibitory Control, $F(1,86) = 4.72$, $p = .033$, $f = .23$, Frustration, $F(1,86) = 4.36$, $p = .040$, $f = .22$, and Perceptual Sensitivity, $F(1,85) = 2.87$, $p = .094$, $f = .18$. Girls ($M = 3.69$, $SE = .15$) were rated as possessing more inhibitory control than boys ($M = 3.26$, $SE = .13$). Also, boys ($M = 3.60$, $SE = .13$) were rated as showing more negative affect when being frustrated than girls ($M = 3.20$, $SE = .14$), and boys ($M = 4.85$, $SE = .13$) were rated as being more perceptually sensitive than girls ($M = 4.51$, $SE = .16$). At 3 years of age, results revealed differences in Inhibitory Control, $F(1,85) = 4.11$, $p = .046$, $f = .22$. Again, girls ($M = 4.77$, $SE = .12$) were rated as possessing more inhibitory control than boys ($M = 4.41$, $SE = .13$).

3.2 Theory of Mind

The ToM scale score ranged from 0 to 3 ($M = 1.25$, $SD = 0.81$, n = 88). Eighteen children (20.5%) did not pass any ToM task, 32 (36.5%) children passed one task, 36 children (41%) passed two tasks and two children (2%) passed all three tasks (see Table 7 for rate of success in single tasks). No gender difference was found in the total ToM scale score ($p > .79$).

Table 7: *Theory of Mind performance at 3 years of age*

Frequencies (and percentages) of children who passed the individual ToM tasks (n=88)

ToM Scale	*n*
Diverse Desires	57 (64.8%)
Diverse Beliefs	47 (53.4%)
Knowledge Access	6 (6.8%)

3.3 Language

At 18 months of age, the ELFRA-1 Production score ranged from 7 to 134 ($M = 42.26$, $SD = 30.21$, n = 86), the Comprehension score ranged from 30

to 171 (M = 122.71, SD = 31.31, n = 85). The data of one child was excluded in word production because of an extreme value (167 words) that was 3 standard deviations above the mean value. A gender difference was found in the Production score, $F(1,84) = 16.54$, $p < .001$, $f = .44$, but not in the Comprehension score ($p > .56$). Girls (M = 56.03, SE = 5.84) were rated as having a greater active vocabulary than boys (M = 31.35, SE = 2.80).

At 3 years of age, children's T-values were in the normal range of language development on the four subscales of the SETK 3–5: Sentence Comprehension score (M = 50.04, SD = 9.01, range 29 – 71, n = 85), Encoding of Semantic Relations score (M = 53.47, SD = 9.79, range 32 – 75, n = 85), Phonological Memory score (M = 47.43, SD = 9.81, range 29 – 67, n = 72), and Morphological Rules score (M = 50.11, SD = 10.68, range 33 – 72, n = 80). The SETK total score ranged from 34.5 to 68.25 (M = 49.75, SD = 9.21). No gender differences were found (ps > .17).

Spearman Rho correlations revealed a stability of language abilities across the two visits: both the ELFRA-1 Production score (r_{sp} = .369) and the Comprehension score (r_{sp} = .302) were related to the SETK total score (ps < .005). No significant relation between any of the language scores and the ToM scale score was found (ps > .70).

3.4 Siblings and Theory of Mind

The presence or absence of siblings was not significantly related with the ToM scale score (p = .778). Also, there was no significant relation between the number of siblings, or number of only older siblings, and children's performance on the ToM scale (ps > .69).

3.5 Parental education and Theory of Mind

Neither maternal nor paternal educational level was significantly related with the ToM scale score (p = .660 and p = .746, respectively). Additionally, there was no difference in ToM performance between children who's mothers or fathers had higher education (Abitur or university degree) compared to a lower educational level (ps = .66).

3.6 Temperament and Theory of Mind

Pearson or Spearman's Rho correlations were performed to assess whether temperament was related to early ToM abilities at 3 years. Two of the seven ECBQ-subscales were found to be correlated with the ToM scale score: Shyness (r_{sp} = .297, p = .005) at 18 months was positively related to ToM, whereas Activity Level (r_{sp} = -.248, p = .020) at 18 months was negatively related to ToM. Cross-sectionally, also two CBQ-subscales were found to be correlated with the ToM scale score: Attention Focusing (r_{sp} = .261, p = .015), and Shyness (r_{sp} = .185, p = .086) at 3 years were positively correlated to ToM, although shyness only revealed a trend.

To accurately assess the contribution of temperament to early ToM development at 3 years, we ran two different multiple regression analyses simultaneously including all seven temperament characteristics, one for each age in order to preclude multicollinearity.

At 18 months, Shyness and Activity Level significantly predicted ToM, $F(7,78)$ = 2.57, p = .020, accounting for 19% of the variance in early ToM abilities (see Table 8, Model A). Consistent with prior findings, higher Shyness was associated with better early ToM abilities, and higher Activity Level was

Table 8: *Relations between children's temperament and their Theory of mind abilities at 3 years of age*

	early ToM abilities		
predictors	*B*	*SE B*	*β*
	Model A: Temperament at 18 months, n = 86		
Shyness	.259	0.099	.309*
Activity Level	-.320	0.110	-.336*
	Model B: Temperament at 3 years, n = 87		
Shyness	.160	0.073	.230*
Attention Focusing	.348	0.135	.270*

Note. * p < .05

associated with poorer early ToM abilities. At 3 years, the model including all 7 temperament characteristics was not significant, so we ran a model with only Shyness and Attention Focusing that correlated with ToM. Shyness and Attention Focusing significantly predicted ToM, $F(2,84) = 4.92$, $p = .010$, accounting for 10.5% of the variance in early ToM abilities (see Table 8, Model B). Higher Shyness and Attention Focusing was associated with better early ToM abilities.

In two hierarchical regression analyses, one for each age, we assessed the relation between temperament and ToM controlling for other influences. At step 1, gender, language abilities, number of siblings and maternal education were entered, then at step 2, the 7 temperament characteristics. In both models, the additional control variables did not explain a significant portion of the variance of the ToM scale score. At 18 months, the R of the overall model was significantly different from zero, $F(7,68) = 2.352$, $p = .033$, $R^2 = .20$. Still, only Shyness ($B = .28$, $t = 2.48$, $p = .015$) and Activity Level ($B = -.31$, $t = -2.61$, $p = .011$) were significantly related to ToM. The unique contributions to the total variance of the ToM scale score were 7.2% for Shyness and 8% for Activity Level (indicated by squared part correlations sr^2). At 3 years, the overall model was not significantly different from zero ($p = .75$) and the relations between ToM and Shyness ($B = .16$, $t = 1.79$, $p = .078$) as well as Attention Focusing ($B = .33$, $t = 1.70$, $p = .093$) were only a trend.

Following Banerjee and Henderson (2001) and Wellman and colleagues (2011), we also added some temperamental interaction effects, shy-fearful and shy-angry temperament, in a regression analysis, but this, too, had no further predicting effect on ToM.

To control beyond, if the reported relations between temperament and ToM were special for social cognition and that there were no relations to a non-social cognition task, we assessed the relation between temperament and language abilities. Attention Focusing ($r_{sp} = .221$, $p = .041$) at 18 months was positively related to the SETK total score, whereas Fear ($r = -.291$, $p = .007$) at 18 months was negatively related to the SETK total score. No relations were found for temperament at 3 years (ps > .20). In sum, there were no relations

found between language and the ToM-related temperament characteristics Shyness and Activity Level. Only Attention Focusing seemed to be important for both, non-social as well as social cognition, as Attention Focusing at 18 months was related to language and Attention Focusing at 3 years to ToM.

4 Discussion

The aim of the present study was to longitudinally examine the impact of infant temperament at 18 months of age on Theory of Mind development in children at 3 years of age. The main findings showed that children rated as shyer and less active at the end of infancy, showed a better ToM understanding one and a half years later. Also, children rated as shyer and more attentionally focused at 3 years, showed a better ToM performance. The findings extend previous work by suggesting a relation between individual differences in temperament and ToM development well before the emergence of false belief understanding.

Theory of Mind. Children's performance on the first three tasks of the ToM scale replicated previous patterns of results suggesting a developmental progression from understanding discrepant desires to understanding discrepant beliefs (Henning et al., 2011; Kristen et al., 2006). Also consistent with this previous work (Henning et al., 2011; Kristen et al., 2006 for Order 1), the majority of 3-year-olds did not consider visual access when asked about the story character's knowledge about the content of a box (knowledge access). In fact, while about two-thirds of the children understood discrepant desires and about half of them understood discrepant beliefs, only six children passed the knowledge access task. This great drop in success rate may be explained by differences between tasks in the role of the actual state of affairs in relation to the mental state reasoned about. According to Searle (1983), epistemic states such as beliefs have a mind-to-world fit (i.e., the belief may be at fault, but not the world), whereas desires and volitional states have a world-to-mind fit (i.e., the world may be at fault in that it does not comply with or fulfill the desire or volitional state). In all three tasks employed here, the child needs to understand the subjectivity of mental states (the child's desire, belief or knowledge state differs from that of the story character). However, only in the knowledge

access task is the child also required to understand that this difference results from a difference in experience of – or relation to – the real state of affairs (a toy figurine in the box; please note that in the discrepant belief task no information is given on the actual hiding place of the cat). Overall, the average scores on the ToM scale were below the performance of 3-year-olds reported in other German samples (Henning et al., 2011; Kristen et al., 2006). Binomial tests for each task revealed that only in the first task performance is significantly above chance level (p = .007). This discrepancy can be explained by differences in average age and age range. Whereas the average age in the two previous German studies was 43 and 42 months, respectively, children were on average 37 months old in the current study. Also, while children's age in the previous samples ranged up to 3 years and 9-11 months, children in the current sample were tested around their third birthday.

Temperament. Parental ratings of child temperament showed both stability and changes from second to third year of life. Correlation coefficients for 6 out of the 7 considered temperament subscales, ranged between .23 and .56, indicating moderate to high normative stability in infant temperament between 18 and 36 months of age. Overall, and consistent with Putnam and colleagues (Putnam, Rothbart, & Gartstein, 2008), results indicate a continuity of temperament from infancy to early childhood. In addition to the reported moderate to high normative stability, there were significant changes in average ratings for 5 out of the 7 subscales suggesting that the temperament domains considered here are also subject to developmental change between 18 and 36 months of age. There was a significant increase in the ratings for Attention Focusing, Fear, Frustration, Inhibitory Control, and Perceptual Sensitivity. No change was observed in the ratings for Activity Level and Shyness. Saudino and Cherny (2001) discuss genetic factors as well as new nonshared environmental influences as possible factors that might have an impact on developmental change at this age. One should also keep in mind that the observation window differs respectably between the ECBQ (2 weeks) and the CBQ (6 months). Hence, it is more likely to observe a specific behavior when the observation window is longer, which might also explain the overall increase in means between visits.

Influence of temperament on ToM. Correlational analyses as well as confirmatory regression analyses showed that Shyness at 18 months as well as at 3 years of age was positively related to children's performance on the ToM scale at 3 years. Children rated by their parents as shy were better in reasoning about others' mental states. These results are consistent with those found by Wellman and colleagues (2011) who postulate that shyness in terms of a more quietly observant stance towards human interactions yield to a better insight into interpersonal processes.

Different to Wellman and colleagues (2011) we did not find a relation between Perceptual Sensitivity and ToM nor between aggressiveness and ToM. It is possible that Perceptual Sensitivity becomes more important for an observational attitude in interactions later in preschool age, or that it is more strongly related to the understanding of false-belief tasks and therefore contributes only little to early ToM understanding at the age of 3. Additionally, the lack of a relation between aggressiveness and ToM in our study might be explained by the fact, that we only investigated one component of aggressiveness, namely Anger, but no other component as for example harmful behavior towards others.

Furthermore, analyses showed that Activity Level at 18 months was negatively related to children's performance on the ToM scale at 3 years. As this relation was only found for Activity Level at 18 months, but not at 3 years, we inspected the Activity Level scores at both ages more thoroughly and conducted further analyses. The Activity Level scores scattered more at 18 months and there were four extreme values that differed more than 2 standard deviations from the mean value. By omitting these four scores from analysis, the correlation between Activity Level at 18 months and ToM at 3 years was no longer significant (r_{sp} = -.179, p = .102; n=84). Hence, the relation between Activity Level and ToM is not a meaningful one and will not be further discussed.

Moreover, analyses showed that Attention Focusing at 3 years of age was positively related to children's performance on the ToM scale. However, as we also found relations between Attention Focusing and language abilities,

Attention Focusing might not be an influencing factor specific to social-cognitive abilities but rather influences cognitive development per se.

Influence of siblings, parental education and language on ToM. In addition to investigating the influence of child temperament on children's early ToM abilities, we controlled for three further factors – siblings, parental education and language abilities – that are thought to influence ToM development. Previous research pointed to a beneficial effect of siblings (e.g., Cassidy et al., 2005; Ruffman et al., 1998), maternal education (e.g., Cutting & Dunn, 1999; Pears & Moses, 2003) as well as language (e.g., Milligan et al., 2007).

In the current study, the presence and number of siblings as well as the presence of one or more older siblings did not have a positive effect on children's ToM development. An explanation for the absence of a sibling effect in our study might be that the children were simply too young. Ruffman and colleagues (Ruffman et al., 1998) reported that children younger than 39 months were not able to benefit from their siblings regarding belief understanding. In our sample children averaged only 36.7 months. Following their argument, it is likely that children have to reach a certain precondition before siblings can exert a positive influence on them. This precondition might be maturational readiness, knowledge acquired through learning, a certain threshold of interaction or a change in interaction among siblings due to children's maturation (Ruffman et al., 1998).

Neither maternal nor paternal education had an impact on ToM development in the current study. In sum, parental educational level in our sample was relatively high with 83% of mothers and 86% of fathers holding Abitur or a college or university degree, what leaded to restricted variance. But also, maybe the social interactions in these families did not substantially differ from families whose parents had a lower educational level and therefore, children did not show differences in ToM development. Still, the current findings are consistent with some previous studies that did not find a relation between maternal education and ToM (Dunn et al., 1991; Ruffman et al., 1999).

Similarly, in the current study there was no evidence for an influence of language abilities on ToM development, neither longitudinally at 18 months of age nor cross-sectionally at 3 years of age. This finding, however, is in line with earlier studies that also found no correlation between ToM and language abilities (Aschersleben et al., 2008; Thoermer et al., 2012; Wellman et al., 2004; Wellman et al., 2008). Thus evidence concerning the relation between language abilities and ToM competencies is somewhat inconsistent and further research is needed to clarify this point.

In summary, our results extend earlier findings by Wellman and colleagues (Lane et al., 2013; Wellman et al., 2011) who demonstrated that certain temperament factors reflecting social-emotional reactivity predict children's ToM development from early childhood to preschool age. In the current work, such a relationship was even found from infancy to early childhood, but only for Shyness. As Shyness was not related to language abilities, the reported relations between Shyness and ToM likely are specific for social-cognitive development, and do not hold for cognitive development in general. It seems that, observant attitude children benefit in their social-cognitive development despite—or perhaps precisely because of— this kind of passive attitude. Lane and colleagues (2013) argue that even though shy or withdrawn children may not actively participate in social interactions, they still learn from them by merely attending to them. Additionally, ToM development is by default assessed via cognitive insights into other's mental states and is not assessed in ongoing interactions with others where children have to use their social-cognitive skills (Wellman et al., 2011).

In sum, there is evidence that shyness fosters the social-cognitive development already in infancy. Even at this young age, and well before the emergence of false belief understanding, inferences about ToM development from individual differences in temperament can be drawn.

Study 3 – Dropout in visual habituation paradigms: The role of temperament

The first aim of this study was to assess whether dropout rates in two visual habituation tasks varied as a function of individual differences in infant temperament assessed at 6 and 12 months of age. To the best of our knowledge, all previous studies examining the influence of temperament on dropout rate were cross-sectional. The second aim of this longitudinal work was therefore to assess stability of infant temperament as well as stability of dropout rate by employing the same tasks at each age tested. The third aim was to assess whether dropout rates differed as a function of specific content of the task. The two visual habituation tasks employed in the present study only differed in content while experimental set-up and procedure were kept constant across tasks. In addition, we explored whether dropout rates differed as a function of temporal order of tasks. Finally, given that previous work points to a possible effect of temperament in interaction with infant gender on task performance, infant gender was included into the analyses.

Infants were invited to the lab when they were 6 and 12 months old and tested in two visual habituation experiments, one on perception of human goal-directed behavior and one on perception of physical causality. At both age points, parents rated their infants' temperament by completing the IBQ-R prior to testing. Based on previous studies (Slaughter & Suddendorf, 2007), we expected a moderate number of dropouts in the current sample. Also based on previous work (Carranza Carnicero et al., 2000), we predicted a general stability of temperament across the two age points as well as an impact on dropout of the temperamental domains Distress to Limitations, Duration of Orienting and Activity Level. Distress to Limitations includes fussing and crying while being motor constrained or frustrated. As in habituation experiments infants are typically held on their parents' lap or placed in a baby seat, we predicted non-completers to have higher scores in Distress to Limitations than completers. Duration of Orienting serves as an indicator for infant attention to a specific object or event which is the behavior assessed in

looking time tasks. Thus, we predicted completers to have higher scores in this subscale compared to non-completers. Finally, since an increased Activity Level seems to have an impact on dropout (Miceli et al., 1998; Treiber, 1984), we predicted a higher dropout rate at 12 months of age, when infants have progressed in their motor development, as compared to 6 months. Given the exploratory nature of the study with regard to content of task and temporal order of tasks, no specific predictions were made.

2 Method

2.1 Participants

A total of 80 infants (34 females, 46 males) were included in the final sample and tested at 6 months (*M* = 190.73 days, *SD* = 8.27, *range* = 159-213 days) and 12 months (*M* = 364.33 days, *SD* = 11.58, *range* = 324-397 days) of age. An additional 11 infants participated in the study but were excluded because one of the two habituation tasks could not be administered to infants due to fussiness (n = 4), or because parents did not complete the temperament questionnaire for one of the two age points (n = 7). The data from further 22 infants were excluded because the number of missing items (no response to the item) in the questionnaire exceeded 25% or because the number of items that were reported as not applicable exceeded 50% in at least one subscale. With regard to siblings, 52.5% of the children were single children, 32.5% had one sibling, and 15.0% had two or more siblings. Participants came from a medium-sized city and surroundings in the southwest of Germany, were predominantly Caucasian and from middle-class backgrounds. Paternal educational level was relatively high: 60% of both mothers and fathers held a college or university degree, 21.3% of mothers and 20% of fathers completed secondary school at top track (Abitur), and 18.7% of mothers and 20% of fathers completed secondary school at lower or middle track.

Parents were recruited by telephone from a list of families who had earlier expressed interest in volunteering for research on infant development. At each visit, they received a recompense for travel expenses and infants were given a small gift and a certificate for participating.

2.2 Tasks and materials

2.2.1 Temperament

To assess infant temperament parents completed the German version of the Revised Infant Behavior Questionnaire (IBQ-R, Gartstein & Rothbart, 2003; Kristen, Eisenbeis, Thoermer, & Sodian, 2007) at both age points. Questionnaires were mailed to parents 1-2 weeks prior to the respective test session. At test scheduling, the primary caregiver was advised to fill out the questionnaires and to hand them back to the experimenter at testing. The IBQ-R includes the following 14 scales: Activity Level (15 items), Distress to Limitations (16 items), Fear (16 items), Duration of Orienting (12 items), Smiling and Laughter (10 items), High Pleasure (11 items), Low Pleasure (13 items), Soothability (18 items), Falling Reactivity/Rate of recovery from distress (13 items), Cuddliness (17 items), Perceptual Sensitivity (12 items), Sadness (14 items), Approach (12 items) and Vocal Reactivity (12 items). For each item, parents were asked to rate on a 7-point Likert scale, how frequent a specific infant behavior occurred during the preceding 1-2 weeks (1 = never, 7 = always).

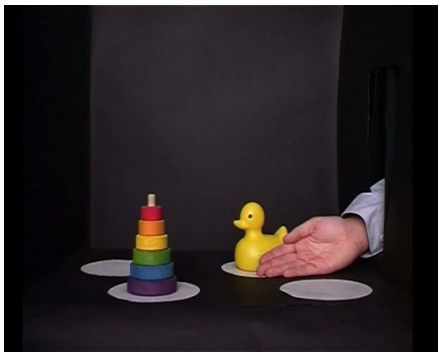
2.2.2 Visual habituation tasks

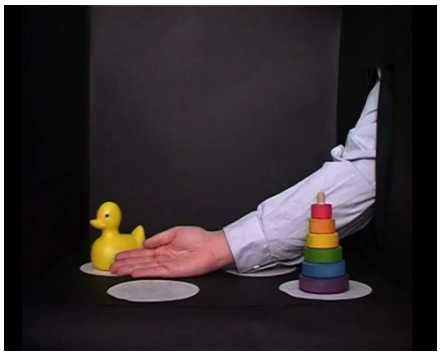
At both age points, infant perception of human goal-directed action (Back-of-Hand task, BoH) as well as their perception of physical causality in a collision event (Causality task, Caus) was assessed employing a visual habituation paradigm. In addition to the two habituation tasks, infant motor and cognitive development was assessed with the Bayley Scales and mother-infant-interaction was recorded during a 5-minute free play (the latter two assessments will not be further discussed in the current study). The cognitive scale of the Bayley Scales and the free play were always administered in between the two habituation tasks such that the two habituation tasks were never presented consecutively. There were two orders of tasks: 44 infants at 6 months and 37 infants at 12 months saw the BoH task first (BoH first), and 36 infants at 6 months and 43 infants at 12 months saw the Caus task first (Caus first). The presentation order of the two habituation tasks was counterbalanced

across age points. Testing sessions lasted between 60-90 minutes, with each single habituation task amounting to about 10 minutes of testing time. The first habituation task was conducted circa 10 minutes after beginning of testing (second task), the second habituation task was conducted circa 45 minutes after beginning of testing (fifth task).

Figure 6: Setup of the visual habituation tasks.

Back-of-Hand task. The same video clips as in Hofer and colleagues' (Hofer et al., 2008) adaption of the Woodward paradigm (1999) were used (see Figure 7). There were four action clips and two reversed-position clips without action. In each action clip, an arm appeared from behind a curtain at the right side of a stage, lowered its hand with the palm facing up onto one of two objects positioned side-by-side on the front part of the stage (duck and tower), then pushed the target object smoothly to the back of the stage and came to a halt. The four action clips differed with regard to the target object chosen (duck

a) 

b) 

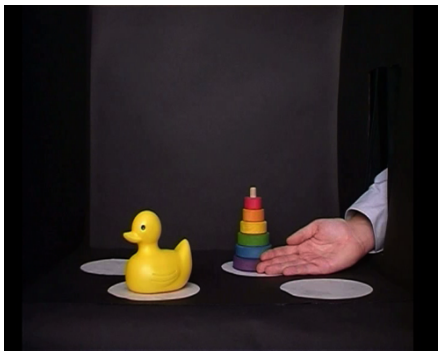
c) 

Figure 7: Back-of-Hand Task.
a) Habituation phase b) Path change test event c) Object change test event.

or tower) and the position of the target object (left or right). In the habituation phase, infants were presented with one of the four action clips (e.g., displacing the duck on the right). The habituation phase was infant-controlled and lasted between a minimum of 4 and a maximum of 14 trials. Each trial ended as soon as the infant looked away for 2 seconds or after 60 seconds had elapsed (also in test phase). The criterion was computed relative to the sum of the preceding three trials and was reached when infants' total looking time in one trial decreased below 50% of the sum of the preceding three trials (i.e., floating-point 50% decrement criterion). After the last habituation trial, infants saw one of the reversed-position clips (object positions reversed compared to habituation, no action) in order to familiarize with the new arrangement. In the test phase, infants saw a path change test event and an object change test event, each for three times in an alternating order. In the *path change* test event (e.g., displacing the duck on the left), the hand's back pushed the same target object as during habituation but due to the change in object positions, the arm took a different movement trajectory. In the *object change* test event (e.g., displacing the tower on the right), the arm performed the same trajectory as during

habituation but contacted and pushed the other object that had previously not been the target. Action clip conditions and order of test trials were counterbalanced across infants.

Causality task. The same collision event videos were used as in Hohenberger, Elsabbagh, Serres, de Schoenen, Karmiloff-Smith, and Aschersleben (2012), which were closely modeled to the looking-time task developed by Kotovsky and Baillargeon (1994, 1998). The stimuli were computer generated color film clips (2D flash animation) showing a ramp and a horizontal track with three stylized houses in the background (see Figure 8). In the habituation clip, a middle-sized blue ball rolled down the ramp, hit a middle-sized red ball, causing it to roll along the track and stop in front of the middle house. In the *possible* test event, a big yellow ball rolled down the ramp, hit the red ball that then rolled along the track to the end of the display. In the *impossible* test event, a small orange ball rolled down the ramp, hit the red ball that then rolled along the track to the end of the display. The possible

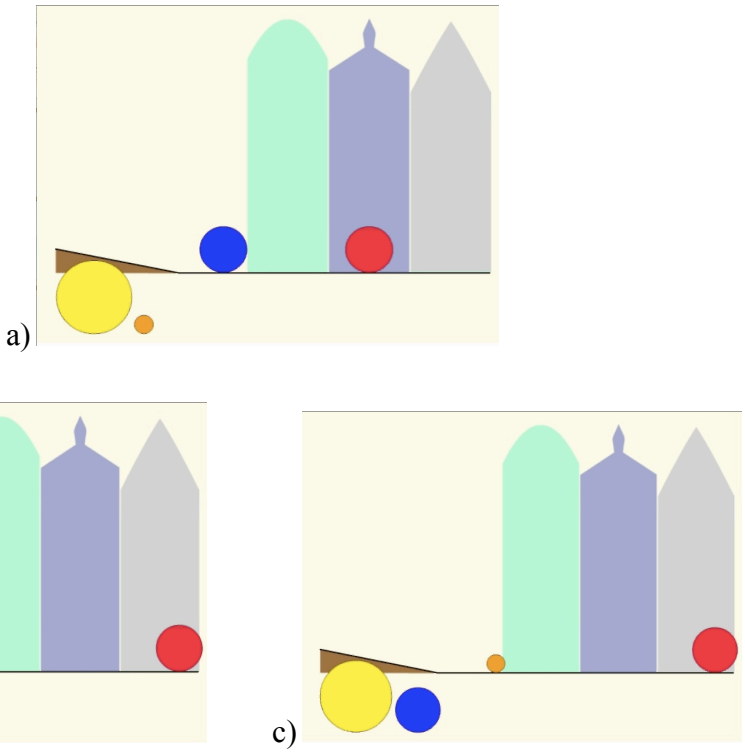


Figure 8: Causality Task.
a) Habituation phase. b) Consistent test event. c) Inconsistent test event.

and the impossible test event were presented each for three times in an alternating order. Order of test trials was counterbalanced across infants. Maximum number of habituation trials, habituation criterion, maximum length of trial, and looking time criteria were the same as in the BoH task. The only procedural difference was the additional reverse position trial in the BoH task.

2.2.3 Procedure

The test room was unfurnished except for the test equipment. Infants sat on their parent's lap in front of a table (60 x 110 cm) facing a 16" Apple computer screen (distance 70 cm). Black curtains were draped around the screen so that only the monitor was visible for the participants. Parents were instructed not to interact with their infants and to look down on infants' heads throughout the presentation of the video action clips. The video presentation and the succession of trials were controlled by the computer program Habit 2000 on a Mac OS 9.2. The whole procedure was controlled by a trained experimenter who observed the infants' looking behavior from behind the curtains. Test sessions were videotaped for later offline coding.

To shorten testing time, number of trials in the habituation phase was reduced to 6 trials in the course of the study. At the same time, number of test trials was increased to 8 trials. Thirty-seven out of 80 infants were tested using this altered procedure at 6 months and 45 infants at 12 months. Groups did not differ in dropout rate ($p > .11$ and $p > .22$ for 6 and 12 months, respectively). As to temperament, groups significantly differed in Cuddliness ($p = .020$) and Approach ($p = .026$) at 6 months of age and in Activity Level ($p = .011$) and Distress to Limitations ($p = .016$) at 12 months of age. However, a Fisher's omnibus test run on the 24 *p*-values indicated that the statistical significance of these four tests likely resulted by chance (Haccou & Meelis, 1994). Hence, data were collapsed for subsequent analyses.

2.3 Scoring

2.3.1 Temperament

According to the scoring procedure of the IBQ-R (Rothbart & Gartstein, 2000), scores for each subscale were averaged by dividing the total by the number of items receiving a numerical response. Parents seemed to have difficulties in responding to the scales Perceptual Sensitivity and Falling Reactivity. For Perceptual Sensitivity, parents responded with "does not apply" to more than 50% of the items for 10 infants at 6 months and 2 infants at 12 months. For Falling Reactivity, parents responded with "does not apply" to more than 50% of the items for 2 infants at 6 months and 2 infants at 12 months. Due to the relatively high number of missing values, the data pertaining to these two scales were excluded from analyses, except for the factor analyses to compare results with Gartstein and Rothbart (2003) and Vonderlin and colleagues (2012).

2.3.2 Visual habituation tasks

Looking times. Two trained coders recoded infant looking times in the habituation and test trials offline from the video recordings. Note that in the BoH task, as durations of reaching to the object differed slightly between test events as a function of position of the object, infant looking times in test trials were counted once the hand had contacted the object. Intra-class correlation revealed an inter-observer reliability for looking durations in the BoH task of $r = .963$ and $r = .997$, and in the Caus task of $r = .956$ and $r = .971$ (all ps $< .001$) at 6 and 12 months of age, respectively. Looking time analyses were not within the scope of this work and are thus not reported in the following.

Dropout. If infants showed excessive fussiness or crying during the habituation task, the task was aborted by the experimenter. These infants were assigned to the group of non-completers in the present study regardless of the abort's moment in experiment (habituation vs. test phase).

3 Results

3.1 Strategy of Data Analysis

The main objective of this study was to investigate the impact of individual differences in various temperamental dimensions on the dropout rate

in two different visual habituation tasks. First, chi^2-tests were conducted to check for differences in dropout rate between boys and girls as well as between the two habituation tasks. Second, temporal stability of dropout rate in the habituation tasks was assessed via McNemar-tests. In order to compare the present data to another German sample (Vonderlin et al., 2012) a factor analysis was carried out on the IBQ-R scales for each age point. Following Carranza and colleagues (Carranza Carnicero et al., 2000), both stability in parental ratings of infant temperament as well as developmental change were assessed. Stability of the IBQ-R scales was reviewed via Pearson product-moment correlations and developmental changes were reviewed via paired *t*-tests or Wilcoxon-tests if data were not normally distributed. To test whether dropout was systematically related to temperament when taken temporal position of the respective task into account, point-biserial correlations were conducted between temperament and dropout separately for each age point and each temporal position of task. Please note that here order of tasks (BoH first, Caus first) was not of interest but correlations between 1) temperament and dropout in the first task administered to the infant, and between 2) temperament and dropout in the second task administered. Finally, four multivariate analyses of variance were conducted, one for each experiment and age point, with mean scores of the 12 IBQ-R scales and the dimensions as dependent variables, and with dropout (yes, no), infant gender (boys, girls), and order of tasks (BoH first, Caus first) as independent factors. In case of significant interaction effects, linear contrasts were conducted to test the significance of the single effects. In addition to significant results ($p < .05$), marginally significant results ($.05 < p < .06$) and trends ($.06 < p < .10$) are reported as well. Effect sizes are reported as *f* with the ranges small: $.10 \leq f < .25$, medium: $.25 \leq f < .40$, and large: $f \geq .40$ (Cohen, 1988).

3.2 Dropout

Dropout in the present sample ranged from 21% in both tasks at 12 months to 42% in the Caus task and 68% in the BoH task at 6 months. Table 9 displays number of infants completing and not completing each of the two visual habituation experiments at 6 and 12 months of age, separated by gender

and order of tasks. There was no difference in dropout rate between boys and girls ($p > .17$). At the age of 6 months, there were more non-completers in the BoH task compared to the Caus task ($\chi^2 = 9.80$, $p = .002$). At the age of 12 months, there were overall more completers than non-completers in both experiments ($\chi^2 = 26.45$, $p < .001$). Two McNemar-Tests, one conducted for each habituation task, revealed no stability of dropout between 6 and 12 months (BoH: $p < .001$, Caus: $p = .006$). For both experiments, there were more infants who were non-completers at 6 months and completers at 12 months of age than vice versa.

Table 9: *Number of completers in the two habituation experiments*

Absolute (and relative) frequencies of infants completing and not completing the experiment for each of the two visual habituation experiments at 6 and 12 months of age, separated by gender and order of tasks (n= 80).

	Completer		**Non-Completer**	
	BoH first	Caus first	BoH first	Caus first
		BoH 6 months		
Male	10 (13%)	5 (6%)	15 (19%)	16 (20%)
Female	7 (9%)	4 (5%)	12 (15%)	11 (14%)
		Caus 6 months		
Male	13 (16%)	14 (18%)	12 (15%)	7 (9%)
Female	9 (11%)	11 (14%)	10 (13%)	4 (5%)
		BoH 12 months		
Male	19 (24%)	16 (20%)	2 (3%)	9 (11%)
Female	15 (19%)	13 (16%)	1 (1%)	5 (6%)
		Caus 12 months		
Male	15 (19%)	20 (25%)	6 (8%)	5 (6%)
Female	15 (19%)	13 (16%)	1 (1%)	5 (6%)

Note. BoH = experiment to test perception of goal-directed human action. Caus = experiment to test perception of physical causality.

3.3 Temperament

Table 10 displays descriptive information regarding parental ratings of infant temperament at 6 and 12 months of age in the current study as well as ratings reported by Gartstein and Rothbart (2003) and Vonderlin and colleagues (2012). Overall, average ratings and standard deviations of the IBQ-R scales were similar to those reported for 6-9 months old US infants (Gartstein & Rothbart, 2003) and for 7-9 months old German infants (Vonderlin et al., 2012). Ratings for High Pleasure and Soothability were lower in the current sample.

A principal axis extraction factor analysis with an oblimin rotation was conducted for each age in order to evaluate the underlying factor structure of the IBQ-R. The two-factor solution derived for the current sample is nearly identical to that of Vonderlin and colleagues (2012, see Table 11). At 6 months, the first factor mainly included the loadings for Activity Level, Sadness, Distress to Limitations, Fear, and negative loadings for Falling Reactivity and Soothability. This first factor may thus be interpreted as Negative Affectivity (Gartstein & Rothbart, 2003). The second factor mainly contained the loadings for Approach, Vocal Reactivity, High Pleasure, Low Pleasure, Smiling and Laughter, Perceptual Sensitivity, Cuddliness and Duration of Orienting. This second factor may thus be interpreted as Surgency/Extraversion (Gartstein & Rothbart, 2003). Similarly, at 12 months, the Negative Affectivity factor included mainly the loadings for Activity Level, Sadness, Distress to Limitations, Fear, and negative loadings for Falling Reactivity, but not for Soothability (different from 6 months of age). The Surgency/Extraversion factor mainly contained the loadings for Approach, Vocal Reactivity, High Pleasure, Low Pleasure, Smiling and Laughter, Perceptual Sensitivity, Cuddliness, Duration of Orienting and Soothability. Similar to Vonderlin and colleagues (2012), and for ratings at both age points, the current analysis did not yield the third factor Self-regulation postulated by Gartstein and Rothbart (2003).

Table 10: *Parental temperament ratings at ages 6 and 12 months*

Means (*M*) and standard deviations (*SD*) of parental temperament ratings at ages 6 and 12 months (*n*=80) and ratings reported by Vonderlin et al. (2012)[1] and Gartstein and Rothbart (2003)[2]. *P*-values indicate significant changes between 6 and 12 months.

		M (SD)		*M (SD)*[1]	*M (SD)*[2]
Scale	Number of items	6 months	12 months	7-9 months	6-9 months
Approach[b)]	12	5.41** (.81)	5.69 (.67)	5.70 (.60)	5.35 (.81)
Vocal Reactivity[b)]	12	3.80*** (.91)	4.61 (.91)	4.45 (.98)	4.67 (.80)
High Pleasure	11	5.48*** (.78)	5.75 (.71)	5.86 (.64)	6.03 (.59)
Smiling & Laughter	10	4.17** (.97)	4.47 (.97)	4.34 (1.0)	4.66 (.88)
Activity Level	15	3.86** (.93)	4.10 (.80)	4.09 (.84)	4.37 (.73)
Perceptual Sensitivity[a)]	12	3.95 (1.17)	4.46 (1.03)	4.40 (1.15)	4.14 (1.05)
Sadness[b)]	14	3.46 (.84)	3.52 (.88)	3.48 (.95)	3.45 (.98)
Distress to Limitations	16	3.41*** (.77)	4.24 (.77)	3.90 (.93)	3.56 (.87)
Fear[b)]	16	2.16*** (.72)	2.83 (.95)	2.57 (1.05)	2.46 (.97)
Falling Reactivity[a)]	13	4.94 (.93)	5.13 (.91)	5.05 (1.01)	5.30 (.77)
Low Pleasure	13	5.07*** (.87)	4.53 (.98)	4.96 (.84)	5.07 (.82)
Cuddliness[b)]	17	5.62*** (.88)	5.15 (.90)	5.43 (.74)	5.72 (.63)
Duration of Orienting[b)]	12	3.31 (1.0)	3.20 (1.02)	3.26 (.99)	3.60 (1.13)
Soothability	18	4.56*** (.94)	3.34 (.52)	5.24 (.69)	5.32 (.75)

Note. **$p < .01$. ***$p < .001$. All *p*-values two-tailed. [a)] Scores for Perceptual Sensitivity and Falling Reactivity are provided for our sample but not included in further analyses due to parental difficulties in responding to the items of these scales. [b)] Scales are not normally distributed why non-parametric comparisons were used.

Table 11: *IBQ-R factor loadings at ages 6 and 12 months*

IBQ-R factor loadings are additionally compared to the loadings obtained by Vonderlin et al. (2012)[1] and Gartstein and Rothbart (2003)[2]

	F1				F2				F3
Scale	*6 m*	*12 m*	*1*	*2*	*6 m*	*12 m*	*1*	*2*	*2*
Approach	.23				.70	.65	[.39]	(.74)	
Vocal Reactivity					.63	.59	[.72]	(.74)	
High Pleasure					.66	.76	[.51]	(.69)	
Smiling & Laughter	.22				.63	.69	[.65]	(.55)	
Activity Level	.46	.39	[.52]			.23	[.31]	(.49)	
Perceptual Sensitivity	.27		[.20]		.63	.51	[.55]	(.45)	
Sadness	.54	.63	[.79]	(.79)					
Distress to Limitations	.82	.81	[.83]	(.69)					
Fear	.19	.29	[.36]	(.31)					
Falling Reactivity	-.40	-.30	[-.51]	(-.56)		.23			
Low Pleasure	-.34		[-.39]	(-.25)	.69	.55	[.54]		(.70)
Cuddliness			[-.36]		.34	.34			(.56)
Duration of Orienting					.47	.39	[.43]		(.43)
Soothability	-.37		[-.21]			.19	[.29]		(.43)

Note. F1 = Negative Affectivity, F2= Extraversion/Surgency, F3 = Selfregulation. Loadings < .20 are not included (except for Fear at 6 months and Soothability at 12 months for our own sample). Loadings obtained by Vonderlin et al. (2012) are listed in square brackets, loadings obtained by Gartstein and Rothbart (2003) in round brackets.

Pearson product-moment correlations were computed to assess the stability of infant temperament ratings (see Table 12). Except for two of the 12 considered IBQ-R scales (Cuddliness and Soothability), results indicated moderate to high normative stability in parental temperament ratings from 6 to 12 months. To assess developmental change, paired *t*-tests and Wilcoxon tests were computed depending on whether the data were normally distributed. Scores significantly increased from 6 to 12 months for Approach, Vocal Reactivity, High Pleasure, Smiling and Laughter, Activity Level, Distress to Limitations, and Fear; scores significantly decreased for Low Pleasure, Cuddliness, and Soothability (all *p*-values < .01). No significant change was observed for Sadness and Duration of Orienting (see Table 10).

Table 12: *Correlations of parental temperament ratings*

Pearson product-moment correlations (two-tailed) of parental ratings at ages 6 and 12 months for 12 IBQ-R scales (n=80).

Scale	*r*	*p (two-tailed)*
Approach	.496	< .001
Vocal Reactivity	.547	< .001
High Pleasure	.609	< .001
Smiling & Laughter	.620	< .001
Activity Level	.598	< .001
Sadness	.566	< .001
Distress to Limitations	.455	< .001
Fear	.388	< .001
Low Pleasure	.609	< .001
Cuddliness	.210	.061
Duration of Orienting	.607	< .001
Soothability	-.070	.539

Note. Scores for Perceptual Sensitivity and Falling Reactivity are not provided for our sample due to parental difficulties in responding to the items of these scales.

3.4 Temperament and Dropout

Point-biserial correlations were performed to assess whether temperament was related to dropout rate. Correlations were performed separately for each age point and for temporal position of task (first or second task administered). There were no relations between temperament and dropout in the first habituation task at 6 months (*p*s > .07). At 12 months, there was a marginally significant correlation showing that non-completers in the first task were rated lower in Duration of Orienting than completers (r_{pb}= -.214, p = .057). In contrast, significant relations were found between temperament and dropout in the second habituation task administered to the infants. At 6 months of age, non-completers in the second task were rated higher in Sadness than completers (r_{pb}= .228, p = .042). At 12 months of age, non-completers in the second task were rated lower in Duration of Orienting than completers (r_{pb} = -.308, p = .005).

To test whether dropout was systematically related to individual differences in infant temperament, following Vonderlin and colleagues (2008), four multivariate analyses of variance were conducted, one for each task and age point, with dropout (yes, no), gender (boys, girls) and order of tasks (BoH first, Caus first) as factors and mean scores of the 12 IBQ-R scales and the 2 dimensions (Surgency/Extraversion, Negative Affectivity) as dependent variables.

BoH task at 6 months. Results revealed gender effects for the IBQ-R scales Distress to Limitations, $F(1,72) = 3.94$, $p = .051$, $f = .23$, and High Pleasure, $F(1,72) = 4.91$, $p = .030$, $f = .26$. Boys ($M = 3.54$, $SE = .12$) were rated as being more distressed in confining situations than girls ($M = 3.23$, $SE = .14$). Also, boys ($M = 5.62$, $SE = .13$) were rated as having more pleasure in games with high stimulus intensity than girls ($M = 5.28$, $SE = .15$). In addition, there was a gender X order of tasks interaction for Activity Level, $F(1,72) = 5.20$, $p = .026$, $f = .27$, and for Negative Affectivity, $F(1,72) = 4.47$, $p = .038$, $f = .25$. Boys in the group BoH first ($M = 3.63$, $SE = .18$) were rated as less active than boys in the group Caus first ($M = 4.37$, $SE = .21$), $p = .006$. Girls in the group BoH first ($M = 3.38$, $SE = .12$) were rated as having more Negative

Affectivity than girls in the group Caus first (M = 3.03, SE = .15), p = .046. Finally, there was a dropout X order of tasks interaction for Smiling and Laughter, $F(1,72)$ = 5.87, p = .018, f = .28. In the group BoH first, non-completers (M = 4.40, SE = .18) were rated as showing more Smiling and Laughter than completers (M = 3.80, SE = .24), p = .042.

Caus task at 6 months. Results revealed a gender effect for the IBQ-R scale High Pleasure, $F(1,72)$ = 4.18, p = .045, f = .24. Boys (M = 5.62, SE = .12) were rated as having more pleasure in games with high stimulus intensity than girls (M = 5.28, SE = .15). In addition, for Activity Level, there was a gender X order of tasks interaction, $F(1,72)$ = 4.29, p = .042, f = .24, as well as a gender X dropout X order of tasks interaction, $F(1,72)$ = 3.92, p = .052, f = .23. Boys in the group BoH first (M = 3.63, SE = .18) were rated as less active than boys in the group Caus first (M = 4.37, SE = .20), p = .006, and female non-completers in the group BoH first (M = 3.42, SE = .28) were rated as less active than female completers (M = 4.23, SE = .29), p = .048. Also, there was a gender X order of tasks interaction for Negative Affectivity, $F(1,72)$ = 3.69, p = .059, f = .23. Girls in the group BoH first (M = 3.38, SE = .12) received higher scores in Negative Affectivity than girls in the group Caus first (M = 3.03, SE = .15), p = .046. Furthermore, there was a gender X dropout interaction for Fear, $F(1,72)$ = 3.13, p = .081, f = .21, and for Soothability, $F(1,72)$ = 3.70, p = .058, f = .23. Male non-completers (M = 2.36, SE = .17) were rated as more fearful than male completers (M = 1.94, SE = .13), p = .058, and female non-completers (M = 4.12, SE = .28) were rated as being harder to soothe than female completers (M = 4.69, SE = .21), p = .077.

BoH task at 12 months. Dropout effects were found for Sadness, $F(1,72)$ = 5.86, p = .018, f = .28, Negative Affectivity, $F(1,72)$ = 3.05, p = .085, f = .21, and Duration of Orienting, $F(1,72)$ = 2.93, p = .091, f = .20. Completers (M = 3.60, SE = .10) were rated as being more sad than non-completers (M = 3.21, SE = .28), and completers (M = 3.70, SE = .07) received higher scores in Negative Affectivity than non-completers (M = 3.36, SE = .18). Additionally, completers (M = 3.35, SE = .13) received higher scores in Duration of Orienting than non-completers (M = 2.63, SE = .33). There was a

main effect of order of tasks for Approach, $F(1,72) = 3.11$, $p = .082$, $f = .20$, and a dropout X order of tasks interaction for Approach, $F(1,72) = 6.98$, $p = .010$, $f = .31$, High Pleasure, $F(1,72) = 3.30$, $p = .073$, $f = .21$, and Negative Affectivity, $F(1,72) = 2.94$, $p = .091$, $f = .20$. In the group Caus first, non-completers ($M = 6.02$, $SE = .18$) were rated as showing more approach than completers ($M = 5.66$, $SE = .12$), $p = .031$, and non-completers ($M = 6.02$, $SE = .20$) were rated as having more pleasure in games with high stimulus intensity than completers ($M = 5.65$, $SE = .13$), $p = .039$. In the group BoH first, completers ($M = 3.83$, $SE = .10$) received higher scores in Negative Affectivity than non-completers ($M = 3.07$, $SE = .34$). This last result might be due to chance because the non-completer group consisted of only three infants and will thus not be discussed further.

Caus task at 12 months. There was a gender effect for Cuddliness, $F(1,72) = 3.0$, $p = .090$, $f = .20$. Girls ($M = 5.34$, $SE = .26$) were rated as being more cuddly than boys ($M = 5.01$, $SE = .16$). In addition, there was a gender X dropout interaction for Duration of Orienting, $F(1,72) = 2.79$, $p = .099$, $f = .20$. Male completers ($M = 3.49$, $SE = .17$) received higher scores in Duration of Orienting than male non-completers ($M = 2.60$, $SE = .31$), $p = .003$.

As order of tasks depends on study design and gender is not random, interaction effects including gender and order of task are not discussed further in the following section.

4 Discussion

The aim of the present study was to longitudinally examine the impact of infant temperament on the dropout rate in visual habituation experiments in infants at 6 and 12 months of age. For both habituation tasks and both age groups, there were no differences between groups of completers and non-completers for the majority of the 12 considered IBQ-R subscale ratings. Only few subscales were found to have an influence on dropout in the current sample (22 out of 384 possible effects were significant at $p < .10$). Our specific hypotheses were only confirmed for Duration of Orienting in that a higher score in this subscale related to completion of the task. Overall, the present

data are in line with the findings of the newer studies (Slaughter & Suddendorf, 2007; Vonderlin et al., 2008) suggesting that only few temperamental factors influence the completion of visual habituation experiments in infancy. In addition, this is, to our knowledge, the second study that evaluated the factor structure of the IBQ-R in a German sample. Consistent with Vonderlin and colleagues (2012), a two-factor solution was found in the current sample with Surgency/Extraversion and Negative Affectivity as underlying temperament dimension both at 6 and 12 months of age. Together with previous findings showing a three-factor solution in other countries (e.g., Casalin et al., 2012, Gartstein & Rothbart, 2003), these results point to potential cultural differences in the structure of infant temperament.

Dropout rates: infant age and task type. Compared to other studies using visual habituation or violation-of-expectation techniques (Slaughter & Suddendorf, 2007), the dropout rate in the current sample was in part quite high: 68% in the BoH task and 42% in the Caus task at 6 months, and 21% in both tasks at 12 months. These dropout rates are similar to those reported by Hohenberger and colleagues (2012) who used the same two habituation tasks also within a larger battery of tasks at 6 and 10 months of age. Possibly, methodological differences in paradigms employed (e.g., habituation, familiarization, violation-of-expectation, paired-comparison techniques) might explain differences in dropout rate in infant looking time studies. In the current work, length of trial and length of habituation phase were infant-controlled, thus length of test sessions varied with infant interest. However, we spared fancy attention getters at the beginning of each trial, which might account in part for higher dropout rates. Number of trials and looking time criteria were held constant across the two habituation tasks but dropout rate differed between tasks at 6 months. This difference together with similar dropout rates in Hohenberger and colleagues (2012) suggest that the high dropout rate especially in the BoH task at 6 months might be due to other factors like infant age and type of stimuli rather than the specific procedure used. Dropout was less frequent at the age of 12 months compared to 6 months. One reason why infants at 6 months are more likely to dropout in the habituation tasks might be that their self-regulation is still developing compared to 12 months of age

(Zentner & Bates, 2008). At the younger age, infants might have more difficulties to stay calm and to cope with the task demands of the habituation tasks, bringing contextual factors into question. In fact, although the two visual habituation tasks shared the same procedure and setup, results suggest that the BoH task was more difficult to complete than the Caus task at 6 months, but not at 12 months. This apparent difference in task difficulty at 6 months may have been related to surface properties of the video stimuli and/or to the understanding infants had of the content of the events presented. In the Caus task, stimuli consisted of a 2D scene with brightly colored simple shapes with clear contours that were presented against a white background. Some of the shapes (spheres) rolled across the screen. In contrast, the BoH stimuli consisted in a 3D scene, in which a human hand (and white-sleeved arm) moved towards one of two brightly colored toys (a duck and a tower) on a stage and pushed this toy to the back of the stage. The stage and the background were dark in color. In terms of surface features, the Caus stimuli were less complex and more colorful than the BoH stimuli. It might therefore be that at 6 months, infants were overall more attracted to the Caus stimuli compared to the BoH stimuli. As to conceptual content, previous research showed that infants perceive the Back-of-Hand action as directed towards a goal by 6 months (e.g., Jovanovic et al., 2007), whereas infants understand the violation of the physical principle in the Caus task only by 10 months of age (Hohenberger et al., 2012). It is therefore unlikely that dropout rates at 6 months were related to a differential understanding of the conceptual content of the stimulus events, given that in this case, the BoH task should have been less difficult than the Caus task. Please note that in both tasks, duration of trials were infant-controlled.

Underlying temperament dimensions. The average scores of the IBQ-R scales were overall comparable to those reported by Gartstein and Rothbart (2003) and Vonderlin and colleagues (2012). However, findings regarding underlying dimensions as revealed by the factor structure are only in part consistent with previous work. Whereas a three-factor solution was found for US American (Gartstein & Rothbart, 2003), Polish (Dragan, Kmita, & Fronczyk, 2011), Russian (Gartstein, Knyazev, & Slobodskaya, 2005) and

Italian (Montirosso, Cozzi, Putnam, Gartstein, & Borgatti, 2011) infants, the current work confirmed the two-factor solution found by Vonderlin and colleagues (2012) with Surgency/Extraversion and Negative Affectivity as the only two underlying dimensions (please see Nakagawa & Sukigara (2005) for a comparable two-factor solution in a Japanese sample). Vonderlin and colleagues (2012) suggested that development of self-regulation with respect to age ranges assessed in the different studies might explain the differences in factor structure between their German sample and the above-cited work. Whereas the studies finding a three-factor solution based their analyses on data comprising the first 12 months of age, the age ranges assessed in the two German samples were rather narrow. Since self-regulatory components begin to develop later in the second half of the first year of life, it might be possible that they cannot be assessed separately from affectivity at a younger age (Putnam, Ellis, & Rothbart, 2001). Thus, a two-factor solution would be predicted in younger infants and a three-factor solution in older infants, or in samples comprising also a substantial group of older infants. However, the findings of the current study are inconsistent with this prediction, as analyses yielded nearly the same two-factor solution both when infants were 6 as well as 12 months of age. Still, the scales contained in the IBQ-R were designed to cover a wide age range (3-12 months), which might also account for differences in factor structure between the two German samples and the above-cited work. This explanation is supported by a change between age points in the frequency of mother's non-applicability ratings that also differs between single scales. For example, number of caregivers rating more than 50% of the items as non-applicable changed for Perceptual Sensitivity from 10 caregivers at 6 months to 2 caregivers at 12 months, whereas there was no change in frequency for Falling Reactivity (2 caregivers at each age point). It is therefore still an open question for future work to address, whether a three-factor solution would result from ratings of a German sample that includes infants ranging from 3 to 12 months of age.

A second explanation put forward by Vonderlin and colleagues (2012) regards the considerable correlations found (Putnam et al., 2001) between Orienting/Regulation and the two other dimensions, Surgency/Extraversion and

Negative Affectivity, that question the assumption of Orienting/Regulation as a third independent dimension (please see also Evans & Rothbart, 2009, for arguments in favor of a higher-order two-factor model). Finally, cultural differences may explain the differences found in factor structure. It might be that the everyday situations described in the IBQ-R differ in their occurrence between countries. For example, parents in the present sample seemed to have difficulties in responding to the two subscales Perceptual Sensitivity and Falling Reactivity. It might be that these subscales contain items or situations that are more common in daily life for US Americans than for Germans. It may also be that the German translation did not perfectly correspond to the specific situation referred to in some of the items. Thus differences in factor structure may reflect also differences in parents' understanding of the items. Related to this, parents' interpretation of infant behavior as indicating a specific temperament characteristic might also differ between cultures. Nakagawa and Sukigara (2005) showed that Japanese mothers did not correctly assign the single IBQ-R items to the 14 subscales, failing especially for the subscales Activity Level and Distress to Limitations. It is therefore favorable that the given items in a temperamental questionnaire validly indicate the underlying temperament characteristics across different cultures (Gartstein et al., 2006). Finally, cultural differences in parenting might account for differences in factor structure. First, despite a genetic component of temperament, child rearing fosters temperament characteristics that are in line with the respective cultural values (Kohnstamm, 1989; Gartstein et al., 2006). Second, parents potentially rate their infants' temperament characteristics by referring to culture-specific standards on, e.g., desirability and normativity (Slobodskaya, Gartstein, Nakagawa, & Putnam, 2012). In line with this argument, Vonderlin and colleagues (2012) suggested a revision or even an extinction of some item translations due to low discrimination coefficients.

A limitation of the present study is that temperament was only assessed via parental report. Research suggests that parental report might be biased by parental depression and anxiety as well as by parental expectations formed during pregnancy (Leerkes & Crockenberg, 2003; Wolk, Zeanah, Garcia-Coll & Carr, 1992). Also, while some studies assessing the agreement between

parental report on temperament and observational data collected in laboratory assessments (Parade & Lerkes, 2008) or home observations (Stifter, Willoughby, Towe-Goodman & The Family Life Project Key Investigators, 2008) yielded support for the validity of the IBQ-R, other studies found little concordance between parental report and observational measures (e.g., Seifer, Sameroff, Barrett & Krafchuk, 1994). However, moderate agreement is expected when raters differ in interactional contexts and size of reference population (e.g., Funder & West, 1993). Parents have the opportunity to observe their children in a variety of different situations for a long period of time, whereas an observation at home or in the lab only allows a rather brief glimpse on behavior indicating temperament (Stifter et al., 2008). Despite the limitations of parental report measures, it is important for the current work that a study by Pauli-Pott, Mertesacker and Beckmann (2005) reported a satisfying convergence between parental ratings in the German version of the IBQ and observational data.

Dropout rates: longitudinal analyses. To our knowledge, this is the first study to longitudinally assess long-term stability in dropout rates by employing the same two tasks across age points. In the current study, there was no evidence for longitudinal stability in dropout rates between the 6 and 12 months visit. These findings are inconsistent with Bell and Slater (2002) who reported a long-term stability of dropout rate in two different tasks. Infants who did not complete a habituation task at 4 months were also those who did not complete a problem-solving task at 13 months. The authors assumed "that there might be something relatively stable in the infant […] *across* certain environments, and across time, underlying these findings" (p. 157). However, instead of some stable infant characteristic, stability in task demands may also explain this stability in dropout. Although the two tasks employed by Bell and Slater (2002) differed in content, it may be that they posed comparable demands on infants across testing sessions exactly because the type of task was adjusted to infants' interests at the respective point in development. In the current study, differences in dropout rates suggest that the same tasks may have indeed posed different demands on infants at the different age points. Still, the relatively high normative stability in parental temperament ratings supports the

interpretation that dropout in the current study was due to age-related differences in perceived task characteristics or other contextual factors rather than infant temperament.

Temperament: longitudinal analyses. Parental ratings of infant temperament showed both stability and changes across the second half of the first year. Correlation coefficients for 11 out of the 12 temperament subscales considered ranged between .21 and .62, indicating moderate to high normative stability in infant temperament between 6 and 12 months of age. Soothability was the only subscale where no stability in parental ratings was found. At the end of the first year, the emergence of stranger anxiety and the consolidation of attachment relationships may account for this lack in stability. By this time, differences in attachment style also account for differences in soothability. In fact, soothability constitutes a crucial criterium for attachment classification and varies between caretakers as a function of attachment relationship (e.g., van den Boom, 2004). Furthermore, developmental change in infant self-regulation may also explain the lack of stability in Soothability. One may speculate that with better self-regulatory abilities, infants in the current sample mostly needed their parents in very disturbing situations when they were unable to self-regulate and consequently appeared harder to soothe. This explanation would be consistent with the finding that as a group parents rated infants as less soothable at 12 compared to 6 months of age.

In addition to the reported moderate to high normative stability, there were significant changes in average ratings for 10 out of the 12 (14) subscales suggesting that the temperament domains considered here are subject to developmental change between 6 and 12 months of age. There was a significant increase in the ratings for Approach, Vocal Reactivity, High Pleasure, Smiling and Laughter, Activity Level, Distress to Limitations, and Fear, whereas the ratings for Low Pleasure, Cuddliness, and Soothability showed a significant decrease. No change was observed in the ratings for Sadness and Duration of Orienting. Carranza Carnicero and colleagues (2000) reported similar developmental changes in the first year of life in a study employing the five subscales of the original IBQ (Rothbart, 1981). Between 3

and 12 months, they observed an increase in Activity Level, Distress to Limitations, Fear and Smiling and Laughter and no changes in Duration of Orienting.

Influence of temperament on dropout rate. Results of the multivariate analyses suggest that dropout rate seemed not to be systematically influenced by differences in infants' temperament. Analyses yielded only few significant main and interactions effects that were rather unsystematically scattered across the IBQ-R scales. Furthermore, these relations were not consistent across age points and tasks. Most support was found for an influence of Duration of Orienting on dropout rate. At 12 months of age, completers in the BoH task as well as male completers in the Caus task received higher scores for Duration of Orienting. These results are consistent with Fagen et al.'s (1987) work showing that at 9 months, females who did not complete an operant conditioning task due to crying scored lower on Duration of Orienting than non-criers. These findings suggest that Duration of Orienting may be a crucial temperament dimension, influencing dropout in visual habituation tasks, especially at the end of the first year of life. The other main effect of dropout regarded Sadness. At 12 months of age, infants that completed the BoH task were rated as lower in mood than non-completers. A tentative explanation might be that parents evaluated not only crying and unresponsivity as sadness, but also passiveness and quietness and therefore slightly overestimated the sadness score of the completers.

In addition to differential effects of task demands across age and tasks, the time of administration of the specific task within the testing session seems to be a further contextual factor influencing likelihood of dropout. In the present work, dropout was related to both 1) temporal position of the task within the test session (about 10 or 45 min after the beginning) and 2) order of tasks, that is, to temporal position in combination with type of task (BoH first, Caus first). Importantly, only in the second habituation task administered to infants, significant relations between infant temperament and dropout were found. At 6 months of age, dropout in the second habituation task was related to higher rating scores in Sadness. Infants with a generally lower mood related

to personal suffering and/or inability to perform a desired action seemed to have more difficulties to complete the second habituation task. Infants with these temperament characteristics might have experienced the second task as more demanding not only because of fatigue but also because the other tasks might have been more interesting to the infant (e.g., playing with mom, presence of toys) such that they would have liked to return to these activities. At 12 months of age, dropout in the second habituation task administered to infants was related to lower rating scores in Duration of Orienting. Infants who were rated as more able to attend to a single object for an extended period of time seemed to have fewer difficulties to complete a habituation task when the task was administered after a series of other potentially fatiguing tasks. As to order of tasks, individual differences in Approach and High Pleasure were related to dropout at 12 months of age in the BoH task, but only if the BoH task was presented second. Non-completers were rated as showing more rapid approach and positive anticipation of pleasurable activities as well as having more pleasure in games with high stimulus intensity than completers. Possibly, in the group of non-completers there were mostly infants with a higher arousal level who get easily excited and who were therefore more rapidly frustrated when confronted with a boring task. Interestingly, the dimension of High Pleasure is related to the construct of sensation seeking (Putnam et al., 2001; Zuckerman, 1990). It is possible that these non-completers might have been bored by the task, given that they had been already administered one of a similar kind. Also, a crucial characteristic of habituation tasks is to repeatedly present infants with the same stimulus until they loose interest.

Infant gender. Prior research suggests that a potential impact of infant temperament on dropout may vary as a function of gender (Fagen et al., 1987; Wachs & Smitherman, 1985). Infant gender was therefore included as control variable in each multivariate analysis of variance. Gender main effects were revealed for three subscales. At the age of 6 months, boys were rated as having more pleasure in games with high stimulus intensity than girls. This effect was also observed by Gartstein and Rothbart (2003) and reported in a meta-analysis by Else-Quest and colleagues (Else-Quest, Shibley Hyde, Hill Goldsmith & Van Hulle, 2006). Additionally, 6 months old boys were rated as being more

fussy and distressed in a confining situation. This result is consistent with Gunnar and colleagues (Gunnar, Porter, Wolf, Rigatuso & Larson, 1995) who also reported higher parental ratings of distress to limitations for boys compared to girls at 6 months. There is evidence that already as neonates, girls show a greater degree of cuddliness than boys (Benenson, Philippoussis & Leeb, 1999). In the current study, this difference was only found at 12 months but not at 6 months of age. As to interaction effects of infant gender and dropout, Fear and Soothability show relations to dropout rate dependent on gender in the Caus task at 6 months of age. On the one hand, male non-completers were rated as more fearful and more easily startled by novel situations than male completers. Since the whole testing procedure was a novel, unfamiliar situation these boys might have been too aroused to complete the habituation task. On the other hand, female completers were rated as more easily soothed by a parent than female non-completers. Since infants were sitting on their parent's lap during the habituation task, it might be that easily soothed females benefitted from this closeness. Alternatively, this result might be explained by the fact that these girls were also better in self-regulation and were therefore more likely to complete the habituation task. Finally, male non-completers were rated lower in Duration of Orienting at 12 months in the Caus task compared to male completers. Though direction of effects are consistent with the idea that infants with a difficult temperament are more likely to be a dropout, these few interaction effects do not suggest a systematic relationship between gender, temperament and likelihood of dropout. It also remains an open question why these gender differences were specific to dropout in the Caus task.

In sum, the present study supports the assumption that dropout in infant visual habituation studies is not systematically related to a different temperament between infants who complete the task under investigation and infants who show heightened distress such that testing ends before completion of the task. This suggests that the relatively high dropout rates reported in infant looking time studies are not systematically related to infant temperament. However, findings also suggest that temperament might have an impact on likelihood of dropout in the presence of specific contextual factors

such as task characteristics (e.g., attractiveness of stimuli) and time of task administration within a test session. Differences in temperament seem to have an impact on infants' performance especially in more demanding and stressful situations, for example, when a habituation task is conducted at the end of a longer test session. It is therefore recommendable to conduct visual habituation tasks rather at the beginning than at the end of a test session, and to only conduct one task per test session.

Chapter 3: General Discussion

From early infancy, children are attentive interaction partners that gain increasing insight into other people's minds with advancing age. It is well investigated that around the age of 4 years children develop a so called Theory of Mind that enables them to understand the causal relation between people's mental states and actions and consequently to correctly predict their actions (Wellman et al., 2001). Around two decades ago, earlier social-cognitive development has gained attention. Research has shown that already infants and toddlers pay regard to the subjectivity and directedness of mental states while interpreting human behavior (e.g. Meltzoff, 1995; Woodward, 1998, 1999). Theorists have argued that some of these abilities are indicators of children's understanding of intentionality (Tomasello, 1999; Woodward, 1998). As to early understanding of goal-directed action, children develop an understanding of the relation between an actor's intention and the manipulated object. Joint attention skills implicate the understanding that a person's actions are pursuing a certain goal. Beyond, Perner (1991) has associated further early social-cognitive abilities to later ToM. He argued that pretend play and mirror self-recognition both rely on a capacity for secondary representation. Therefore, they might also provide the foundation for the later meta-representation and the ability to attribute mental states to self and others. If these abilities are related in fact, we would indeed expect longitudinal relations from the single early abilities to later ToM as well as interrelations between these early social-cognitive abilities.

Subsequently, researchers tried to reveal a relation between some of the early social-cognitive abilities and later ToM and found a relation for early action-understanding (Aschersleben et al., 2008; Wellman et al., 2004, 2008; Yamaguchi et al., 2009), joint attention (Charman et al., 2000; Sodian & Kristen-Antonow, 2015), intention-based imitation (Colonnesi et al., 2008; Olineck & Poulin-Dubois, 2005), and pretend play (Youngblade & Dunn, 1995). These results support the assumption of continuity in social-cognitive development (e.g. Aschersleben et al., 2008; Wellman et al., 2008) but only

few studies have investigated the interrelations between these early social-cognitive abilities so far. Investigating how the single abilities were also longitudinally interrelated is important to further clarify the supposed continuity, as interrelations would suggest that these abilities are part of the same ability. It could provide some evidence for Perner's (1991) assumption that the social-cognitive abilities rely on the maturation of secondary representation (Nielsen & Dissanayake, 2004). Yet, it is still an open question if these mentioned social-cognitive abilities are early manifestations of a developing ToM or if these abilities are first required to enable one to develop a ToM (Henning, Daum, & Aschersleben, 2009).

Partial support for the continuity hypothesis concerning the early social-cognitive abilities and the assumption that the social-cognitive abilities establish consecutively is provided by the following findings: The understanding of other's intentions in intention-based imitation tasks was found to be related to earlier infant's understanding of intentional actions (Olineck & Poulin-Dubois, 2009), to infant's declarative pointing (Camaioni et al. 2004; Colonnesi et al., 2008; Kristen et al., 2011) and to infant's performing in goal detection tasks (Charman et al., 2000). Additionally, children's mirror self-recognition related to infant's declarative pointing (Sodian & Kristen-Antonow, 2015) as well as to pretend play (Lewis & Ramsay, 2004). Nevertheless, a general relation between the single abilities could not be found (e.g. Nielsen & Dissanayake, 2004). That is why additional studies providing further evidence for continuity in social-cognitive development are needed. To my knowledge, study 1 was the first longitudinal study that assessed cross-sectional and longitudinal relations between five early social-cognitive abilities in the same group of children. Based on the number of investigated abilities, our results revealed only few interrelations but these were in consistence with previous findings. Mainly, a relation between intention-based imitation and two joint attention abilities, namely declarative pointing and detection of experimenter's goal while teasing, were found as well as a relation between pretend play and mirror self-recognition. Partially, in some aspects, our results support previous research (Camaioni et al., 2004; Charman et al., 2000; Kristen et al., 2011; Lewis & Ramsay, 2004) and they complement findings in support

of continuity in social-cognitive development. Moreover, our results make clear that this continuity in social-cognitive development is not global but rather task- and age-specific. It is still an open question whether social cognition develops within one domain or whether there exist different domains of social cognition. It might also be, as Charman and colleagues (Charman et al., 2000) postulated, that these abilities are possibly part of an underlying "social cognitive representational ability to understand and interact with people […] and objects" (p. 492) that emerges within a specific timeframe, after which abilities develop along diverging pathways. Therefore, as I mentioned before, future studies should necessarily assess the respective abilities not only with a single task at one age but with a variety of complementing tasks at different ages in continuous, shorter intervals to clarify the supposed continuity. Another example that otherwise relations could not be discovered by only assessing single abilities might be the study by Nielsen and Dissanayake (2004). They longitudinally examined imitation, pretend play and mirror self-recognition at three-monthly intervals between the ages of 12 and 24 months, but only conducted one task per ability. For example, they tested pretend play in a single scripted drinking task with dichotomous scoring instead of a free play episode. In contrast to other studies using the latter (Lewis & Ramsay, 2004; study 1), relations between pretend play and mirror self-recognition could not been found using this procedure (Nielsen & Dissanayake, 2004). This might indicate that detecting (age- and task-) specific cross-sectional and longitudinal associations between the early social-cognitive abilities requires a more continuous assessment rather than conducting unique tasks. Another explanation, at least for the few significant relations between the abilities emerging at 18 months of age, might be that these abilities are not expressed simultaneously (Nielsen & Dissanayake, 2004). As intention-based imitation, mirror self-recognition and pretend play all demand to represent two different representations of the same object or situation (Perner, 1991), they are all supposed to rely on secondary representation. Yet, children might have to separately learn how to apply their knowledge in the different abilities. The same could be assumed for the different joint attention abilities emerging around the end of the first year of life. Mundy and colleagues (Mundy et al., 2007) reported, for example, different development patterns for initiating and

responding to joint attention and behavior request. Consequently they questioned a common underlying concept of social cognition. This view would speak against a domain-general change in social-cognitive development in infancy (see Nielsen & Dissanayake, 2004). It seems more likely to assume that children's ToM abilities subsequently develop from infancy on. Further long-term longitudinal studies are necessary to adequately investigate the continuity hypothesis and the social-cognitive development before the emergence of ToM in preschool age with a fully developed false belief understanding.

Also study 2 aimed to extend previous findings, focusing on the influence of temperament on social-cognitive development. Wellman and colleagues (Lane et al., 2013; Wellman et al., 2011) have originally advanced this idea by suggesting a developmental link between childhood temperament and ToM abilities that is specific to a less reactive, more observant temperament. To explain how child temperament might influence ToM development they refer to the emotional reactivity hypothesis (Hare, 2007) that originally accounts for social-cognitive capacities in dogs vs. wild canines and chimpanzees. It holds that, as a result of selecting dogs for domestication based on their nonaggressive and non-fearful temperament towards humans, development of human-like social-communicative skills in these animals was supported. In convergent evolution with humans, level of emotional reactivity has modulated domesticated dogs' social-cognitive performance. Wellman and colleagues (Wellman et al., 2011) assumed that this hypothesis is transferable on human development. Hence, an initial difference in child temperament may lead to differences in interactive behavior and social experiences, which in turn may foster or interfere with the development of mental understanding. Yet, supporting evidence is provided by two studies showing a relation between shy, nonaggressive and perceptually sensitive behavior and false belief understanding in preschoolers (Wellman et al., 2011), even across different cultures (Lane et al., 2013). Study 2 aimed to investigate this relation in toddlers. In fact, children with shy temperament at 18 months as well as at 3 years of age showed better ToM abilities at 3 years of age. Without actively participating, shy children may still gain insights in the processes of social

interactions. Thus, our results support the assumption that shyness fosters the ability to reason about others' mental states, actually already before the emergence of false belief understanding. Our findings contribute to the understanding of ToM development in early childhood. From infancy on, one might predict the ToM development of children based on their temperament. Still, little is known about the relation between ToM and temperament. It would be interesting to know if certain temperament factors reflecting social-emotional reactivity also influences early social-cognitive abilities already in infancy. Future research should investigate the relations between temperament and the early social-cognitive abilities we have assessed in study 1.

In addition to examining the influence of temperament on children's early ToM abilities, we analyzed the influence of language, siblings, and parental education in study 2. Contrary to temperament, the influence of these three variables on ToM is well investigated. Yet, findings are controversial. Children's language skills are often discussed in the literature as an important predictor of ToM development. Whereas most previous research pointed to a beneficial effect of language on ToM (e.g., Milligan et al., 2007), our results showed no evidence that language abilities at the ages of 18 months or 3 years influence ToM development although we assessed different language skills. However, our finding is in line with other studies analyzing language when studying ToM (e.g., Aschersleben et al., 2008; Wellman et al., 2004, 2008) who also did not find such a relation. There exist various reasonable hypotheses about how language affects ToM development. Although almost every aspect of language seems to have an essential influence on ToM development, it might be that certain language abilities are particularly more important for different aspects of ToM (see Cutting & Dunn, 1999). On the one hand, perhaps, potential relations between language and ToM abilities could have been detected using another instrument for testing children's language skills. On the other hand, potential relations between language and ToM abilities might be detected when in fact investigating false belief understanding. In contrast to the 4-year-olds tested within the broader project, we only conducted those ToM tasks that demand less language skills than more complex false-belief tasks with the 3-year-olds. Maybe a relation between

language and ToM is therefore not to be found in study 2. Nevertheless, further (longitudinal) investigation is needed to clarify the general assumption that ToM is influenced by language, as evidence concerning this relation is somewhat inconsistent.

In addition, our results are inconsistent with previous findings reporting a positive impact of siblings on ToM abilities (e.g., Cassidy et al., 2005; McAlister & Peterson, 2006, 2007, 2013). Although interactions with siblings seem to be relevant for ToM development, in that siblings share closer interests and feelings with the child as parents do (Dunn et al., 1991), neither number of siblings nor presence of one or more older siblings had a positive effect on ToM in study 2. However, our finding is in line with other studies that did not find a sibling effect when studying ToM (e.g., Cutting & Dunn, 1999; Henning et al., 2011). As mentioned earlier, it might be that children have to reach a certain age before siblings can have a positive influence regarding the understanding of mental states. Ruffman and colleagues (Ruffman et al., 1998) specified the age of 39 months as critical to benefit from siblings in this area. This assumption might be verified in a further analysis including the data of the 4-years-old's ToM abilities that were assessed in the broader project, too. Alternatively, following the assumption by Cutting and Dunn (1999) it is also likely to assume that not the number or age of siblings per se, but the quality of interaction and relationships with siblings is relevant for children's ToM development (Cutting & Dunn, 1999).

Finally, results for parental education in study 2 are inconsistent with previous findings, in that an impact on ToM abilities was reported (e.g., Cutting & Dunn, 1999). Neither maternal nor paternal education had a positive impact on ToM. However, despite the expectation that children of families with low socio-economic status lag behind children of families with middle socio-economic status in ToM development (e.g., Cutting & Dunn, 1999) other studies analyzing parental socio-economic status when studying ToM (e.g., Lucariello, Durand, & Yarnell, 2007) did not find such a relation. We report on this finding, as parental education is one of the markers for socio-economic status of families (Bornstein, Hahn, Suwalsky, & Haynes, 2003). One reason

might be that the reported gap between children with different socio-economic status is restricted to certain ToM tasks, especially false-belief tasks (Lucariello et al., 2007). In study 2, we neither investigated false belief reasoning in the 3-year-olds. Furthermore, it seems important which measure is selected to indicate SES (see Bornstein et al., 2003). Beyond, our result that parental education had no impact on ToM might be result of the parents in our sample being predominantly highly educated, except for some early school dropouts. One can speculate that well-educated parents are more willing to participate with their children in psychological studies taking place at Saarland University that is also situated out-of-town and therefore variance might be restricted in questions regarding parental factors. Besides this phenomenon, one can question what mechanisms are initiated by high parental education or socio-economic status that in turn might affect children's ToM development (Cutting & Dunn, 1999). It might be the way in which parents talk or interact with their children or the activities they take part in. Possibly, social interactions and activities in Saarland families might not crucially differ depending on high or lower educational level so that differences in ToM development were not detectable in the tested children. Even if a lot of parents in our sample were academics, they predominantly come from homes with working class background in mining and metallurgy. Thus, our sample might be a more homogenous group of people independent of different education. As for the impact of siblings on ToM, the influence of parental education might not be that crucial at the beginning of the third year of age and only begins to matter during preschool age. Still, these findings are also consistent with previous research that lack to find a relation between maternal education and ToM (e.g., Farhadian et al., 2011; Ruffman et al., 1999). In sum, even if language, siblings, and parental education all are well-investigated factors, their influence on ToM development is not finally resolved, yet. There exist still some outstanding aspects that need to be considered and detect in future research.

Study 3 investigated the influence of temperament on task performance and focused on its impact on dropout in visual habituation tasks. As we recognized in the course of the study that especially the habituation tasks had to be aborted ahead of time, we further investigated this phenomenon of

dropout. Habituation tasks per se pose special requirements to children: not only have they to deal with an unfamiliar environment and interact with strangers, but also the habituation task as such demands them to remain calm and focused on the presented stimuli. According to Slaughter and Suddendorf (2007) dropout rates up to more than 60% are common in habituation tasks. In their review, temperament has been discussed as one possible influencing factor. If children are excluded for fussiness that might be caused by certain temperamental traits hindering these children to regulate during habituation task, this dropout could systematically bias research findings. At last, Slaughter and Suddendorf (2007) could not find an influence of temperament on dropout whereas older studies did so (e.g., Miceli et al., 1998). In our project, the visual habituation tasks had to be aborted by the experimenter plenty of times, usually due to infants' fussy behavior. As the reported findings are controversial, we examined post-hoc if the high dropout rate might be caused by children's temperament. One of our main concerns, as we followed the same children over a variety of tasks in this longitudinal project, affected the question if the results of the remaining children completing the tasks were generalizable or if the dropouts systematically influenced them. Moreover, our study design and the fact that we employed two visual habituation tasks that only differed in content enabled us to longitudinally investigate the dependence of dropout on task content. Our results showed that temperament has only little impact on dropout; a finding that is in line with the more recent studies on this topic (e.g., Slaughter & Suddendorf, 2007). Only the ability to attend to something for an extended time seems to be related to completion of a habituation task. This is an important finding ensuring that study results are not specific for children with a special temperament but are generalizable on the population. Nevertheless, the results comparing both types of habituation tasks suggest that specific task characteristics, as for example attractiveness of stimuli, and time of task administration might be important to regard. Hence, depending on requirement and time of conduction, temperament might indeed play a role on dropout in infant studies. Consequently, visual habituation tasks should be conducted rather at the beginning of a test session to minimize a potential impact of temperament. Anyway, a negative influence of temperament on task performance might be negligible.

In addition, assessing temperament longitudinally with the same questionnaire enabled us to investigate the stability of temperament in infancy. Parental ratings of infant temperament showed both stability and change across the second half of the first year. Our findings indicate moderate to high normative stability in infant temperament between 6 and 12 months of age. Anyhow, in consistence with previous findings (Carranza Carnicero et al., 2000), most of the considered temperament domains are also found to be subject to developmental change between 6 and 12 months of age. Finally, study 3 provides converging evidence for the assumption that parental report is a valid but economic mean for assessing children's temperament especially in large samples (Rothbart & Bates, 1998). Although some research suggests that parental report on temperament might be biased (Leerkes & Crockenberg, 2003; Wolk et al., 1992), other studies assessing the agreement between parental report and observational data (Parade & Lerkes, 2008; Stifter et al., 2008) yielded support for the validity of the IBQ-R and IBQ, also in German translation (Pauli-Pott et al., 2005).

In summary, in this dissertation, a large number of children were investigated with comprehensive assessments of social cognition during the first three years of life. Still not all of the questions raised at the beginning of this work could be answered. In all three studies, hypotheses were proven true only in part. Further developmental research is needed to clarify the development of early social cognition and possible influencing factors. Still, continuity in social-cognitive development contains much more topics to address as for example the debate whether early social-cognitive abilities are early manifestations of a developing ToM or whether they are a required precondition for a ToM development. Also, further research should include temperament that was only recently considered as influencing factor on social cognition.

In conclusion, longitudinal research per se involves an enormous effort that cannot be undertaken by a single person. Not only does it consume a lot of time and money but also does it demand a huge amount of personal resources. Despite our collective effort in keeping in touch with the parents of our sample,

e.g. by sending birthday cards to the participating children, we were not able to prevent dropout from our project due to moving, parents' time pressure or simple disinterest to take part in further assessments over the course of the studies. As the studies cannot be conducted alone, research assistants have to be employed for different tasks. Due to the duration of a longitudinal project, many different research assistants have to be trained as they probably finish their degrees before the end of the assessments as we have experienced plenty of times. Also, it has to be emphasized that the single assessments have to be planned and sophisticated more thoughtfully and carefully than in cross-sectional research. Once decided on a certain method, the course has to be maintained until data collection is finished. You cannot change your study design without risking losing a part of your sample or the possibility to compare data between the tested children. However, to answer how children's social-cognitive abilities develop it is necessary to ensure that further longitudinal studies were conducted as only this kind of study fulfills the conditions to answer many research questions thoroughly.

References

Aschersleben, G., Hofer, T., & Jovanovic, B. (2008). The link between infant attention to goal-directed action and later theory of mind abilities. *Developmental Science*, *11*(6), 862-868.

Baillargeon, R. (1987). Young infants' reasoning about the physical and spatial proerties of a hidden object. *Cognitive Development, 2*(3), 178-200.

Baillargeon, R., Spelke, E. Z., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition, 20*(3), 191-208.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False belief understanding in infants. *Trends in Cognitive Sciences, 14*(3), 110-118.

Banerjee, R., & Henderson, L. (2001). Social-cognitive factors in childhood social anxiety. A preliminary investigation. *Social Development, 10*, 558-572.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46.

Barrera, M. E., & Maurer, D. (1981). The Perception of facial expression by the three-month-old. *Child Development, 52*(1), 203-206.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. Oxford university press.

Baudonnière, P. M., Margules, S., Belkhenchir, S., Carn, G., Pepe, F., & Warkentin, V. (2002). A longitudinal and cross-sectional study of the emergence of the symbolic function in children between 15 and 19 months of age: Pretend play, object permanence understanding, and self-recognition. In R. W. Mitchell (Ed.), *Pretending and imagination*

in animals and children (pp. 73-90). Cambridge, UK: Cambridge University Press.

Bellagamba, F., & Tomasello, M. (1999). Re-enacting intended acts: Comparing 12-and 18-month-olds. *Infant Behavior and Development*, *22*(2), 277-282.

Benenson, J. F., Philippoussis, M., & Leeb, R. (1999). Sex differences in neonates' cuddliness. *The Journal of Genetic Psychology, 160*(3), 332-342.

Bell, J. C., & Slater, A. (2002). The short-term and longer-term stability of non-completion in an infant habituation task. *Infant Behavior & Development, 25*(2), 147-160.

Bischof-Köhler, D. (1994). Selbstobjektivierung und fremdbezogene Emotionen. Identifikation des eigenen Spiegelbildes, Emapthie und prosoziales Verhalten im 2. Lebensjahr [Self-objectification and other-oriented emotions: Self-recognition, empathy, and prosocial behavior in the 2nd year] *Zeitschrift für Psychologie, 202*, 349-377.

Blair, C., Peters, R., & Granger, D. (2004). Physiological and neuropsychological correlates of approach/withdrawal tendencies in preschool: further examination of the behavioral inhibition system/behavioral activation system scales for young children. *Developmental Psychobiology*, *45*(3), 113-124.

Bornstein, M. H., Hahn, C. S., Suwalsky, J. T., & Haynes, O. M. (2003). Socioeconomic status, parenting, and child development: The Hollingshead Four-Factor Index of Social Status and The Socioeconomic Index of Occupations. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development. Monographs in parenting series* (pp. 29-82). Mahwah, NJ: Erlbaum

Bornstein, M. H., Haynes, O. M., O'Reilly, A. W., & Painter, K. M. (1996). Solitary and collaborative pretense play in early childhood: Sources of individual variation in the development of representational competence. *Child development*, *67*(6), 2910-2929.

Bornstein, M. H., Krinsky, S. J., & Benasich, A. A. (1986). Fine orientation discrimination and shape constancy in young infants. *Journal of Experimental Child Psychology, 41*(1), 49-60.

Bradmetz, J., & Schneider, R. (1999). Is Little Red Riding Hood afraid of her grandmother? Cognitive vs. emotional response to a false belief. *British Journal of Developmental Psychology*, *17*(4), 501-514.

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

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

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy*, *5*(3), 291-308.

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*(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.

Carranza Carnicero, J. A., Péréz-Lopéz, J., Del Carmen González Salinas, M., Martínez-Fuentes, M. T. (2000). A longitudinal study of temperament

in infancy: Stability and convergence of measures. *European Journal of Personality, 14*(1), 21-37.

Casalin, S., Luyten, P., Vliegen, N., & Meurs, P. (2012). The structure and stability of temperament from infancy to toddlerhood: A one-year prospective study. *Infant Behavior & Development, 35*(1), 94-108.

Cassidy, K. W., Fineberg, D. S., Brown, K., & Perkins, A. (2005). Theory of mind may be contagious, but you don't catch it from your twin. *Child Development*, *76*(1), 97-106.

Charman, T. (1998). Specifying the nature and course of the joint attention impairment in autism in the pre-school years: implications for diagnosis and intervention. Autism: The International Journal of Research and Practice, 2, 61-79.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, *15*(4), 481-498.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cole, K., & Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. *British Journal of Developmental Psychology*, *18*(2), 279-295.

Colombo, J., & Mitchell, D.W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory, 92*(2), 225-234.

Crick, N. R., Bigbee, M. A., & Howes, C. (1996). Gender differences in children's normative beliefs about aggression: How do I hurt thee? Let me count the ways. *Child development*, *67*(3), 1003-1014.

Cutting, A. L., & Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: Individual differences and interrelations. *Child Development*, *70*(4), 853-865.

De Villiers, J. G., & De Villiers, P. A. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. J. Riggs (Eds.) *Children's reasoning and the mind*. (pp. 191-228). Hove, England: Psychology Press

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1432-1443.

Dragan, W. L., Kmita, G., & Fronczyk, K. (2011). Psychometric properties of the polish adaption of the infant behavior questionnaire-revised (IBQ-R). *International Journal of Behavioral Development, 35*(6), 542-549.

Dunn, J., & Brown, J. (1994). Affect expression in the family, children's understanding of emotions, and their interactions with others. *Merrill-Palmer Quarterly, 40*(1), 120-137.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development*, *62*(6), 1352-1366.

Else-Quest, N. M., Shibley Hyde, J., Hill Goldsmith, H., Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin, 132*(1), 33-72.

Evans, A. D., & Lee, K. (2013). Emergence of lying in very young children. *Developmental psychology*, *49*(10), 1958-1963.

Evans, D. E., & Rothbart, M. K. (2009). A two-factor model of temperament. *Personality & Individual Differences, 47*(6), 565-570.

Fagen, J. W., Ohr, P. S., Singer, J. M., & Fleckenstein, L. K. (1987). Infant temperament and subject loss due to crying during operant conditioning. *Child Development, 58*(2), 497-504.

Farhadian, M., Gazanizad, N., & Shakerian, A. (2011). Theory of Mind and Siblings among Preschool Children. *Asian Social Science*, *7*(3), p224.

Farrar, M. J., Lee, H., Cho, Y. H., Tamargo, J., & Seung, H. (2013). Language and false belief in Korean-speaking and English-speaking children. *Cognitive Development*, *28*(3), 209-221.

Flavell, J. H. (1985). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.

Funder, D. C., & West, S. G. (1993). Consensus, self–other agreement, and accuracy in personality judgment: An introduction. *Journal of Personality, 61*(4)*,* 457–476.

Gallup Jr, G. G. (1991). Toward a comparative psychology of self-awareness: Species limitations and cognitive consequences. In J. Strauss & G. R. Goethals (Eds.), *The self: interdisciplinary approaches* (pp. 121-135). Springer New York.

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior & Development, 26*(1), 64-86.

Gartstein, M. A., Knyazev, G. G., & Slobodskaya, H. R. (2005). Cross-cultural differences in the structure of infant temperament: Unites States of America (U.S.) and Russia. *Infant Behavior & Development, 28*(1), 54-61.

Gartstein, M. A., Gonzalez, C., Carranza, J. A., Ahadi, S. A., Ye, R., Rothbart, M. K., & Yang, S. W. (2006). Studying cross-cultural differences in the development of infant temperament: People's Republic of China, the United States of America, and Spain. *Child Psychiatry & Human Development*, *37*(2), 145-161.

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., Hinde, R. A., & McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. *Child development*, *58*, 505-529.

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*, 26-37.

Gopnik, A. and Meltzoff, A.N. (1994) Minds, bodies, and persons: young children's understanding of the self and others as reflected in imitation and theory of mind research. In S. T. Parker, R. W. Mitchell, M. L. Boccia, & A. Whiten (Eds.), *Self-Awareness in Animals and Humans: Developmental Perspectives* (pp. 166–186), Cambridge: University Press

Grimm, H. (2001). *Sprachentwicklungstest für Kinder (SETK 3–5)* [Language development test for 3- to 5-year-old children]. Göttingen, Germany: Hogrefe.

Grimm, H., & Doil, H. (2006). *Elternfragebögen für die Erfassung von Risikokindern* [Parental questionnaire to identify children at risk for developmental language disorders](2nd edn.). Göttingen, Germany: Hogrefe.

Gunnar, M. R., Porter, F. L., Wolf, C. M., Rigatuso, J., & Larson, M. C. (1995). Neonatal stress reactivity: Predictions to later emotional temperament. *Child Development, 66*(1), 1-13.

Haccou, P., & Meelis, E. (1994). *Statistical Analyses of Behavioral Data.* Oxford, UK: Oxford University Press.

Hare, B. (2007). From Nonhuman to Human Mind What Changed and Why? *Current Directions in Psychological Science*, *16*(2), 60-64.

Hare, B., Melis, A. P., Woods, V., Hastings, S., & Wrangham, R. (2007). Tolerance allows bonobos to outperform chimpanzees on a cooperative task. *Current Biology*, *17*(7), 619-623.

Hare, B. & Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences, 9*(9), 439-465.

Harris, P. L. (1991). The work of the imagination. In *Natural theories of mind: evolution, development and simulation of everyday mindreading* (ed. A. Whiten), pp. 283–304. Cambridge, MA: Blackwell.

Harris, P. L. (1992). From simulation to folk psychology: The case for development. *Mind & Language*, *7*(1-2), 120-144.

Harris, P. (1996). Desires, beliefs, and language. In P. Carruthers & P. K. Smith (Eds.). *Theories of theories of mind*. (pp.200-220). Cambridge: Cambridge University Press

Harris, P. L. & Kavanaugh, R. D. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development*, *58*(1, Serial No. 231).

Henderson, H. A., & Wachs, T. D. (2007). Temperament theory and the study of cognition-emotion interactions across development. *Developmental Review, 27*(3), 396-427.

Henning, A., Daum, M. M., & Aschersleben, G. (2009). Frühkindliche Handlungswahrnehmung und Theory of Mind [Infant action perception and Theory of mind]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *41*(4), 233-242.

Henning, A., Spinath, F.M., & Aschersleben, G. (2011). The link between preschoolers' executive function and theory of mind and the role of epistemic states. *Journal of Experimental Child Psychology, 108*(3), 513-531.

Herrmann, E., Hare, B., Call, J., & Tomasello, M. (2010). Differences in the cognitive skills of bonobos and chimpanzees. *PloS one*, *5*(8), e12438.

Hinchcliffe, V. (1996). Fairy stories and children's developing theories of mind. *International Journal of Early Years Education*, *4*(1), 35-46.

Hofer, T., & Aschersleben, G. (2007). *'Theory of Mind'-Skala für 3- bis 5-jährige Kinder* [German theory of mind scale for 3- to 5-year-old children] (3rd edn.). Max Planck Institute for Human Cognitive and Brain Sciences, Munich, Germany.

Hofer, T., Hohenberger, A., Hauf, P., & Aschersleben, G. (2008). The link between maternal interaction style and infant action understanding. *Infant Behavior & Development, 31*(1), 115-126.

Hohenberger, A., Elsabbagh, M., Serres, J., de Schoenen, S., Karmiloff-Smith, A. & Aschersleben, G. (2012). Understanding goal-directed human actions and physical causality: the role of mother-infant interaction. *Infant Behavior & Development, 35*(4), 898-911.

Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology*, *32*(1), 70-78.

Jovanovic, B., Király, I., Elsner, B., Gergely, G., Prinz, W., & Aschersleben, G. (2007). The role of effects for infants' perception of action goals. *Psychologia, 50*(4), 273-290.

Kagan, J., & Lewis, M. (1965). Studies of attention in the human infant. *Merrill-Palmer Quarterly, 11*(2), 95-127.

Kaufman, J., & Needham, A. (1999). Objective spatial coding by 6.5-month-old infants in a visual dishabituation task. *Developmental Science, 2*(4), 432-441.

Kohnstamm, G. A. (1989). Temperament in childhood: Cross-cultural and sex differences. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart

(Eds.), *Temperament in childhood*. (pp. 483-508). Oxford, England: John Wiley & Sons.

Komsi, N., Räikkönen, K., Heinonen, K., Pesonen, A.-K., Keskivaara, P., Järvenpää, A.-L., & Strandberg, T. E. (2008). Continuity of father-rated temperament from infancy to middle childhood. *Infant Behavior & Development, 31*(2), 239-254.

Komsi, N., Räikkönen, K., Pesonen, A.-K., Heinonen, K., Keskivaara, P., Järvenpää, A.-L., & Strandberg, T. E. (2006). Continuity of temperament from infancy to middle childhood. *Infant Behavior & Development, 29*(4), 494-508.

Kotovsky, L. & Baillargeon, R. (1994). Calibration-based reasoning about collision events in 11-month-old infants. *Cognition, 51*(2), 107-129.

Kotovsky, L. & Baillargeon, R. (1998). The development of calibration-based reasoning about collision events in young infants. *Cognition, 67*(3), 311-351.

Kristen, S., Eisenbeis, H., Thoermer, C., & Sodian, B. (2007). Temperamentsfragebogen für Babys – revidierte Form. [Infant Behavior Questionnaire-revised version.] Retrieved Oct. 2007, available from http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/instrument-descriptions/infant-behavior-questionnaire.html

Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Skalierung von Theory of Mind Aufgaben [Scaling of theory of mind tasks]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38*, 186-195.

Kristen, S., Sodian, B., Thoermer, C., & Perst, H. (2011). Infants' joint attention skills predict toddlers' emerging mental state language. *Developmental psychology*, *47*(5), 1207-1219.

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*(5), 825-836.

Lecce, S., & Hughes, C. (2010). The Italian job?: Comparing theory of mind performance in British and Italian children. *British Journal of Developmental Psychology*, *28*(4), 747-766.

Leerkes, EM. & Crockenberg, S.C. (2003). The impact of maternal characteristics and sensitivity on the concordance between maternal reports and laboratory observations of infant negative emotionality. *Infancy, 4*(4), 517-539.

Leslie, A. M. (1992). Pretense, autism, and the theory-of-mind module. *Current Directions in Psychological Science*, 18-21.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity in cognition and culture. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119-148). New York: Cambridge University Press.

Leslie, A.M., Friedman, O., & German, T.P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences, 8*(12)*,* 528–533.

Lewis, C., Freeman, N. H., Kyriakidou, C., Maridaki‐Kassotaki, K., & Berridge, D. M. (2008). Social influences on false belief access: specific sibling influences or general apprenticeship? *Child Development*, *67*(6), 2930-2947.

Lewis, M., & Ramsay, D. (2004). Development of Self‐Recognition, Personal Pronoun Use, and Pretend Play During the 2nd Year. *Child Development*, *75*(6), 1821-1831.

Lewis, M., Stanger, C., & Sullivan, M. W. (1989). Deception in 3-year-olds. *Developmental psychology*, *25*(3), 439-443.

Lewis, M., Sullivan, M. W., Stanger, C., & Weiss, M. (1989). Self development and self-conscious emotions. *Child Development, 60*, 146-156.

Lillard, A. S. (1993). Pretend play skills and the child's theory of mind. *Child development*, *64*(2), 348-371.

Lillard, A. S. (1997). Other folks' theories of mind and behavior. *Psychological Science*, *8*(4), 268-274.

Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin*, *139*(1), 1.

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: a meta-analysis of false-belief understanding across cultures and languages. *Developmental psychology*, *44*(2), 523.

Lockl, K., Schwarz, S., & Schneider, W. (2004). Sprache und Theory of Mind [Language and Theory of Mind]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *36*(4), 207-220.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*(4), 1130-1144.

Lucariello, J. M., Durand, T. M., & Yarnell, L. (2007). Social versus intrapersonal ToM: Social ToM is a cognitive strength for low-and middle-SES children. *Journal of Applied Developmental Psychology*, *28*(4), 285-297.

Mahy, C. E. V, Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience, 9C*, 68-81.

McAlister, A., & Peterson, C. C. (2006). Mental playmates: Siblings, executive functioning and theory of mind. *British Journal of Developmental Psychology*, *24*(4), 733-751.

McAlister, A., & Peterson, C. (2007). A longitudinal study of child siblings and theory of mind development. *Cognitive Development*, *22*(2), 258-270.

McAlister, A., & Peterson, C. (2013). Siblings, theory of mind, and executive functioning in children aged 3–6 years: New longitudinal evidence. *Child development*, *84*(4), 1442-1458.

Meins, E., Fernyhough, C., Russell, J., & Clark-Carter, D. (1998). Security of attachment as a predictor of symbolic and mentalising abilities: A longitudinal study. *Social development*, *7*(1), 1-24.

Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal mind–mindedness and attachment security as predictors of theory of mind understanding. *Child development*, *73*(6), 1715-1726.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838.

Miceli, P. J., Whitman, T. L., Borkowski, J. G., Braungart-Rieker, J., & Mitchell, D. W. (1998). Individual differences in infant information processing: The role of temperamental and maternal factors. *Infant Behavior & Development, 21*(1), 119-136.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-Analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622-646.

Montirosso, R., Cozzi, P., Putnam, S. P., Gartstein, M. A., & Borgatti, R. (2011). Studying cross-cultural differences in temperament in the first

year of life: United States and Italy. *International Journal of Behavioral Development, 35*(1), 27-37.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). Early social communication scales (ESCS). *Coral Gables, FL: University of Miami*.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child development*, *78*(3), 938-954.

Nakagawa, A. & Sukigara, M. (2005). How are cultural differences in the interpretation of infant behavior reflected in the Japanese revised Infant Behavior Questionnaire? [Abstract]. *Japanese Journal of Educational Psychology*, *53*(V), 491.

Nielsen, M., & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior and Development*, *27*(3), 342-365.

Nikolaizig, F. (2007). *Temperamentsfragebogen für Kinder* [Childhood Behavior Questionnaire]. Retrieved Oct. 2007, available from http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/instrument-descriptions/childrens-behavior-questionnaire.html

Okamoto-Barth, S., Call, J., & Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychological Science*, *18*(5), 462-468.

Olineck, K. M., & Poulin-Dubois, D. (2005). Infants' ability to distinguish between intentional and accidental actions and its relation to internal state language. *Infancy*, *8*(1), 91-100.

Olineck, K. M., & Poulin-Dubois, D. (2009). Infants' understanding of intention from 10 to 14 months: Interrelations among violation of

expectancy and imitation tasks. *Infant Behavior and Development*, *32*(3), 404-415.

Parade, S. H., & Leerkes, E. M. (2008). The reliability and validity of the Infant Behavior Questionnaire-Revised. *Infant Behavior & Development, 31*(4), 637-646.

Pauli-Pott, U., Mertesacker, B., & Beckmann, D. (2003). Ein Fragebogen zur Erfassung des «frühkindlichen Temperaments» im Elternurteil. [A questionnaire for the assessment of «infant temperament» via parental report.] *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 31*(2), 99-110.

Pauli-Pott, U., Mertesacker, B., & Beckmann, D. (2005). Ein Vergleich unterschiedlicher Erfassungsmethoden der frühkindlichen Emotionalität. [Comparing assessment methods of infant emotionality.] *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie*, *33*(2), 123-135.

Pears, K. C., & Moses, L. J. (2003). Demographics, parenting, and theory of mind in preschool children. *Social Development*, *12*(1), 1-20.

Perner, J. (1991). *Understanding the representational mind*. The MIT Press.

Perner, J. (1999). Theory of mind. *Developmental psychology: Achievements and prospects*, 205-230.

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3,* 337–344.

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development*, *65*(4), 1228-1238.

Perner, J., Sprung, M., Zauner, P., & Haider, H. (2003). Want that is understood well before Say that, Think that, and false belief: A test of de Villiers's linguistic determinism on German–speaking children. *Child development*, *74*(1), 179-188.

Peterson, C. C. (2000). Kindred spirits: Influences of siblings' perspectives on theory of mind. *Cognitive Development*, *15*(4), 435-455.

Peterson, C., & Slaughter, V. (2003). Opening windows into the mind: Mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development*, *18*(3), 399-429.

Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*(1), 53-78.

Povinelli, D. J., & Eddy, T. J. (1996). Chimpanzees: Joint visual attention. *Psychological Science*, *7*(3), 129-135.

Povinelli, D. J., & Simon, B. B. (1998). Young children's understanding of briefly versus extremely delayed images of the self: Emergence of the autobiographical stance. *Developmental Psychology*, *34*(1), 188-194.

Premack, D., & Woodruff, G. (1978a). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515-526.

Premack, D., & Woodruff, G. (1978b). Chimpanzee problem-solving: a test for comprehension. *Science*, *202*(4367), 532-535.

Putnam. S. P., Ellis, L. K., & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. In A. Eliasz & A. Angleitner (Eds.), *Advances in research on temperament* (pp. 165-182). Lengerich: Pabst Scientific.

Putnam. S. P., Gartstein, M. A., & Rothbart, M. K (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development, 29*(3), 386-401.

Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy,

toddlerhood, and early childhood. *Infant and Child Development*, *17*(4), 387-405.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: evidence from 14-and 18-month-olds. *Developmental Psychology*, *33*(1), 12.

Rink, K. (2006). *Temperamentsfragebogen für Kinder* [Early Childhood Behavior Questionnaire]. Retrieved Sep. 2008, available via Email from k.rink@psychologie.uzh.ch

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52*(2), 569-578.

Rothbart, M. K. (1986). Longitudinal observation of infant temperament. *Developmental Psychology, 22*(3), 356-365.

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*(5), 1394-1408.

Rothbart, M. K., & Bates, J.E. (1998). Temperament. In W. Damon & N. Eisneberg (Eds.), *Handbook of child psychology, Vol. 3. Social, emotional and personality development.* (5th ed., pp. 105-176). New York: Wiley.

Rothbart, M. K., & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology, Vol. 1* (pp. 37-86). Hillsdale, NJ: Erlbaum.

Rothbart, M. K., Derryberry, D., & Hershey, K. (2000). Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 85-119). Mahwah, NJ: Erlbaum.

Rothbart, M. K. & Gartstein, M.A. (2000). Score sheet of the Infant Behavior Questionnaire – Revised. Retrieved Oct. 2007, available from http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/instrument-descriptions/infant-behavior-questionnaire.html

Ruffman, T., Perner, J., Naito, M., Parkin, L., & Clements, W. A. (1998). Older (but not younger) siblings facilitate false belief understanding. *Developmental Psychology*, *34*(1), 161-174.

Ruffman, T., Perner, J., & Parkin, L. (1999). How parenting style affects false belief understanding. *Social Development*, *8*(3), 395-411.

Saudino, K. J., & Cherny, S. S. (2001). Sources of continuity and change in observed temperament. In R. N. Emde & J. K. Hewitt (Eds.), *Infancy to early childhood: genetic and environmental influences on developmental change* (pp. 89-110). New York: Oxford University Press.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind.* Cambridge, UK: Cambridge University Press.

Seifer, R., Sameroff, A. J., Barrett, L. C., & Krafchuk, E. (1994). Infant temperament measured by multiple observations and mother report. *Child development*, *65*(5), 1478-1490.

Slater, A. (1995). Visual perception and memory at birth. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research*, (Vol. 9, pp. 107–162). Norwood, NJ: Ablex.

Slaughter, V. & Suddendorf, T. (2007). Participant loss due to „fussiness“ in infant visual paradigms: A review of the last 20 years. *Infant Behavior & Development, 30*(3), 505-514.

Slobodskaya, H. R., Gartstein, M. A., Nakagawa, A., & Putnam, S. P. (2012). Early temperament in Japan, the United States, and Russia: Do cross-cultural differences decrease with age? *Journal of Cross-Cultural*

Psychology. Published online before print July 20, 2012, doi: 10.1177/0022022112453316.

Sodian, B., Hülsken, C., Ebner, C., & Thoermer, C. (1998). Die begriffliche Unterscheidung von Mentalität und Realität im kindlichen Symbolspiel: Vorläufer einer Theory of Mind? [Children's differentiation of mentality and reality in pretense: precursor to a theory of mind ?]. *Sprache & Kognition*, *17*(4), 199-213

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental Psychology*, *51*(9), 1190-1200.

Sodian, B. & Thoermer, C. (2006). Theory of Mind. In W. Schneider & B. Sodian (Eds.), *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie V: Entwicklungspsychologie, Band 2: Kognitive Entwicklung* (pp. 495-608). Göttingen: Hogrefe

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*(4), 605-632.

Stifter, C. A., Willoughby, M. T., Towe-Goodman, N. & The Family Life Project Key Investigators (2008). Agree or agree to disagree? Assessing the convergence between parents and observers on infant temperament. *Child Development*, *17*(4), 407-426.

Symons, D. K., & Clark, S. E. (2000). A Longitudinal Study of Mother-Child Relationships and Theory of Mind in the Preschool Period. *Social Development*, *9*(1), 3-23.

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child development*, *79*(4), 866-881.

Thomas, A., & Chess, S. (1977). *Temperament and development.* New York: Bruner/Mazel.

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*(1), 172-187.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention. Its origins and role in development* (pp. 103-130). Hillsdale, NJ: Erlbaum.

Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind & Language*, *18*(2), 121-147.

Treiber, F.A. (1984). Temperament differences between infants who do and do not complete laboratory testing. *The Journal of Psychology: Interdisciplinary and Applied, 116*(1), 95-99.

Van den Boom, D. (2004). First attachments: Theory and research. In G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development* (pp. 296-325). Malden, MA: Blackwell.

Vonderlin, E., Pahnke, J., & Pauen, S. (2008). Infant temperament and information processing in a visual categorization task. *Infant Behavior & Development, 31*(4), 559-569.

Vonderlin, E., Ropeter, A., & Pauen, S. (2012). Erfassung des frühkindlichen Temperaments mit dem Infant Behavior Questionnaire Revised. Psychometrische Merkmale einer deutschen Version. [Assessment of infant temperament with the Infant Behavior Questionnaire Revised (IBQ-R) - the psychometric properties of a German version.] *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 40*(5), 307-314.

Wachs, T. D., & Smitherman, C. H. (1985). Infant temperament and subject loss in a habituation procedure. *Child Development, 56*(4), 861-867.

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*(3), 297-312.

Wellman H. M. (2012). Theory of mind: Better methods, clearer findings, more development. European Journal of Developmental Psychology, 9, 313-330.

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

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science*, *17*, 1075-1081.

Wellman, H. M., Lane, J. D., LaBounty, J., & Olson, S. L. (2011). Observant, nonaggressive temperament predicts theory-of-mind development. *Developmental Science*, *14*(2), 319–326.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541.

Wellman, H. M., Lopez-Duran, S., LaBounty, J., & Hamilton, B. (2008). Infant attention to intentional action predicts preschool theory of mind. *Developmental Psychology*, *44*(2), 618-623.

Wellman, H. M., Phillips, A. T., Dunphy‐Lelii, S., & LaLonde, N. (2004). Infant social attention predicts preschool social cognition. *Developmental Science*, *7*(3), 283-288.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Wolk, S., Zeanah, C. H., Coll, C. T. G., & Carr, S. (1992). Factors affecting parents' perceptions of temperament in early infancy. *American Journal of Orthopsychiatry*, *62*(1), 71-82.

Woodward, A.L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*(1), 1-34.

Woodward, A.L. (1999). Infants' ability to distinguish between purposeful and nonpurposeful behaviors. *Infant Behavior & Development, 22*(2), 145-160.

Xu, F., & Spelke, E.Z. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1-B11.

Yamaguchi, M., Kuhlmeier, V. A., Wynn, K., & VanMarle, K. (2009). Continuity in social cognition from infancy to childhood. *Developmental science*, *12*(5), 746-752.

Youngblade, L. M., & Dunn, J. (1995). Individual differences in young children's pretend play with mother and sibling: Links to relationships and understanding of other people's feelings and beliefs. *Child Development*, *66*(5), 1472-1492.

Zentner, M., & Bates, J.E. (2008). Child temperament: An integrative review of concepts, research programs, and measures. *European Journal of Developmental Science, 2*(1), 7–37.

Zuckerman, M. (1990). The psychophysiology of sensation seeking. *Journal of Personality, 58*(1), 313-345.