Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Development of Social-Cognitive Abilities in the First Two Years of Life

Dissertation

zur Erlangung des Doktorgrades
Doctor rerum naturalium

an der
Universität Hamburg
Fakultät für Psychologie und Bewegungswissenschaft
Institut für Psychologie

vorgelegt von

Marianna Jartó

Hamburg, Dezember 2020

PROMOTIONSPRÜFUNGSAUSSCHUSS/COMMITTEE

Vorsitzender/Chair of the Committee

Prof. Dr. rer. nat. Lars Schwabe

1. Dissertationsgutachter/Evaluator of the Dissertation

Prof. Dr. rer. nat. Ulf Liszkowski

2. Dissertationsgutachterin/Evaluator of the Dissertation

Prof. Dr. Aylin Küntay

1. Disputationsgutachterin/Evaluator of the Defense

Prof. Dr. phil. Juliane Degner

2. Disputationsgutachterin/Evaluator of the Defense

Prof. Dr. rer. nat. Tania Lincoln

Tag der Disputation/Date of the Defense: 18. November 2020

To Ilse Lugscheider

Acknowledgements

First of all, I want to thank my supervisors for their helpful input and fruitful discussions. Thank you, Ulf, for always having an open ear and sharing your experience and ideas with me. Thank you, Aylin, for all the great exchange on several conferences and meetings. It was great to be part of this exciting international project. Without both of you I wouldn't have got such deep insight into the discourse of current infant research and got so close to the academic community. Thank you for this unique experience.

Furthermore, I want to thank my colleagues Johanna, Wiebke, Susanna, Sebastian, Ranjani, Mareike, Jessica, Remi, Anke, Marlena and Katharina for their support in all possible ways. Having you around for scientific exchange, technical and programming coaching, regular lunch sessions and after work drinks was an important driver for this thesis and my motivation. Thank you very much!

A big thanks goes to my friends from Vienna as well as from Hamburg. Your regular visits, endless working sessions and parties, open ears and supportive words were important precursors for the development of this thesis. Thanks especially to Puppi and Hermann for motivating me. In particular, I want to thank my partner Jan for his helpful scientific support, his positive words and his automatic and communicated extension of his tolerance limit.

I also want to thank my family for their curious questions in interaction-based settings and their belief in me although I was never able to answer their question, what I exactly do, in a satisfactory way. In particular, I want to thank my brother Stefan, who plays an important role in my development and my grandfather Walter for his patient proofreading.

Finally, I want to thank all the participating families and their infants for their long term support and all the graduate and undergraduate students for their assistance at stimuli preparation and data collection.

Without all of you this thesis wouldn't have been possible. Thank you very much!

Contents

List of Figures

List of Tables

Abbreviations

VPT **V**isual **P**erspective **T**aking
SOA **S**timulus **O**nset **A**synchrony
VOE **V**iolation **O**f **E**xpectation
AOI **A**rea **O**f **I**nterest

Abstract

Infants develop important social cognitive abilities in the first year of life. Interaction changes from dyadic to triadic and infants start to understand goal-directed and communicative actions. However, different theories and findings exist about the emergence of early social cognitive abilities and to date there is little systematic comparison and longitudinal evidence. Some argue that action understanding and communicative understanding are present from birth and emerge early on in behavior (6-months-olds) as separate systems. In contrast, others suggest that infants develop an integrated understanding of humans as goal-directed and perceiving organisms through second-person interaction around their first birthday. While some paradigms assess infants as observers of social situations, other paradigms directly involve infants as interactants in social situations. The aim of this thesis is (1) to systematically compare social cognitive abilities in different situations and age groups (Study 1 & Study 2) and (2) to examine stable manifestations and developmental relations of action understanding and communicative understanding (Study 3).

In Study 1 + Study 2, I used a cueing paradigm with a centrally presented person looking behind one of two barriers to test visual-perspective-taking (VPT) in communicative vs. non-communicative situations. To measure reflective VPT ability in contrast to reflexive spatial cueing, the cue was presented for a long time (3000ms). Infants looking time to the cued barrier revealing an empty box was used as indication for object expectation. Results of Study 1 revealed that 14-months-olds need communicative cues to represent what others can see. At the age of three years, VPT seems to be automatized to non-communicative cues similar to adults. In Study 2, only weak VPT was found for younger infants (8-months-olds). Results speak for the emergence of social cognitive abilities later in development.

In Study 3, I measured action understanding and communicative understanding in interaction-based and eye-tracking paradigms using a longitudinal design. Results indicate an emergence of stable abilities at 11 months. Relations between action understanding and communicative understanding is in support of the hypothesis that they are part of one integrated understanding of humans. Method comparisons revealed earlier competencies in interaction-based measures than in eye-tracking, supporting the assumption that infants develop an understanding through second-person interaction.

Together, results support the view that infants' social cognitive skills emerge at the end of the first year of life, instead of a present competence from early on. This understanding emerges through and first reveals itself in communicative interactions. Future research should focus on social-interactional experiences as predictors of intra-individually stable, emerging social cognition at the end of the first year of life.

Chapter 1

Introduction

In everyday interaction, humans have to coordinate their behavior in very complex ways. When someone steps up to a barrier looking at their phone and we want to warn them, we first have to anticipate their future action (they will hit the barrier, if they do not look up) and take into account what they see or pay attention to (not the barrier). We will shout out to get their attention and point to the obstacle to warn them. They will understand what we want to tell them, and that we want to communicate them something. This shows that we understand humans' actions as intentional, we know that others have their own perspectives, and we can mentally represent their representations. Most importantly we communicate with the intention to change the other person's mental representation, which seems to be specific to our species (Tomasello, 2019)

The foundation of these human-specific abilities seem to be found early on in life (Brooks & Meltzoff, 2015; Sodian et al., 2016; Tomasello, 2019). Around their first birthday, infants start to show objects to their interaction partners or point to outside entities, indicating first communicative actions (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998). Even earlier, they seem to be able to interpret human actions as directed to a goal (Woodward, 1998) and understand communicative actions by following others' gazes and pointing gestures (Carpenter et al., 1998). Some argue that they even know about others' mental representations and underlying goals and intentions (Luo & Baillargeon, 2010), thus, having a mentalistic understanding. Early social cognitive abilities are known to be important precursors for more complex social cognition and language development (Brooks & Meltzoff, 2015). Thus, understanding their development is important for possible early interventions to prevent for example language delay, and brings light into human-specific development.

Even though a variety of theories exists, there is little systematic and longitudinal empirical evidence about the development of early social cognitive abilities. Some argue that infants first need to experience human-specific, second-person interaction in communicative settings to learn about perspectives, i.e. knowing that humans can see and form mental representations (Tomasello, 2019). Others argue that a mentalistic understanding is innate and activated at a certain time point without the need of experience (Baillargeon, Scott, & Bian, 2016; Csibra, 2003). Theories also differ as to when mentalistic action interpretation first emerges. That is to know that humans have mental representations, that they act based on their goals and that they refer to something, when they communicate. In addition, there are divergent assumptions about whether social-cognitive abilities emerge as an integrated understanding of others (Tomasello, Carpenter, Call, Behne, & Moll, 2005) or as separate systems (Csibra, 2003). Thus, it is the aim of this thesis to test the assumptions of these different theories empirically with a systematic cross-sectional design and a longitudinal approach.

In order to do so, I address three major questions. First, in which situations do infants take into account others' perspectives and mental representations indicated by a referential object expectation. Do they need communicative situations, or do they show these skills in less communicative situations as well? Finding evidence in communicative situations would support the need of social interaction to develop an understanding of others' perspectives. Second, when do infants show stable manifestations of early social cognitive abilities? Finding competence from early on would be in line with abilities being innate. Third, how are action-understanding and communicative-understanding developmentally related to one another? Finding relations between abilities would suggest an integrated understanding of others' intentions and perspectives.

In Study 1, I will compare VPT in communicative (pointing gesture) and non-communicative (still face, facial expression) situations. In Study 2, I will test for indications of VPT in very young infants (8-months) in communicative actions (pointing) and instrumental actions (reaching actions). In Study 3, I will examine communicative and action understanding using a longitudinal design to test how these two abilities are related to each other in their early emergence. In addition, I included different methods to validate measures used in infant studies.

In the following sections, I will first discuss different definitions and interpretations of early social cognitive abilities (Chapter 1.1), by addressing action understanding and communicative understanding. Then I will present different methods used in infant research and the importance of considering the used method (Chapter 1.2), followed by an illustration of two different theories on the development of these early social cognitive abilities (Chapter 1.3). As briefly mentioned above, the first theory assumes abilities to be innate and activated by maturation (1.3.1 Nativist view), whereas the second theory leans toward a more constructivist view by developing abilities through experience and interaction (1.3.2 Constructivist view). Finally, I will outline the research questions (Chapter 1.4.) and the work plan of the present research (Chapter 1.4.1).

1.1 Social Cognitive Abilities

Early social cognitive abilities are interpreted in different ways. They range from very rich interpretations, assuming that infants understand the underlying intentions and mental states of others (Luo & Baillargeon, 2010), to very lean interpretations, suggesting that infants only react on behavioral cues without taking into account the underlying goal or representation (Csibra, 2003).

In the following, I will provide different interpretations of action understanding and communicative understanding

1.1.1 Action Understanding

Understanding that humans act intentionally and in a goal-directed manner is fundamental to our understanding of others. In every day interaction we have to interpret and anticipate others' actions constantly. Most of our actions are instrumental, that is they have a purpose and a goal. Already, very young infants seem to have an understanding of others' actions as goal-directed (Woodward, 1998). For example, in a cueing paradigm, where a grasping hand is centrally presented for some milliseconds and an object appears either on the cued side (congruent) or on the not-cued side (incongruent), 7-month-olds detect the object in congruent trials earlier than in incongruent trials (cueing effect) (Daum & Gredebäck, 2011). In addition, 6-month-olds expect adults to reach for the same object they reached before several times, even when the position of the object changed (Woodward, 1998). They are also able to anticipate the goal-object of a reaching action when it is hidden (Applin & Kibbe, 2019). Thus, infants seem to understand some aspects of instrumental actions and in addition may understand the underlying motivations and preferences. However, instead of an understanding of the action or the goal, lower level processes, like attention direction or statistical learning, could be responsible for the demonstrated effects (Uithol & Paulus, 2014). It may be that an involuntary, reflexive gaze shift causes the cueing effect instead of interpreting the grasping hand (Wronski & Daum, 2014). In addition, when infants see a person grasping for an object several times the expectation that the person reaches for the same object again, even if the location of the object changes, could be due to statistical regularity learning (Ruffman, 2014).

One possibility to exclude lower level processes is to test if infants infer the goal of an action, even if they never see the end state of the action (Brandone, Horwitz, Aslin, & Wellman, 2014). Infants are able to anticipate unfulfilled goals only later on in development. Brandone et al. (2014) found that infants predicted a goal by looking at the to-be-grasped object in an unsuccessful reaching action only after they turned ten months old. In interaction-based measures, infants anticipate the goal and correctly react to unsuccessful actions even later (Warneken & Tomasello, 2007). 14-month-olds passed an out-of-reach helping task, where infants had to hand over an object that fell outside the adult's reach (Warneken & Tomasello, 2007). When infants had to complete more complex actions, where an adult failed to do so, like dropping a chain into a container, only 18-month-olds but not 14-month-olds passed (Meltzoff, 1995). Thus, these longer presented sequences where a voluntary gaze shift or an behavior reaction is required may measure a more sophisticated understanding of others' reaching actions (Gredebäck & Daum, 2015).

However, even with respect to these tasks there is some discussion as to whether infants have to understand the underlying intention or goal in order to pass the task, which comprises the mental state of the person. Instead, it might be sufficient to understand that human actions are efficient and goal-directed (Csibra & Gergely, 2013). Uithol and Paulus (2014) argue that the underlying ability could be some kind of pattern completion also visible in non-social physical actions. In contrast, others assume that very young infants take the perspective of an agent into account when interpreting goal directed actions (Luo & Johnson, 2009). Thus, there seems to be no clear picture what infants understand when they show abilities in different tasks and when infants start to understand instrumental actions as directed by underlying goals and mental representations yet.

1.1.2 Communicative Understanding

Another important aspect of social cognition is early communication. Infants communicate long before they start to speak by pointing to things and by following others' pointing gestures and attention (Liszkowski, 2018). Early language development is dependent on knowing what adults are looking at or referring to (Brooks & Meltzoff, 2005). Infants follow others' gazes and pointing gestures beginning at around 4 to 6 months (Farroni, Mansfield, Lai, & Johnson, 2003). This is especially visible in cueing paradigms, where a face or a pointing hand is presented as an endogenous cue for a short time (100 – 1000ms) and an object appears either on the cued or uncued side. Infants detect the object on the cued side earlier than on the uncued side, which is seen as indication for some understanding of gaze or pointing gestures (Bertenthal, Boyer, & Harding, 2014). In addition, when the cue is presented for a longer time and in more realistic settings, e.g. when a whole person is presented, 6-month-olds look more often and longer to the gazed at object than to a distractor (Senju & Csibra, 2008). Within interaction, 9-month-olds follow their parents' looks and pointing gesture to objects within their visual field (Flom, Deák, Phill, & Pick, 2004).

However, one alternative explanation for these results from very young infants is that infants divert their attention to the next visible object, but do not yet understand that the person is referring to or representing something as argued in Pätzold and Liszkowski (2019). Evidence confirming this assumption is that infants do not follow others' gazes or pointing gestures to targets behind them (Flom et al., 2004) or do not take into account whether the person can see the object, until their first birthday (Brooks & Meltzoff, 2002). Nine-month-olds also follow a person's direction of head turn when their eyes are closed, while 10-month-olds consider whether the person can see something (Brooks & Meltzoff, 2005). Following gazes or pointing gestures to hidden objects requires an understanding that the other person is referring to something. Thus, infants would have to show a referential expectation. In an interaction-based task, 12-month-olds expected a gazed-at object behind a barrier (Moll & Tomasello, 2004) or pointed to objects hidden in a box (Behne, Liszkowski, Carpenter, & Tomasello, 2012). In addition, violation of expatiation measure are used to measure referential expectation, indicated by longer looking times (Csibra & Volein, 2008) or higher pupil dilation (Pätzold & Liszkowski, 2019) in situations where infants would have expected something else. When using pupil dilation, 12-month-olds but not 8-month-olds showed a surprise reaction when a box that was pointed at revealed to be empty (Pätzold & Liszkowski, 2019). However, one study was able to find surprise with 8-month-old infants when using looking-time difference as a measure (Csibra & Volein, 2008).

Thus, it is not quite clear to what extent infants understand communicative actions in the first year of life. In the following section, I will illustrate the role of the method, for activating specific abilities and understanding the underlying mechanisms.

1.2 Methodological Approach

A variety of methods is used in infant research. Within interactive experiments or observational settings, infants' behavior is used as an indicator for abilities. When using infants' gaze behavior or neuropsychological methods, abilities may be captured before infants have the motoric abilities to demonstrate them behaviorally (Krogh-Jespersen & Woodward, 2016). Eye-gaze measure include violation of expectation paradigms, where infants' longer looking-time for unexpected events is used as indication for an understanding, as they would have expected something else. In addition, infants' looking direction can indicate infants' knowledge about future events (anticipatory looks) or interpretation of presented events (reactive looks). Abilities found with eye-gaze measures are sometimes referred to as implicit understanding, while infants interactive reactions are seen as indication for an explicit understanding (Krogh-Jespersen & Woodward, 2016). However, not only the measured variables differ between methods, but also the test-settings, which can have a great impact on eliciting abilities or not.

Infants' interactive behavior is mostly tested in very naturalistic and therefore communicative situations. Even though experiments are usually using a standardized protocol of behaviors, testing happens within an interactive setting, which mostly requires a reaction to the infant's actions. For looking-time studies, stimuli have to be very precise in time, and therefore situations are often videotaped. Thus, situations are less natural and may not be as salient or relevant for the infants (Krcmar, 2010). Sometimes theatre scenes are used, where an experimenter demonstrates the experimental conditions in real time to make it more naturalistic and more familiar to infants.

Although communicative cues can be included in video as well as in theatre scenes, there is no true recursive interaction, which infants are very sensitive to from early on (Rochat, Querido, & Striano, 1999). Different theories about the development of social cognitive abilities have divergent assumptions on the situations in which abilities are first present. Thus, for testing social cognitive abilities, it is crucial to consider the test environment of the method used. Systematic cross-sectional comparisons of different settings that vary in their communicative amount would give important insight into this theoretical controversy, but would also be an important validation for methods used in infant research.

Consequently, the question arises whether we are identifying similar abilities when using different methods. Examining relations between methods could potentially give us an answer to this question. Yet, only few studies dealt with this important methodical validation. Some cross-sectional studies found relations between different methods (Brune & Woodward, 2007; Krogh-Jespersen, Liberman, & Woodward, 2015). Brune and Woodward (2007) could show that infants who understand gaze as referential in looking-time tasks, spent more time in jointly attending to outside entities together with their parents (joint attention). In addition, infants who pointed by themselves understood pointing gestures as referential in looking-time tasks. However, both of these abilities (understanding gaze and pointing gestures as referential) were not related to point and gaze following to the object. Krogh-Jespersen et al. (2015) could show that processing speed in goal anticipation is relevant for VPT within interaction. Thus, there is positive and negative evidence for measuring related abilities with different methods and designs. Examining cross-sectional relations between abilities, measured with different methods and designs, would provide important information whether similar abilities are measured or not.

However, with a cross-sectional design it is not possible to examine directional relations between abilities within different methods, although it would be important to know if some early abilities are related to later skills. Most of the longitudinal studies deal with relations between early social cognitive abilities and later language acquisition or later theory of mind (Brooks & Meltzoff, 2008; Sodian et al., 2016). Thus, it remains unknown whether infants first establish an understanding outside of interaction and are then able to implement it within interaction or vice versa. While training studies would allow for causal inferences, longitudinal studies would give hints on directional relations.

To sum up, examining cross-sectional relations between abilities with different methods and designs and untangling directional relations via longitudinal studies would be a crucial empirical test for different theoretical assumptions and an important validation of different methods currently used in infant studies.

1.3 Development of Social Cognitive Abilities

Broadly, there are two distinct accounts regarding the development of social cognitive abilities. The first account states that infants have innate systems of knowledge, which are activated at specific time points (Baillargeon et al., 2016; Csibra, 2003). These theories can be broadly classified as nativist accounts. However, specific theories on the nature of these innate systems vary greatly within nativist account. In the following Section "1.3.1 Nativist View", I will briefly describe these different theories and present empirical evidence for each.

The second account posits that infants construct their understanding through experience (Moore & Barresi, 2017; Tomasello, 2019). Theories based on this proposition have been referred to as constructivist accounts. These theories differ in their conceptualization of the nature of experience potentially shaping infants' early understanding. In Section "1.3.2 Constructivist View" I will present the different theories briefly, with a focus on the most recent account: social constructivist view, highlighting human-specific second-person interaction.

Both views assume specific developmental pathways for social cognitive domains in contrast to domain-general abilities or improvement, which is a different theoretical approach (Heyes, 2016). However, general attentional development has to be taken into account when talking about domain-specific development. I will present some criticism and alternative interpretations of social cognitive abilities in terms of general attention improvement.

1.3.1 Nativist Views

Nativist accounts assume evolutionary adapted system-based abilities for specific domains, which equip infants with limited knowledge for representing objects, actions, numbers, space and in addition social partners (Spelke & Kinzler, 2007). I will focus on social-cognitive systems or domains, which are referred to as teleological and referential understanding (Csibra, 2003), psychological reasoning (Baillargeon et al., 2016), or implicit belief tracking within a two-system theory (Apperly, 2013). Although these theories refer to similar systems, they differ concerning the amount of understanding that is required. Aside from these differences, nativist theories have the common assumption that infants' social cognitive abilities are activated at a specific time point very early in life without the need of experience. Experience is seen as crucial to generalize understanding to a wider range of complex situations, however, it has not been seen as sufficient for basic abilities (Luo & Baillargeon, 2010). In the following, I will illustrate how different theoretical approaches describe these basic abilities within the nativist view.

1.3.1.1 Teleological and Referential Understanding

Csibra (2003) suggests that infants' understanding of humans is based on two separate, independent systems in the first year of life. One is a teleological system for understanding instrumental actions and the other is a referential system for recognizing communicative actions. Abilities within both systems are not necessarily mentalistic since infants do not have to ascribe the agents' underlying mental states, like intention, desire, perspectives etc. According to the teleological system, infants use behavioral cues to interpret agents' actions as goal directed and expect them to use the most efficient way to reach a goal. For example, 6.5-month-olds (Csibra, 2008), but more reliably 9-month-olds (Csibra, Gergely, Bíró, Koos, & Brockbank, 1999) attribute the goal of approaching an object to an agent, when the agent walks around or jumps over a barrier. When the barrier is not there in the test trial, infants are surprised when the agent uses the same path as before instead of approaching the goal directly on a straight path. Interpreting actions as goal directed is not restricted to humans or animals, but to agents or objects who show properties of freedom and choice in their actions (Csibra & Gergely, 2013). Thus, according to this theoretical approach, infants base their interpretation of goal-directedness on behavioral cues, by considering situations and constrains, and not on underlying goals and motivations.

Within the referential communicative system, it is assumed that infants are born with the predisposition to recognize communicative and ostensive cues, such as eye contact, infant-directed speech or contingent responses, and use these cues for human-specific cultural learning (Csibra & Gergely, 2009). When infants are exposed to direct gaze and infant-directed speech, the authors suggest that infants expect to learn something relevant. Thus, when an adult is then referring to something, by looking, pointing or naming an object, infants are able to connect this referent to the person (Gliga & Csibra, 2009). That is why 4- and 6-month-olds have shown to follow others' gazes when they are preceded by eye contact or infant-directed speech (Farroni et al., 2003; Marno et al., 2015; Senju & Csibra, 2008). Furthermore, 8-month-olds have demonstrated referential expectation when someone is gazing behind a barrier, so infants have to imagine the intended referent (Csibra & Volein, 2008). However, it is not assumed that this referential expectation has to be mentalistic during the first year of life; thus, infants do not necessarily have to represent what the other person represents or understands the communicative intentions, but only search for a relevant referent (Csibra, 2003).

Accordingly, different cues elicit either a teleological or a communicative action interpretation (Csibra, 2003). Seeing a self-propelled agent in biological motion, adjusting his behavior if needed leads infants to interpret the outside goal of the action, without the need to understand the underlying goal (Csibra & Gergely, 2013). Communicative cues elicit a referential expectation, without needing to represent what the other person sees (Csibra, 2003). Yoon, Johnson, and Csibra (2008) managed to show that infants remember different features of an object when a person is communicatively looking and pointing to the object, whereas they did not, when the person tries to reach for the object without any communicative cues. These differences are seen as evidence for two separate and independently working action interpretation systems. However, in a recent study these results could not be replicated (Silverstein, Gliga, Westermann, & Parise, 2019). Although a mentalistic action interpretation does not seem necessary for teleological and referential action understanding in the first year of life (Csibra, 2003), infants seem to be able to represent the mental state of others at this age (Kovács, Téglás, & Endress, 2010; Luo & Johnson, 2009).

1.3.1.2 Psychological Reasoning

Baillargeon et al. (2016) assume that infants interpret human actions in a mentalistic way from early on. Thus, infants do not interpret others' actions on a purely behavioral level, but consider motivational and epistemic states (e.g. seeing) as well. For example, when a person reaches for one of two objects several times, infants interpret this action as goal-directed and as a preference only when the person can see both objects (Woodward, 1998). When the person does not have visual access to one of the two objects, although the infant does, 6-month-olds do not interpret the adult's behavior as an expression of preference (Luo & Johnson, 2009). Although most of the studies are based on infants around the time of their first birthday (Luo & Baillargeon, 2007; Sodian, Thoermer, & Metz, 2007), mentalistic action interpretation is ascribed to younger infants as well (Luo & Johnson, 2009). Taking into account what others see or are referring to is not seen as separate from understanding instrumental actions (Luo & Baillargeon, 2007).

Whether abilities are witnessed or not does not seem to be influenced by communicative settings. Baillargeon et al. (2016) main explanation for not finding abilities in younger infants is that younger infants do not have enough knowledge about these specific actions yet or that insufficient information is provided in specific situations. Therefore, experience plays a role when infants learn about more complex and unfamiliar actions or preferences for specific persons, but not for understanding motivational and epistemic states in general. The main measure used to capture these early mentalistic skills is looking-time, with the expectation that infants possess observational abilities long before they acquire motor or language skills to show competencies within an interactive situation (Luo & Baillargeon, 2010). With looking-time measures, infants are also found to be able to track others' false belief (Kovács et al., 2010; Onishi & Baillargeon, 2005). With language dependent task this has been found only at 4 years of age (Wellman, Cross, & Watson, 2001). Abilities based on looking behavior are referred to as implicit, intuitive knowledge, which infants are not aware of. Based on this knowledge they can consciously decide and react flexibly, even to very complex situations. It is assumed, that this implicit understanding gradually develops into an explicit understanding through developmental time (Baillargeon et al., 2016).

1.3.1.3 Two-system Theory

The two-system theory also assumes an implicit ability of belief tracking (Butterfill & Apperly, 2013), but not in a full-blown and flexible way as suggested by Baillargeon et al. (2016). Within this theory, implicit abilities are assumed to be very efficient, but also inflexible and limited. Referred to as minimal mind-reading, this implicit system is especially beneficial to track what others can see, but not flexible enough to track all variations of false belief understanding (Fizke, Butterfill, van de Loo, Reindl, & Rakoczy, 2017; Surtees, Samson, & Apperly, 2016). In addition, Apperly (2013) presents a possible different developmental pathway than Baillargeon et al. (2016). Apperly (2013) proposes that an implicit system is present from birth and may remain in place until adulthood. An explicit system develops in parallel, manifesting only later in development when language and executive functions are developing (Apperly, 2013). This is contrary to the idea that the implicit understanding is developing to an explicit understanding (Baillargeon et al., 2016).

Different evidence is seen as indication for a remaining implicit system from birth until adulthood. Surtees and Apperly (2012) find VPT of irrelevant bystanders already in 6-year-olds and similar limitations in infants and adults (Apperly & Butterfill, 2009). However, another possibility is that explicit abilities are getting automatized and an efficient implicit ability emerges through experience later in development, which is why efficient and inflexible VPT is found in 6-year-olds and adults as well (Apperly, 2013; Apperly & Butterfill, 2009; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010).

1.3.1.4 Summary

In summary, within nativist views, early social cognitive abilities can either comprise a minimalistic or full-blown understanding of others, but all theoretical approaches agree on predisposition for learning human-specific abilities and evolutionary adaptations that emerge through maturation, not experience. The exact time point of emergence is not defined specifically, but it is assumed that infants show abilities from very early on, finding abilities from as early as 4-6 months of age (Farroni et al., 2003; Luo & Johnson, 2009). Further, in most of these approaches, abilities are not assumed to be dependent on communicative situations. Infants track a person's perspective when the person is reaching for something, but never addresses the infant (Luo & Johnson, 2009), or even when a person is only a bystander in the scene (Kovács et al., 2010; Surtees & Apperly, 2012). However, instead of domain-specific systems, a domain-general system could be the reason for early abilities (Grossmann, 2015; Heyes, 2016). For example it may be not the interpretation of the agent's perspective or mental state, but the front features that directs the attention to the side the way it does with an arrow (Heyes, 2014b) or other properties like novel colors or shapes (Heyes, 2014a). Thus, domain general abilities have to be taken into account and controlled for when assuming domain-specific abilities.

From the constructivist assumptions, about the time of emergence and situations in which abilities first present themselves differ greatly from nativist view. I will present different constructivist views in detail in the following section, but mainly focusing on social constructivism.

1.3.2 Constructivist Views

Theories associated with the constructivist approach have the common assumption that infants construct their understanding through experience (e.g. Vygotsky, 1978). The kind of experience seen as crucial differs between theories. Some theoretical views highlight first person experience, for example showing an understanding of actions as goal directed when infants start to execute goal-directed actions by themselves (Woodward, 2009). Other theoretical approaches see third person observation as crucial to build and adapt theories about other people (Gopnik & Wellman, 2012). However, infants are embedded in interaction most of the time and rarely simply observe. Social interaction with someone else (dyadic) and focusing on outside entities together (triadic) may connect both views and be an important setting for learning about others (Liszkowski, 2018; Moore & Barresi, 2017; Tomasello, 2019). In the following, I will focus on this social constructivist view by presenting theoretical assumptions and empirical evidence.

Moore and Barresi (2017) argue that human-specific dyadic and triadic interaction forms an understanding of the relations between an agent and an object. Infants experience emotional and contingent interaction from the moment of their birth. In the second half of the first year, they start to focus on outside entities together with their caregivers. The authors highlight special experience that is only apparent in second person interaction, which makes learning about others possible. Infants experience self-directedness, emotional engagement and they focus on a common goal with another person. Their interaction is contingent in time and content. Some features are similar to Csibra's (2003) communicative cues. However, Moore and Barresi (2017) do not assume an activated system, instead they argue that infants are able to construct their understanding embedded in these specific situations and bridge the gap between first person experience and third person observation. The understanding of others' actions is getting mentalistic over time.

Tomasello (2019) presents a similar account, assuming the importance of species-specific interaction. He argues that infants in deprivation would not develop abilities such as perspective-taking, although specific teaching or mirroring is not seen as necessary. Infants learn through everyday interaction and parents seem to adapt their behavior based on infants' abilities throughout (Liszkowski, 2018). However, Tomasello (2019) also highlights the evolutionary adaption and maturational component of specific abilities. He assumes two evolutionarily adapted pathways: One pathway for action understanding similar to apes and one path with a human-specific motivation to share psychological states. From early on, infants have an understanding of agency and actions in familiar situations. At the same time, humans show a specific ability for dyadic emotion sharing and protoconversations (Tomasello et al., 2005). These two pathways are expected to merge in the first year of life, around 9 months of age, and build the basis of human social cognition. Infants start to share attention and emotion with others toward outside entities and show joint attention for the first time. From this point on, infants' understanding changes fundamentally and Tomasello (2019) refers to this as the '9 month revolution'.

Within joint attention, both parties focus on an object together, and share their attention and emotion elicited by this object. Even when it is not explicit knowledge yet, they are aware that they are both attending to the same thing and sharing the experience (Siposova & Carpenter, 2019). However, first they have to align their perspectives, which happens through communicative gestures. This recursive experience is seen as crucial to enhances the understanding of perspectives (Tomasello, 2019). Parents initially start to follow the attentional frames of the infant, which seems to be influential for the development of infants' abiltiy to directing others' attention, e.g. in form of index finger pointing (Ger, Altınok, Liszkowski, & Küntay, 2018). Thus, after following others attentions, infants start to coordinate joint attentional sequences around their first birthday (Bakeman & Adamson, 1984). Although gaze following may be manifested as a maturational component in the beginning (Tomasello, 2019), it improves over time by first sharing attention to nearby objects, followed by detecting distal objects and finally by explicitly directing attention (Carpenter et al., 1998). This improvement seems to develop through joint interactional episods.

As empirical evidence, Moll and Tomasello (2007); Tomasello and Haberl (2003) could show that 12- and 14-months old infants remember an adult's attention and perspective when they have shared their attention on an object within joint attention. In contrast, infants were not able to understand the perspective when the adult was interacting with another person (Moll, Carpenter, & Tomasello, 2007), experienced the object alone, or was only on-looking from a distance (Moll & Tomasello, 2007). Only by the age of 18 months, infants were able to draw inferences from others' attention, when the adult interacted with the object individually. However, they were still not able to do so when the adult was only observing the object from a distance (Moll & Tomasello, 2007). Thus, infants seem to need experience within social interaction first to understand perspectives, which is also referred to as social perspective-taking (Moll & Kadipasaoglu, 2013). The dual-level structure of sharing and individuality within joint attention is seen as a special humans ability, so-called shared intentionality (Tomasello, 2019).

From 9 months on, infants' interpretation of instrumental actions changes as well. Throughout the first six months, infants only recognize others as agents and anticipate actions in familiar situations. From six months on they start to understand that actions are driven by underlying goals and humans try to pursue these goals (Tomasello et al., 2005). Infants start to distinguish between accident and intention and to identify if an action was successful. To interpret actions this way, they have to take into account what a person can see. Thus, contrary to Csibra (2003), Tomasello et al. (2005) do not assume referential understanding and action understanding to be separate processes. Instead, Tomasello et al. (2005) advocate for an integrated understanding of others as goal directed acting and perceiving organisms. Humans' special motivation to share psychological states make human cognition fundamentally different, in the way that humans want their interaction partners to understand their intentions (Tomasello, 2019). It is assumed that infants need to demonstrate some intention understanding to be able to take part in joint attention. However, their intention understanding skills transforms when joint attention is possible. Brandone, Stout, and Moty (2019) show that joint attention episodes predict later action understanding, which is demonstrated by the ability to anticipate unsuccessful reaching actions. In addition, Colonnesi, Rieffe, Koops, and Perucchini (2008) find relations between intention understanding and point comprehension and production in the beginning of the second year of life.

In sum, from a social constructivist point of view, infants develop their understanding of others through interacting with them (Moore & Barresi, 2017). A human-specific motivation and a sensitivity to social cues seems to be responsible for the development of these human-specific abilities (Tomasello, 2019). Thus, it is assumed that infants first show abilities within communicative and interactive situations and that there is a gradual development of abilities only later on in development. According to Tomasello et al. (2005) abilities are expected only after 9 months of age, when both pathways (action understanding and motivation to share psychological states) merge and infants start to interact triadically. However, again, domain general abilities can play a role in the development of specific abilities, e.g. infants are getting better in

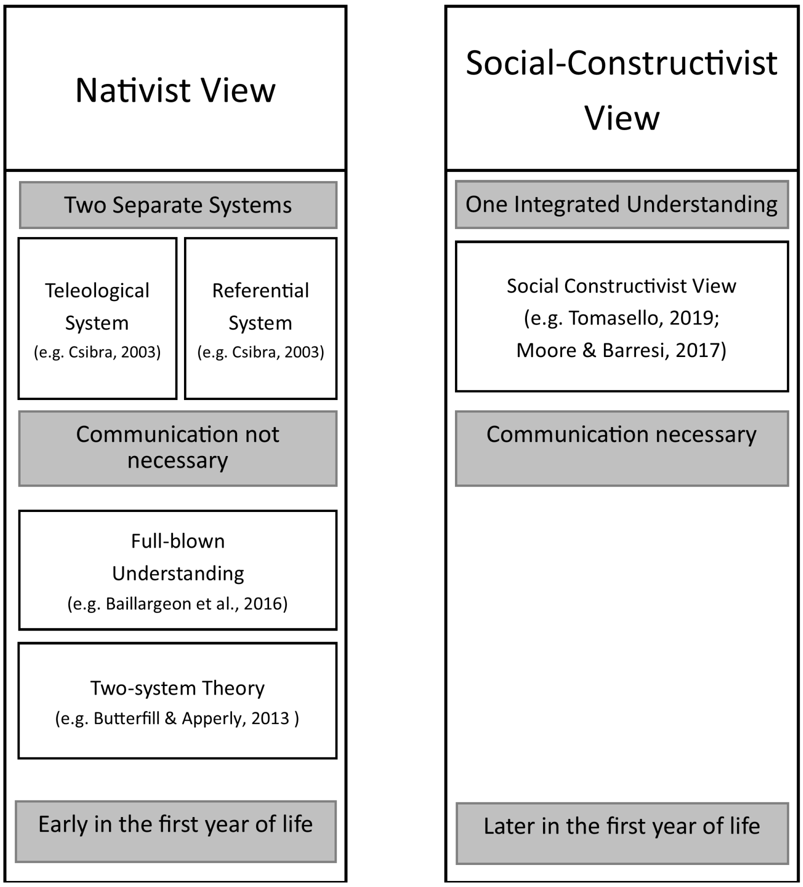

Figure 1.1. Schematic overview of the summarized theoretical approaches and their assumptions

orienting their attention in general (Del Bianco, Falck-Ytter, Thorup, & Gredebäck, 2019). Thus, for this possibility has to be controlled when examining relations within domain-specific abilities.

1.4 Summary and Open Questions

To date, a wide range of studies has examined the emergence of different social cognitive abilities. However, findings differ between studies and provide empirical evidence for different theoretical approaches. Some find abilities first in communicative settings

(Moll & Kadipasaoglu, 2013), others find it in non-communicative or observational settings from early on (Luo & Johnson, 2009). A mentalistic understanding is sometimes found in early months (Luo & Johnson, 2009), and sometimes only later on in the first year of life (Pätzold & Liszkowski, 2019). Different skills are either seen a separated and independent abilities (Csibra, 2003) or as an integrated understanding of others (Tomasello et al., 2005). Thus, systematic cross-sectional comparisons is an important approach to test which settings and situations help infants to show abilities. In addition, measuring diverse skills in a longitudinal design first helps to identify the time point of stability for specific skills and second to examine developmental relations between different abilities. Further, using different methods to measure social cognitive abilities is an important methodological validation. Shedding light onto the development of these early abilities is crucial for understanding the ontogeny of uniquely human social cognition and their role as important precursors for complex social interactional abilities later on in development. Thus, it is the aim of this thesis to capture early social cognitive abilities in the first year of life in detail, focusing on three main questions.

Firstly, it is important to consider the situations in which infants take into account another person's perspective, and therefore show a referential expectation. Social constructivist approaches highlighting second-person interaction assume perspective-taking abilities first within joint attention (Moll & Tomasello, 2007). Later, through experience within interaction, they also show perspective-taking in less communicative situations (Moll et al., 2007). One theory proposed by Csibra (2003), which is in line with the nativist account also suggests referential expectation when communicative cues are present. In contrast, most theories associated with nativist accounts expect to find perspective-taking even when infants are not embedded within interaction, that is, when they observe others for example in video or theatre scenes without any communicative cue (Luo & Johnson, 2009). The two-system theory assumes that implicit belief tracking is present even for bystanders (Samson et al., 2010). Thus, finding abilities in communicative settings first supports the social constructivist view. Abilities in non-communicative settings arising early on would provide evidence for nativist accounts. I will discuss this issue in Chapter 2 (Study 1) and Chapter 4 (Study 3).

The second question focuses on the age of emergence of stable abilities. Nativist accounts expect social cognitive abilities in early development (Baillargeon et al., 2016), while the social constructivist view assumes that abilities emerge gradually only later on in development (Tomasello et al., 2005), as they have to be constructed. Tomasello (2019), for example, suggests a sophisticated social understanding only after 9 months of age, when infants start to focus on outside entities with their interaction partners. There is some empirical evidence for both approaches. For instrumental action interpretation, infants can anticipate the target of an unfulfilled reaching action only after 10 months of age (Brandone et al., 2014) and anticipate the goal in contrast to the path only with 11 months (Cannon & Woodward, 2012). However, as mentioned above, 6-month-olds also seem to be able to interpret reaching actions by taking into account others' perspectives (Luo & Johnson, 2009) or can anticipate hidden goals (Applin & Kibbe, 2019). For communicative understanding, a referential understanding of others' attention and communicative intention has mostly been found at the end of the first year (Pätzold & Liszkowski, 2019). However, Csibra and Volein (2008) did not find a significant difference in referential expectation between eight- and 12-month-olds, arguing for an early ability. Testing 8-month-olds individually in a violation of expectation paradigm suitable for such young infants would enable us to examine if perspective-taking is present early on in the first year of life. I will come back to this issue in Chapter 3 (Study 2). With a longitudinal approach, it is possible to pinpoint when abilities show stability between measuring time points. This is an important addition to the individual cross-sectional evidence at different months. I will present a study using a longitudinal design in Chapter 4 (Study 3).

The third research question examines the developmental relations between instrumental and communicative action understanding. Csibra (2003) presents them as two separate and independent systems in the first year of life and assumes an integration only in the second year of life. However, the exact time point of this integration has not been empirically verified. Since the social constructivist view assumes that an understanding is constructed within interaction, especially in joint attention, consequently as soon as infants are able to follow others' attention, their understanding of their actions would change as well (Siposova & Carpenter, 2019). At 9 months, when infants start to exhibit joint attention, they also start to understand that humans' actions are driven by the underlying goals and perception. Simultaneously, Tomasello (2019) assumes that infants need some basic understanding of others as intentional beings to engage in joint attention. Thus, when these two pathways are getting integrated at 9 months of age, both abilities seem to influence each other bidirectionally. A critical test for an integrated understanding would be whether both abilities are related during their emergence. In addition, using different methods and examining relations within subjects is an important validation of methods used in infant studies in general. I will present a Study focusing on this topic in Chapter 4 (Study 3).

1.4.1 Work Plan and Hypothesis

Study 1 refers to the first question – in which situations do infants show referential expectation. Here, I used a cueing paradigm with different cues varying in their communicative amount. I measured the time to first fixation to the object after the cue disappeared. In addition, I captured referential expectation, with a violation-of-expectation measure indicated by a longer looking-time when the cued side revealed itself to be empty. I compared a communicative pointing cue with a non-communicative cue, where the person was simply directed to one side, similar to a bystander in other VPT tasks (Kovács et al., 2010; Samson et al., 2010). I tested 14-month-olds, as they are known to be able to represent what agents see in communicative settings. If VPT does not depend on communicative situations, as minimal mindreading (Apperly, 2013) and psychological reasoning (Baillargeon et al., 2016) assume, I expect to find referential expectation for the non-communicative cue as well. In another experiment, I added facial expression to the non-communicative cue to improve salience. If infants need more information (Luo & Baillargeon, 2010), but cues do not have to be communicative, 14-month-olds might show referential expectation in cases where it is more obvious, that the person is referring to something. In a third experiment, I tested 36-month-olds, to test if children represent an agent's perspective in non-communicative cues at all. If referential expectation were only to be found in this age group, it would support the assumption, that infants automatize perspective-taking over time and experience, instead of having an innate VPT ability from birth.

In Study 2 + Study 3, I will address the second research question: when do infants start to show a referential understanding of others' actions? In Study 2, I tested 8-month-olds in the cueing paradigm as presented above. If infants show referential expectation for communicative pointing gestures or instrumental reaching actions at this young age, it would be in line with nativist theories assuming a mentalistic understanding from early on (Baillargeon et al., 2016). If infants do not show referential expectation at this young age, this would support constructivist view assuming perspective-taking only after 9 months of age (Tomasello et al., 2005). In Study 3, I tested different social-cognitive abilities in a longitudinal design to capture when abilities emerge to be stable over time and to be able to strengthen findings from my cross-sectional studies.

In Study 3, I will focus on relations between communicative understanding and instrumental action understanding. Therefore, I captured both abilities in a longitudinal design using different methods. If there are developmental relations between both skills this would be in line with developing an integrated understanding of others as goal-directed, acting and perceiving organisms (Tomasello et al., 2005). As relations might be mediated by general attention improvement, I control for a disengagement ability in non-social tasks. Finding relations between action understanding and communicative understanding would not fit with the assumption that these systems emerge independently (Csibra, 2003). In addition, I considered the different methods I used to capture social cognitive abilities. Adding to the first research question, I compared abilities in very communicative settings (interacting based tasks) to less communicate settings (video scenes). I further examined if there are relations between abilities when using different methods, as an indication of measuring similar abilities with different methods.

Chapter 2

Inferring hidden objects from still and communicative onlookers at 14- and 36-months of age[1]

RESEARCH HIGHLIGHTS

- 14-month-olds expect objects behind an occluder following a centrally presented adult's communicative pointing; but not her still onlooking.
- Adding facial expression on the still onlooker do not help 14-month-olds
- 36-month-olds expect occluded objects also when the adult is a still onlooker.
- Visual-perspective-taking emerges gradually through communication and becomes automatized with age.

Acknowledgements: We want to thank all families who participated in these studies. In particular, we want to thank Dr. Wiebke Pätzold for her conceptual input. We thank Nicola Ballhausen for her help in creating the stimuli and running some of the experiments. We are grateful to Jula Brüning-Wessels, Flora Kelmendi, Jana Klose, and Lydia Rihm for their assistance with data collection.

[1]This study is part of a paper currently under revision entitled "Inferring hidden objects from still and communicative onlookers at 8-, 14- and 36-months of age" (Jartó & Liszkowski, under revision) in the Journal of Experimental Child Psychology.

2.1 Abstract

Adults seem to be influenced by others' perspectives even when they are not relevant for them. This automatic visual-perspective-taking (VPT) is assumed to be present from early on as implicit belief tracking. However, VPT abilities around infants' first birthday are mostly tested in very communicative settings and not for non-ostensive, non-relevant bystander. The current study investigated across three eye-tracking experiments under which conditions toddlers are able to infer the presence of an object behind an occluder as indication for VPT. A centrally presented person was either pointing or still looking behind one of two occluders. 14-month-olds expected a hidden object behind an occluder, if the onlooker had communicatively pointed to it (Exp.1), as revealed by shorter latency to target detection in congruent trials and longer dwell-times to the empty side in incongruent trials. This was not the case when the onlooker was only still oriented to one side (Exp.1). Adding emotional facial expression to the still onlooker (Exp.2) did not help produce the effect. However, at 36 months of age (Exp.3), children showed the effect even when the onlooker remained still. Findings reveal that automatic perspective-taking develops after communicative perspective-taking.

Keywords: Visual-perspective-taking; referential expectation; social cognition; eye-tracking

2.2 Introduction

Everyday social interaction often require taking other persons' perspectives, i.e. perceiving a situation from another person's point of view. A basic requirement of perspective taking is to understand that another person perceives something through her eyes and forms a mental representation (mentalizing). However, there is an ongoing debate about (1) to what extent findings on visual-perspective-taking (VPT) involve mental attributions (Furlanetto, Becchio, Samson, & Apperly, 2016) opposed to lower level perceptual spatial mechanisms (Heyes, 2014b), and (2) to what extent it is instigated by communicative social relevance (Moll & Kadipasaoglu, 2013) opposed to automatic processes (Samson et al., 2010). From a developmental perspective, the question is whether there is a change from simpler perceptual to more complex cognitive processes, and whether skills are automatic from the beginning or derive from social communicative relevance and become automatized only later through development (Apperly, 2013). The current study therefore investigated, what kind of cues and behaviors may instigate infants to expect cognitively that a person sees something indicated by a referential object expectation, as opposed to simply orient spatially to visible targets to which a person orients.

In adult research, evidence comes from VPT tasks and cueing paradigms. In VPT tasks, the line-of-sight of an avatar interferes when participants have to judge the number of dots visible to them. That is, when the avatar sees a different number of dots, reaction time is reduced and the error rate is higher than when the avatar sees the same number of dots (Samson et al., 2010). Apparently, the mere presence of an avatar is sufficient to spontaneously adopt their perspective. However, it is discussed whether simple attention orienting and general purpose mechanisms (Cole, Smith, & Atkinson, 2015; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014) or higher mentalizing abilities underlie this so-called altercentric effect (Ferguson, Apperly, & Cane, 2017; Furlanetto et al., 2016). In cueing paradigms, a face or a person is presented as central cue for a specific time before a target appears either on the cued congruent side or the non-cued incongruent side (Gardner, Hull, Taylor, & Edmonds, 2018; Okamoto-Barth & Kawai, 2006). Adults detect congruent targets faster than incongruent targets (cueing effect), but it is crucial that the attention is drawn to the cue before the target appears (stimulus onset asynchrony [SOA]; (Bukowski, Hietanen, & Samson, 2015)). Short SOA (100 ms) seem to elicit reflexive bottom-up effects (involuntary cue-driven attention direction). The longer the SOA gets, the more a reflective voluntary top-down process (by interpreting the cue) is recruited (Gardner et al., 2018).

In infant research, cueing paradigms reveal abilities from early on. When presenting a face (Farroni et al., 2003), a pointing hand (Rohlfing, Longo, & Bertenthal, 2012) or a grasping hand (Daum & Gredebäck, 2011) infants lead their attention to the cued side. At this age, cues have to be very communicative or include some movement to elicit a cueing effect. As in adult research, the SOA plays a crucial role. A reflexive bottom up cueing effect with a very short SOA (100 ms) seems to be present from early on (4-6 months). A reflective top down cueing effect (measured with a SOA longer than 300 ms) is present at 12 months of age, if the cue is communicative and multimodal, but not at 10 months (Daum, Ulber, & Gredebäck, 2013). However, from this line of research it remains somewhat unclear whether infants form referential cognitive expectations as opposed to reflexively orient, or covertly attend, in reaction to a directional moving cue (Rohlfing et al., 2012).

Other research has employed occlusion paradigms, which go beyond perception and find that after their first birthdays, infants understand seeing as a mental act and expect a referent when someone is looking at or pointing to something (e.g. Moll & Tomasello, 2004). For example, infants follow a gaze less, when the view of an agent is blocked by either closed eyes (12-month-olds), a blindfold (14 and 18 months; (Brooks & Meltzoff, 2002) or barriers (18 months; (Butler, Caron, & Brooks, 2000), and they follow expressive communicative gaze behind barriers (Moll & Tomasello, 2004). Further, by 12 months, infants search for an object at its hiding location when it is disambiguated through a communicative pointing gesture (Behne et al., 2012) and show surprise when the object is not at the indicated location, as revealed by longer looking (Csibra & Volein, 2008) and larger pupil dilation (Pätzold & Liszkowski, 2019) in violation-of-expectation (VoE) paradigms. Thus, from their first birthday on, infants seem to take into account what others can or cannot see, however, it remains an open question, what properties the cues have to have to elicit VPT.

While the adult evidence suggests that a mere onlooker is sufficient to elicit VPT, the theory of natural pedagogy assumes that infants have to be addressed ostensively to trigger a referential expectation (Csibra, 2010; Csibra & Gergely, 2009). Similarly, shared intentionality theory argues that infants' understanding of others' perspectives emerges within social interaction. As supportive evidence, Moll and

Tomasello (2007) found that 14-month-old infants have to be engaged in joint action to represent the perspective of an interaction partner. Further, 9-month-olds follow others' gazes more often when gaze is accompanied by pointing and verbalizing (Flom et al., 2004), and 12-month-olds need communicative speech to elicit a cueing effect with a pointing hand (Daum et al., 2013).

However, there is also evidence for perspective taking abilities in less communicative situations. von Hofsten, Dahlström, and Fredriksson (2005) found that 12-month-olds orient their attention to a cued side on a static picture of a person looking or pointing to an object, without communicative ostension or joint engagement at all, suggesting a rather automatic process. Similarly, in automatic VPT tasks with 7-month-old infants, a neutral passive bystander seems to be sufficient to yield an altercentric perspective interference (Kampis, Parise, Csibra, & Kovács, 2015; Kovács et al., 2010; Southgate & Vernetti, 2014). Further, VoE looking-time studies suggest that without being explicitly addressed, 6-month-olds infants represent another person's perspective when they observe the person act on one of two objects of which only one is visible to the person (Luo & Johnson, 2009).

Against this background, the current study investigated across three experiments at which cues infants and toddlers show cognitive referential expectations of occluded objects. We implemented an eye-tracking based cueing paradigm with a long SOA to test for endogenous reflective processes. In the videos, participants saw a person sitting in profile between two occluders, disappearing at test, and then, after a brief central fixation cross, the occluders opened to reveal an object either at a cued or uncued side. We compared the latency to look at the object in congruent and incongruent trials and the looking time to the empty location (dwell-time) as indications for referential expectation. We expected that infants show a shorter latency in congruent than incongruent trials; and we expected a longer looking time in incongruent trials than congruent trials if infants have a referential expectation. The paradigm rests on the logic that participants form a cognitive representation of the occluded object in expectation of seeing it once occlusion ceases (Behne et al., 2012; Csibra & Volein, 2008). Because the cue has ceased before the object is revealed (SOA), and the attention is reoriented to the center before the object is revealed, effects on the speed of orienting to the object are mediated by a cognitive representation. Dwell-time is a measure of cognitive expectation in infancy research such that longer dwell-time reveals the violation of an expectation (Csibra & Volein, 2008).

Across experiments, we manipulated the communicativeness of the cues that the person in the videos provided. The experiments followed up on each other, with Experiment 1 being the main experiment which addressed the main question whether infants can equally well represent another's perspective following a communicative cue or a still, directional cue (as e.g. in the adult visual-perspective-taking tasks). In Experiment 1, we compared a person pointing communicatively behind one of the two occluders (pointing condition) with a still person simply looking ahead in the direction of one of the two occluders (still condition). We tested 14-month-olds, because the literature reviewed above shows that they have referential expectations about occluded objects in response to communicative cues. However, we do not know if they show it in non-ostensive situations like adults do, or if the need communicative situations still at this age. In Experiment 2, we added facial still expressions of disgust/excitement to the person in the still condition to test whether the referential nature of emotions would instigate referential expectations about non-perceivable objects. In Experiment 3, we tested older toddlers (36-month-olds) to assess whether a still condition had an effect in an older sample (as, ultimately, adult studies would suggest). All experiments used the same general paradigm. However, each experiment is based on hypotheses and experimental manipulations in their own right, containing mild variations in timing or sampling rates where appropriate, therefore mandating separate analyses, and enabling directed hypothesis testing.

2.3 Experiment 1

Experiment 1 aimed at first, establishing cognitive referential expectations following a communicative cue in 14-month-olds, and second, to then test whether these referential expectations would also be present following a still, non-ostensive cue, as VPT paradigms using still onlookers would suggest. We employed a visual eye-tracking paradigm closely modeled after Csibra and Volein (2008). However, the actress never looked directly at the infant and was only seen in profile as in VPT paradigms.

If infants indeed have cognitive referential expectations, they should have a shorter latency to look at the object in the cued than non-cued location, as a measure of anticipation. In addition, they should look longer at the empty location when it had been cued than when it had not been cued, as a measure of a VoE. Given previous findings, this pattern should apply to the communicative cue condition. If infants' referential expectations were rather automatic as VPT tasks suggest, then the pattern should also hold for the non-ostensive still condition, in which the actress simply looked ahead to one of two sides.

2.3.1 Methods

2.3.1.1 Participants

Parents were contacted via birth-register in a metropolis in Western Europe. Our study is based on the study from Csibra and Volein (2008), where they found an effect size of $\eta_p^2 = .30$ for dwell-time difference in congruent and incongruent trials when testing 8-12 month-olds in a communicative condition. With a p-value of $p = .05$ it requires a minimum sample size of N = 18 infants to have an 95% chance of detecting significant effect in a repeated measurement analysis. We tested slightly more infants as we had no reference for the still condition and the final sample size was reduced according to exclusion criteria.

The final sample consisted of $N = 32$ infants (15 females) with a mean age of 14 months and 15 days, which ranged from 14;1 to 14;27. Additional 16 infants were invited but excluded due to fussiness ($n = 5$), bad calibration (more than 2 °deviance; $n = 10$) or recording problems ($n = 1$). We note that the calibration procedure ran apparently less smooth in the current compared to subsequent experiments but we cannot pinpoint this to any obvious factor(s) – at any rate, these drop-outs occurred before, and thus independent of the experiment itself. According to exclusion criteria (see Data Analysis and Reduction) sample size differs slightly in each analysis.

2.3.1.2 Set-up and Procedure

Families were individually invited to the laboratory at a time during the day, which they anticipated to be optimal for their infants (awake; not hungry; in a good mood). During a 10 minute warm-up in a welcome area, parents were informed about the procedure and data privacy. Eye-tracking experiments were approved by the local ethical committee and in accordance with the declaration of Helsinki for the protection of human participants. Then, they followed with their infant into the eye-tracking room. A Tobii x120 eye tracker (Tobii Technology, Stockholm, Sweden) was used to record infants' gaze with 60 Hz sampling rate. Stimuli were presented with Tobii Studio running on a computer with hardware components recommended by Tobii Eye-Tracking System (Intel Core 2.60G Hz CPU, 16GB RAM, Nvidia Graphic Card, SSD 500GB Hard Disk; Tobii System Recommendations for Tobii Studio).

Two screens were connected to the computer via DVI-D (18+1) and set to extended mode. One screen was not visible for the infant and used to monitor calibration and infants behavior during the session. The other screen was the presentation screen with 1920x1200px resolution (52 x 34 cm) and positioned above the eye-tracker. Refresh rate of the screen was 60 Hz and response time (time until the picture arrived) was 13.5 ms. The presentation screen was protruded through a black canvas, which was about 2.5 meters high and 1.5 meters in width and had triptych-like wings to the left and right side to minimize distraction for the infant. Above the screen was a web-camera to monitor the behavior of the infant during the session. Parents sat on a swivel chair in front of the screen with their infants on their laps at a distance of approximately 62-65 cm. They were told not to interact with the infant or direct the infant's attention to the monitor during the session and hold it in the same position. Parents wore opaque glasses during calibration and video presentation in order not to record parents' gaze and to keep them ignorant about the presented events to prevent any interference. A five point calibration with rotating balls and infant friendly music was used.

2.3.1.3 Stimulus and Design

Participants watched self-made video-based stimuli. Videos were extracted with 25 frames per second (fps) and as recommended for Tobii-Studio software converted to AVI-format with 15fps and video codec WMV1. One video (a congruent trial) was unintentionally not converted to 15 fps and presented with 25fps. Timing for this trial was adjusted (every event happened 120 ms earlier). Figure 2.1 illustrates the unfolding and the timing of the test stimuli events. A person was displayed in the middle of the screen in profile, sitting between two equidistant tables. During the recording of the stimuli, the person had been instructed to fixate a predefined spot on the table. In the process of editing the videos, two occluders were superimposed onto the tables to cover the predefined spot. They were animated to lower forward like a drawbridge with a door-like frame remaining. The video display was 21.7°vertical x 37.7°horizontal. The occluders measured 6.7x6.7°, and the distance between the occluders was 23.5°. On the screen, the person in profile measured 16.6°and her head about 4.8x4.8°. The distance between the middle of the head and the middle of the occluder was 17.1°. After the person disappeared a fixation cross (1.4x1.4°) appeared instead at the position where the person's head had previously been. A lego-duplo® object was superimposed onto the video, at the location of the predefined fixation spot, behind one of the two occluders, so that it was revealed inside the door-like frame when the occluders opened. There were eight different lego-duplo® objects, a different one for each of the eight trials.

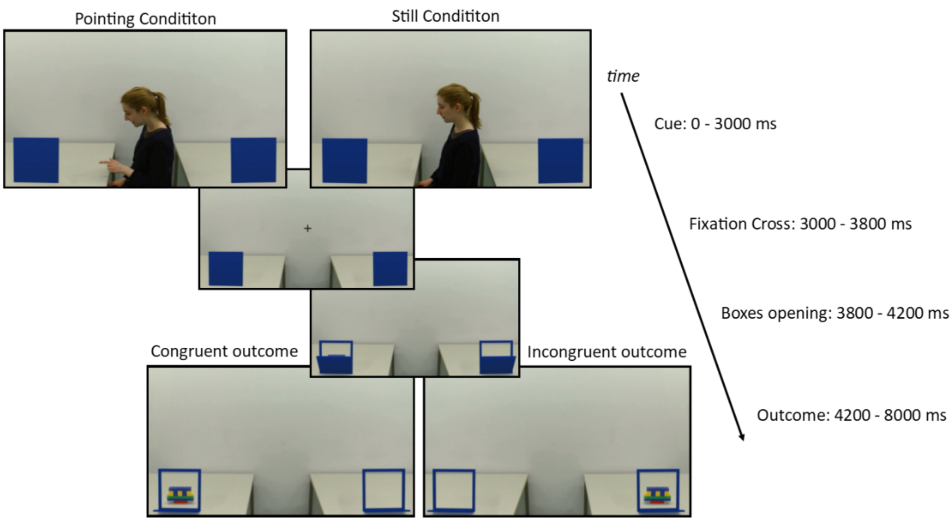

Figure 2.1. Sequence of the video stimuli for both conditions (pointing and still) used in Experiment 1.

There were two types of cues presented in two conditions - a pointing and a still condition. In the pointing condition, a person pointed behind one of the two occluders, starting from her lap and saying "Ah". The person was 10.5°wide at the position of the pointing hand. The distance between the end of the pointing finger and the middle of the pointed occluder was 8.6°. An object appeared either on the side of the cued occluder in front of the person (congruent trial) or on the side of the uncued occluder behind the person (incongruent trial). In the still condition, the person in profile remained still while staring behind the occluder.

Each session started with a 30 s animation video with music (Barbabapa) to increase infants' attention and motivation. Then two familiarization videos without a person were presented to familiarize infants with the opening of the occluders and the appearance of a toy behind one of the two. Object sides alternated. Then the two different conditions were presented blockwise and the presentation order was counter-balanced across subjects. Each condition included four congruent and four incongruent trials. Cueing side, object side and congruency was randomized across trials but never the same in more than two consecutive trials. Between the two condition blocks, another 30 s Barbapapa video was presented. Every trial started with a small moving attention-getter with short sound in the middle of the screen for 3 seconds to center the infant's attention before the beginning of a trial.

2.3.1.4 Data Analysis and Reduction

Our main dependent variables were latency to the object side and dwell-time on the empty side of the occluders. As additional information, we measured the looking behavior in the cueing phase. Therefore, we calculated the proprotion of trials in which infants looked longer to the cued occluder than to the not cued occluder; and vice versa. In addition, we calculated the proportion of trials in which infants did not look away at all from the person. We defined four square areas of interest (AOI). Two AOIs covered each occluder with a horizontal and vertical dimension of 7.7°, another AOI was positioned around the fixation cross with a dimension of 2.4x2.4°one was around the whole video 21.7x37.7°.

Latency was measured from fixation cross off-set until the first fixation in the AOI of the occluder which contained the object. A fixation was defined by Tobii-IVT-filter with following settings: Data were filled (interpolation) for smaller than 75 ms time windows. No noise reduction was used for data sampled with 60 Hz. A fixation has to have a velocity threshold below 30°/s measured in a 20 ms time window and a minimum duration of 60 ms. To identify the fixation in the correct AOI 67 ms had to be detected in a 100 ms time window. To exclude unnaturally high latency values caused by inattentiveness and distraction, we excluded latency measures when infants looked away from the screen after the fixation cross offset for more than 500 ms, because then the object was already visible and infants were clearly not on task to detect the object.

Dwell-time was measured to the empty side of the occluder as the sum of gaze registrations while the object was visible. The value was converted into milliseconds by multiplying it with the duration of one frame (1 s/60 Hz).

We employed conservative exclusion criteria to reduce noise. First, infants had to watch the cue for at least one third of its duration (i.e. 1000 ms) to make sure that infants attended to the communicative features of the cue and reflectively processed it, and that the processing time was comparable between conditions. Second, infants had to fixate the cross for a minimum of 60 ms to exclude that infants' gaze only swayed pass the cross. Latency and dwell-times were measured only after this requirement, to exclude fixations on the boxes from the cueing phase. Results indeed revealed no latency shorter than 200 ms. The following numbers of trials were excluded due to insufficient attention during the cueing phase ($n = 26$ trials of 10 infants in the pointing condition, $n = 26$ trials of 13 infants in the still condition), or inattention of the fixation cross ($n = 49$ trials of 12 infants in the pointing condition, $n = 40$ trials of 20 infants in the still condition). Infants had to have 50% of valid trials per congruency trial type to calculate a mean.

To assess overall effects, we submitted our dependent measures latency and dwell-time to 2 x 2 repeated-measures analysis of variance (ANOVA) with Congruency [congruent, incongruent] and Condition [pointing, still] as within-subject factors. To control for order an additional 2x2x2 ANOVA was run, with order [pointing-condition first, still condition first] as between subject factor. Our hypothesis, based on existing findings was that 14-month-olds do have referential object expectations and thus show latency and dwell-time effects. Of question was whether latency and dwell-time effects are also present in the still condition. Accordingly, we subsequently ran planned t-tests for each condition. The two main dependent measures are based on directed hypotheses such that shorter latencies (not longer) are predicted against the null model when comparing congruent and incongruent trials; and longer dwell-times (not shorter) are predicted against the null model when comparing incongruent and congruent trials. Accordingly, planned t-tests are reported one-tailed. To assess the effects on an individual level, we used binomial tests to compare the number of infants who had a faster mean latency in congruent than incongruent trials to the number of infants who showed the opposite effect; and the number of infants who had longer dwell-times to the empty side in incongruent than congruent trials with the number of infants who showed the opposite effect. We winzorized variables with extreme outliers (3 standard deviation above the mean) by changing values above the 95% percentile to a value within the 95% percentile (Reifman & Keyton, 2010) using an online winsorizing-calculator (Hemmerich, 2019). We measured ANOVAs even when some variables were not normal-distributed, because no extreme outliers were present after winsorizing.

2.3.2 Results

Infants' total dwell-time on the screen did not differ significantly between the conditions (Pointing: M = 6756.55, SD = 804.55; Still Condition: M = 6488.30, SD = 727.01; $t(27) = 1.45$, $p = .160$, $d_z = 0.35$). Infants were thus equally attentive across conditions.

2.3.2.1 Cueing Phase

A 2 x 2 Congruency [cued, uncued box] x Condition [still, pointing] repeated-measures ANOVA showed a significant main effect of Condition ($F(1,\ 28) = 25.99$, $p < .001$, $\eta_p^2 = .481$), and Congruency ($F(1,\ 28) = 7.07$, p = .013, $\eta_p^2 = .201$). The condition effect shows that infants disengaged more often from the person and looked to the boxes in the non-ostensive condition ($M = 18.2\%$, $SE = 2.6$) than in the communicative condition ($M = 4.6\%$, $SE = 1.1$). In both condition, infants looked more to the cued box

($M = 14.7\%$, $SE = 2.3$) than to the uncued box ($M = 8.2\%$, $SE = 1.5$). The Congruency x Condition interaction was not significant ($F(1, 28) < 0.01$, $p > .99$, $\eta_p^2 < .001$). Infants did not disengage from the person in most of the trials ($M_{pointing} = 90.7\%$, $SD = 11.5$; $Mstill = 64.6\%$, $SD = 28.0$).

2.3.2.2 Latency

Due to one outlier, latency was winsorized in incongruent trials in both conditions. There were no significant effect for order. A 2 x 2 (Congruency [congruent, incongruent] x Condition [still, pointing]) repeated-measures ANOVA showed no significant main effect of Congruency ($F(1, 18) = 1.06$ $p = .316$, $\eta_p^2 = .056$), and no significance for Condition ($F(1, 18) = 1.37$, $p = .257, \eta_p^2 = .071$). The Congruency x Condition interaction was significant ($F(1, 18) = 5.12$, $p = .036$, $\eta_p^2 = .221$).

The first column of Figure 2.3 displays the latencies for the t-test of the current Experiment 1. Planned paired t-tests for each condition according to our hypothesis revealed that in the pointing condition, infants were significantly faster to detect the object in congruent than incongruent trials, $t(22) = 2.37$, $p = .014$ (one-tailed) $d_z = 0.65$. No such effect was found in the still condition, $t(19) = 0.08$, $p = .470$ (one-tailed), $d_z = 0.02$. The pattern of statistical findings held when including the same number of infants from the omnibus ANOVA who had sufficient trial numbers in both conditions.

In the pointing condition, 16 out of 23 infants looked faster to the congruent object than to the incongruent object, binomial test, $p = .047$ (one-tailed). In the still condition there was no difference in the number of infants (10 out of 20, p = .500 (one-tailed)).

2.3.2.3 Dwell-time

The 2 x 2 x 2 (Congruency [congruent, incongruent]) x Condition [pointing, still]) x Order [pointing-first, still-first]) mixed-design ANOVA with order as between-subject factor and all others as within-subject factor showed no main effects (Congruency: $F(1, 21) = 0.79$, $p = .383$, $\eta_p^2 = .036$); Condition: $F(1, 21) = 0.20$, $p = .658$, $\eta_p^2 = .010$). The interaction between Congruency and Condition was not significant, $F(1, 21) = 1.89$, $p = .184$, $\eta_p^2 = .082$). Of all other interactions only Condition x Order was significant ($F(1, 21) = 6.13$, $p = .022$, $\eta_p^2 = .226$); pointing first: $M_{pointing}$= 395.64, SE = 68.29, M_{still}=264.70, SE = 62.54; still first: M_{still}= 339.34, SE = 50.15, $M_{pointing}$= 248.61). This seems due to fussiness and less total dwell-time in the condition presented in the second block.

The first column of Figure 2.4 displays the mean dwell-time to the empty location for the t-test of the current Experiment 1. To test directly for the predicted effect, a planned paired t-test revealed that infants in the pointing condition looked significantly longer to the empty side of the occluder in incongruent than congruent trials, $t(23) = 1.91$, $p = .035$ (one-tailed), $d_z = 0.42$. No such effect was found in the still condition, $t(23) = 0.38$, $p = .353$ (one-tailed), $d_z = 0.10$. The pattern of statistical findings held when including the same number of infants from the omnibus ANOVA who had sufficient trial numbers in both conditions.

In the pointing condition, 18 out of 24 infants looked longer to the empty side in the incongruent videos than in the congruent videos, binomial-test (p = .012 (one-tailed)). In the still condition the difference was not significant (10 out of 24 infants, binomial-test p = .271 (one-tailed)).

2.3.3 Discussion

Findings from the communicative condition of Experiment 1 demonstrate that by 14 months of age, infants have cognitive referential expectations following a communicative cue, which is consistent with previous literature. In extension, these findings show that direct gaze towards the participant is not necessary, although we note that the condition remained clearly ostensive and communicative through the actress' pointing and voice display. A remaining question is when in development these expectations emerge. I will address this question in Chapter 3 (Study 2), where we tested 8-month-old infants in the communicative pointing condition.

In contrast, findings from the non-ostensive still condition revealed that 14-month-olds do not have such referential expectations when the actress is simply directed to one side, as the avatar in VPT tasks. This finding adds to the literature, which has suggested that infants first adopt others' visual perspective in communicative settings and initially do not infer the perspective of an onlooker (Moll & Tomasello, 2007). Thus, seeing a person only looking into a direction (onlooker) is not sufficient for 14-month-olds to infer a hidden referent. While static cues (von Hofsten et al.,

2005) or non-ostensive gaze cues (Collicott, Collins, & Moore, 2009) may elicit simple gaze following, as also evident from our analyses of the cueing phase before the test outcome, they do not seem to instigate referential expectations in 14-month-olds. While automatic VPT findings with infants may appear not easily compatible with this view, an alternative interpretation of those findings is that an onlooker only enhances infants' perspective during the cue, but does not instigate the expectation after the cue that the onlooker has seen something.

However, it is also possible that it was not a lack of communication or motion that hindered infants to form referential expectations, but that the cue of the still, expressionless face was simply not salient enough. In Experiment 2, we explored this possibility and added emotional expressions to the still face to test whether this would trigger referential expectations despite the absence of communication or further movements.

Another question, which follows from the pattern of findings in Experiment 1 is whether a still onlooker may indeed at all yield referential expectations in the observer, as suggested by adult findings. One possibility is that an understanding of others' visual perspectives becomes automatized in development such that older children will infer a hidden object in the still condition of our paradigm. We investigate that possibility in Experiment 3, in which we tested 3-year-olds in the non-ostensive still condition. Arguably, if 3-year-olds would not show referential expectations in the still condition, then 14-month-olds' failure in the current Experiment would seem less meaningful.

2.4 Experiment 2

Because it is at least theoretically possible that the negative finding from the still condition of Experiment 1 stemmed from insufficient salience or relevance of the cue, in Experiment 2 we followed up on the still cue and tried to enrich its salience. Therefore, we added a still emotional expression to the onlooker's face and tested infants of the same age

Previous research has assessed infants' understanding of facial emotions when looking at the frontal face, not its profile. By 12 months, infants interpret emotional expressions in communicative settings as referring to objects (Flom & Johnson, 2011; Hertenstein & Campos, 2004; Moses, Baldwin, Rosicky, & Tidball, 2001). Studies measuring gaze following in younger infants find that 6 and 9-month-olds follow the gaze of happy and fearful faces (de Groote, Roeyers, & Striano, 2007), although 7-month-olds seem to have difficulties to disengage from sad faces (Flom & Pick, 2005). While these studies assessed conditions under which infants follow gaze to a visible target, they did not assess object expectations indicative of VPT as in our current paradigm.

Given that infants link an adult's emotion to a referent before 14 months of age, we reasoned that in our paradigm, emotional expressions may help 14-month-olds to expect a referent when someone is looking behind a barrier. Note, however, that our primary question was still to test whether still, not ostensive-communicative expressions, would yield the effects. We used two different emotions for stimulus variety (joy and disgust) but our focus was not to test which emotion works better. Thus, we had no specific expectation that one or the other emotion would yield a stronger object expectation.

2.4.1 Methods

2.4.1.1 Participants

Parents were contacted as described in Experiment 1. The final sample consisted of N = 26 infants (13 females). Infants age range was between 14;2 and 14;29 month (M = 14;16, SD = 9.14). Additionally, n = 4 infants were invited but excluded due to fussiness. According to exclusion criteria (see Data Analysis and Reduction) each analysis is based on different Ns.

2.4.1.2 Set-up and Procedure

Set-Up and procedure remained the same as in Experiment 1. The sampling rate was set to 120 Hz.

a) b)

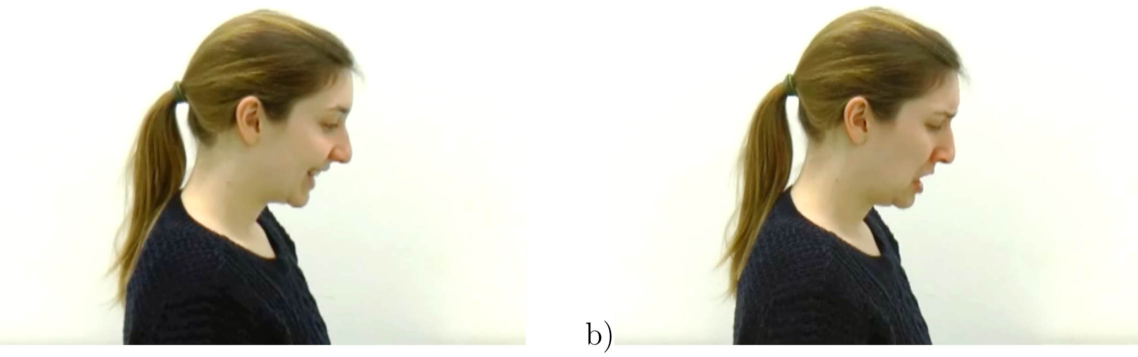

Figure 2.2. Facial expression used in Experiment 2, a) joy, b) disgust. Set-up of the video stimuli remained the same as in Experiment 1.

2.4.1.3 Stimulus and Design

Setting and size of the video stimuli were the same as in the still condition of Experiment 1. The person in profile showed a still facial expression while staring at the spot behind the occluder (see Figure 2.2). We presented two different emotions – joy and disgust, mainly for variation in the stimulus material. The person presented the facial expression from the first frame on and held the expression. Instead of a still frame we presented a video sequence to maintain validity, such that lips and body were moving slightly in a natural way. Video format and codec were the same as in Experiment 1. As in Experiment 1, videos were converted to AVI format, and the video codec was WMV1. Due to recommendations from Tobii Eye-tracking systems support we did not convert videos to 15fps. Thus, timing was the same as in the one 25fps video of Experiment 1. The fixation cross was extended to 1000 ms, to give infants enough time to fixate it. To keep the timing the same, we had the occluder come down slightly faster (200 ms than 400 ms). Presentation started with two familiarization trials and then both emotions were presented blockwise in a within design. Number of congruent and incongruent trials, presentation order and attention-getter were the same as in Experiment 1.

2.4.1.4 Data Analysis and Reduction

AOIs and dependent variables were the same as in Experiment 1. As data were sampled with 120 Hz in this experiment, we used a noise reduction (Moving Median, 3 sample window size in Tobii I-VT filter) to adjust the higher sampling rate to the smaller one from Experiment 1 (60 Hz).

Exclusion criteria were the same as in Experiment 1. For looking to the cue less than 1000m, $n = 17$ trials of 8 infants were excluded in the disgust condition and $n = 26$ trials of 10 infants in the joy. For looking at the fixation cross less than 60 ms, $n = 21$ trials of 7 infants were excluded in the disgust condition and $n = 21$ trials of 8 infants in the joy condition.

Based on these criteria, 12 infants watched enough videos of both emotions, making a within-analysis of the type of emotion less feasible. Eleven infants watched the joy condition first, and seven infants watched the disgust condition first.

2.4.2 Results

2.4.2.1 Cueing Phase

Across both emotions, infants followed to the cued occluder in $M = 18.1\%$ ($SD = 19.0$) of the trials and to the not cued occluder in $M = 11.2\%$ ($SD = 12.0$) of the trials ($t(25) = 1.67$, $p = .054$ (one-tailed), $d_z = 0.43$). Infants did not disengage from the person in 70.7% ($SD = 23.7$) of the trials.

2.4.2.2 Latency

A repeated-measures ANOVA with Congruency [congruent, incongruent] as within-subject factors, and Emotions [joy, disgust] as between-subject factor showed no significant main effect of Congruency, $F(1, 16) < 0.01$, $p = .985$, $\eta_p^2 < .001$), and no significant interaction for Congruency and Condition, $F(1, 16) = 0.20$, $p = .661$, $\eta_p^2 = .012$).

The second column of Figure 2.3 displays mean latencies for the current Experiment 2. T-test revealed no significant difference between congruent and incongruent trials, $t(19) = 1.31$, $p = .103$ (one-tailed), $d_z = .37$. Eleven out of 20 infants looked faster to the object side in the congruent videos than in the incongruent videos ($p = .412$, one-tailed).

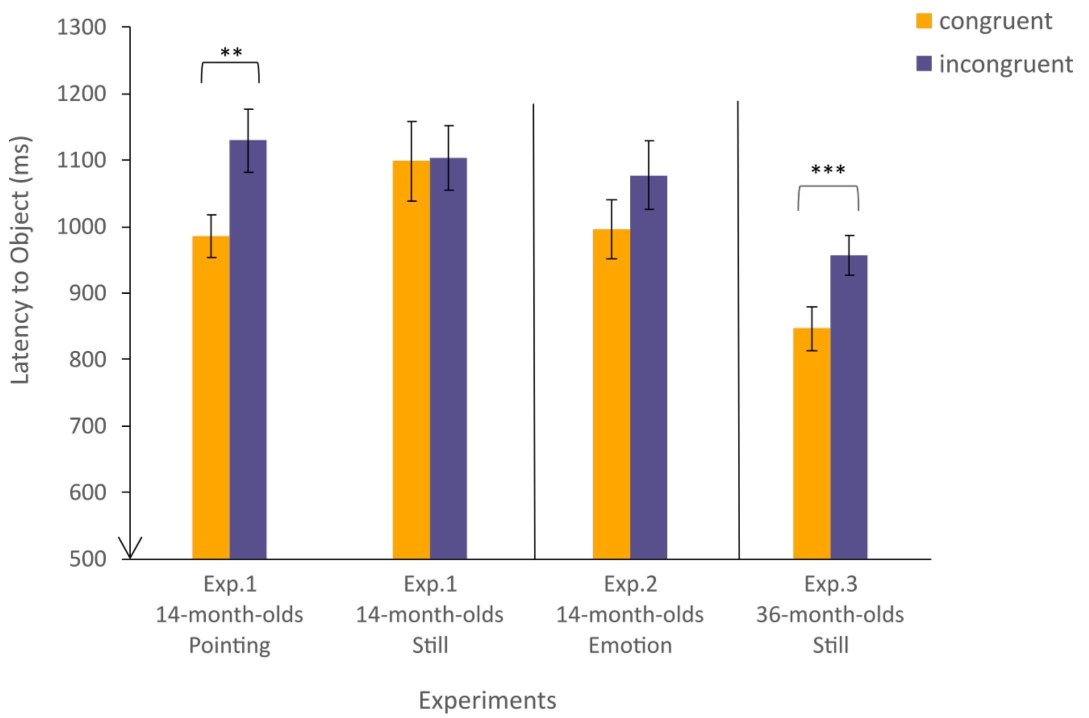

Figure 2.3. Mean latency (time to first fixation) to the object after cross off-set in congruent and incongruent trials, with standard error bars in the 3 Experiments. P-values: $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$.

2.4.2.3 Dwell-time

A repeated-measures ANOVA with Congruency [congruent, incongruent] as within-subject factor and Condition [joy, disgust] as between-subject showed no significant main effect of Congruency, ($F(1, 17) = 1.97$, $p = .179$, $\eta_p^2 = .104$), no main condition, $F(1, 17) = 0.15$, $p = .702$, $\eta_p^2 = .009$) and no interaction effect of Congruency and Condition, ($F(1, 17) = .32$, $p = .324$, $\eta_p^2 = .057$).

The second column of Figure 2.4 displays mean dwell-time for the current Experiment 2. T-test revealed no significant difference between congruent and incongruent trials ($t(21) = 0.26$, $p = .397$ (one-tailed), $d_z = 0.06$. Eight out of 22 infants looked longer to the empty side in the incongruent videos than in the congruent videos ($p = .143$, one-tailed).

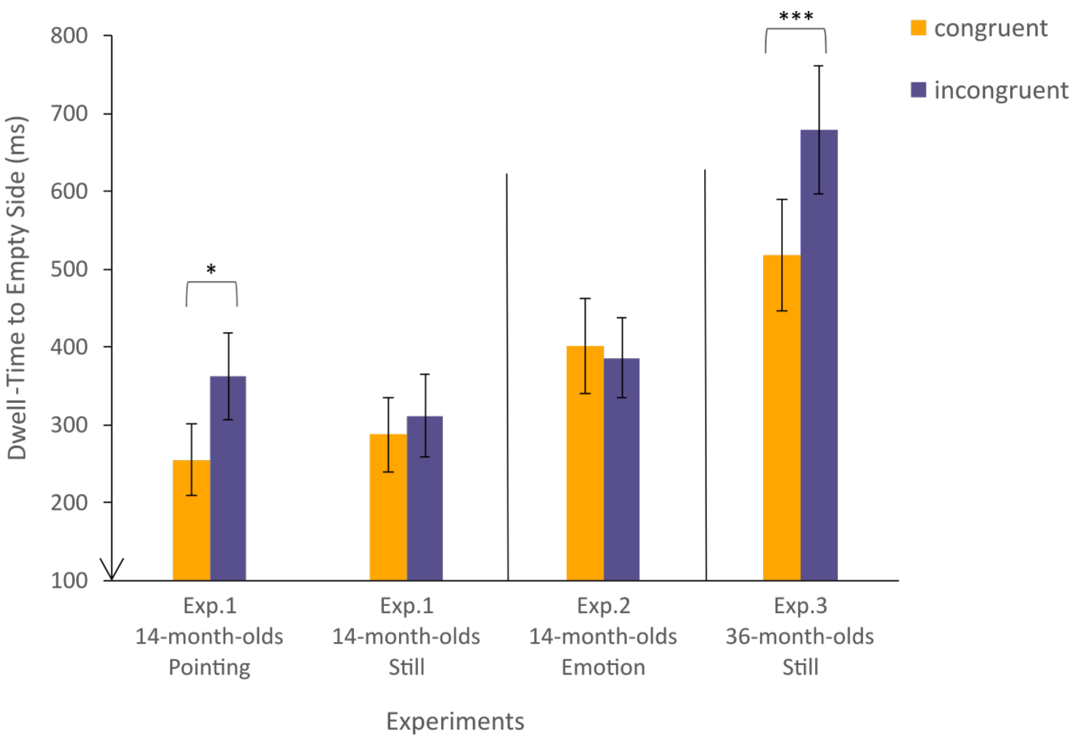

Figure 2.4. Mean of total dwell-time to the empty side of the occluder in the outcome phase in congruent and incongruent trials, with standard error bars in the 3 Experiments. P-values: $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$.

2.4.3 Discussion

Adding emotional expressions to the still cue did not help 14-month-olds to instigate object expectation. Unlike following the explicit communicative pointing cue of Experiment 1, the current emotional still cue was as ineffective as the still cue of Experiment 1 in instigating a referential expectation of an object. These findings are thus in support of our interpretation of Experiment 1 that at 14 months, infants do not automatically take another's perspective but likely require communicative cues to do so.

The findings may appear at odds with studies showing that facial expressions enhance gaze-following already in 9-month-olds (de Groote et al., 2007). However, in contrast to the current paradigm, gaze following and social referencing paradigms present the face frontal and usually include a variety of social-communicative cues like eye-contact and vocal expression, which were absent in our paradigm by intended design. Further, none of the previous studies on emotion cueing tested whether these effects instigate cognitive representations. Nevertheless, it is important to stress that while the negative findings support our interpretation of Experiment 1, they do not question infants' ability to process and understand facial emotions in general.

2.5 Experiment 3

While Experiment 1 showed that communicative cues instigate object expectations, it could be that the absence of such expectations following the directional still cue were an artefact of the experimental design. Experiment 2 did not provide support for that interpretation, but it could be that the still cue does not work at all in the current paradigm. Therefore, in the current Experiment 3 we tested whether the still cue would produce an effect in children older than 14 months of age. We tested preschoolers at 36-month-olds, who have developed sophistical social cognitive abilities and competencies in language use, and understand the referential nature of points and words. If our paradigm was indeed sensitive to revealing spontaneous, non-ostensive visual-perspective-taking, we expected that 36-month-olds should show cueing effect and a dwell-time difference in the still condition of Experiment 1.

2.5.1 Methods

2.5.1.1 Participants

Parents were contacted as described in Experiment 1. The final sample consisted of $N = 30$ toddlers (14 female) with a mean age of 36 month and 15 days which ranged from 36;2 to 36;27. Additionally, six toddlers were invited but excluded due to fussiness ($n = 2$), bad calibration ($n = 3$) or failed calibration ($n = 2$). According to exclusion criteria (see Data Analysis and Reduction) sample size differs across analysis.

2.5.1.2 Set-up and Procedure

Set-Up and procedure remained the same as in Experiment 1. Tobii eye-tracker X120 was sampling with 120 Hz.

2.5.1.3 Stimulus and Design

Setting, size, frame-rate and timing of the video stimuli were the same as in the still condition of Experiment 1. As we were only interested in the still condition, there was no further within-subject factor other than congruency.

2.5.1.4 Data Analysis and Reduction

Dependent variables and AOIs were the same as in Experiment 1. As in Experiment 2 we enabled a Moving Median noise reduction (3 sample window size in Tobii I-VT filter) to adjust the higher sampling rate (120 Hz) to the smaller one from Experiment 1 (60 Hz).

Exclusion criteria were the same as in Experiment 1. For watching the cueing phase less than 1000 ms, n = 15 trials of 10 toddlers were excluded. For looked at the fixation cross less than 60 ms, n = 23 trials of 11 toddlers were excluded.

To compare latency and dwell-time in congruent and incongruent videos we used paired-sample t-test, which we report one-tailed according to our hypotheses. To assess the effects on an individual level, we used binomial tests.

2.5.2 Results

2.5.2.1 Cueing Phase

Toddlers followed the cue in M = 42.2% (SD = 27.8) of the trials and looked to the not-cued occluder in M = 13.6% (SD = 14.8) of the trials. These values differed significantly from each other, $t(29) = 4.64$, $p < .001$ (one-tailed), d_z = -1.30. Toddlers did not disengage from the person in 44.2% (SD = 29.1) of the trials.

2.5.2.2 Latency

The last column of Figure 2.3 displays the mean latencies of the current Experiment 3. Toddlers were significantly faster to detect the object in congruent than incongruent trials, $t(25) = 3.01$, $p = .003$ (one-tailed), $d_z = 0.69$. Nineteen of 26 toddlers looked faster to the object in the congruent videos than in the incongruent videos (binomial test, $p = .015$, one-tailed).

2.5.2.3 Dwell-time

The last column of Figure 2.4 displays the mean dwell-time of the current Experiment 3. Toddlers looked significantly longer to the empty side in incongruent than congruent trials, $t(26) = 2.72$, $p = .006$ (one-tailed), $d_z = 0.39$. Twenty of 27 toddlers looked longer to the empty side in the incongruent than congruent videos (binomial test, $p = .010$, one-tailed).

2.5.3 Discussion

When 36-month-olds watched videos of a person simply directed to one side, without communication, they spontaneously and seemingly automatically adopted her perspective and expected to see an object, as evident in our measures of latency and dwell-time. On both these measures, the effects were significant not just on the group level but also for a significant majority of children. Thus, by 36 months of age, a still person elicits visual-perspective-taking without communicative cues. The findings demonstrate that the still condition of the current paradigm is a viable way of assessing automatic visual-perspective-taking. The pattern of findings from Experiments 1-3 supports the interpretation that visual-perspective-taking is an emerging skill, which initially relies on communicative cues. It is reasonable that it derives from social engagement and interaction, and then becomes automatized with development and experience.

2.6 General Discussion

The current study investigated the age and conditions under which children begin to engage in visual-perspective-taking of the kind that enables representing that others represent something in their line of sight. This basic form of visual-perspective-taking is conceptually different from more advanced forms of representing what exactly someone else represents (Flavell, Everett, Croft, & Flavell, 1981), or (falsely) believes (Tomasello, 2018). It is conceptually also different from simpler forms of only following others' line of sight perceptually, or orienting covertly to a cued location. The paradigmatic case of the form of visual-perspective-taking under investigation is that a participant cannot see an object, even when following another's line of sight, but can infer the presence of an object given the person's behavior. Previous research had shown that this ability is in place around 12 months of age (Behne et al., 2012; Csibra & Volein, 2008; Moll & Tomasello, 2004; Pätzold & Liszkowski, 2019). While these paradigms have all employed communicative cues, other paradigms, primarily from adult visual-perspective-taking tasks have revealed that non-ostensive, directed cues, like a person or face in profile, are sufficient to cue a perspective, even in infancy (Kampis et al., 2015; Kovács et al., 2010; Samson et al., 2010; Teufel, Alexis, Clayton, & Davis, 2010). The current study investigated to what extent and at what ages in development, infants rely on socially directive, communicative cues versus non-ostensive still displays of directed perspectives.

The main Experiment 1 established in a within-subject design that communicative pointing indeed, as expected from the literature, induced cognitive representations of a hidden object. In addition, it revealed the relative absence of this effect following a still directed perspective of a person in profile. The two subsequent experiments were conducted to follow up on this absence of the effect. They established on the one hand that adding still emotional expressions did not help 14-month-olds; and on the other hand that the effect is significant at an older age. Together, the pattern of findings of the first three experiments thus provides a firm basis to suggest that seemingly automatic visual-perspective-taking following non-ostensive still directive perspectives is a developmentally emerging skill rather than a starting point of automatic VPT (e.g. Apperly & Butterfill, 2009). At the same time, these current findings enforce previous interpretations on which visual-perspective-taking first emerges in communicative situations through joint engagement (Moll & Tomasello, 2007).

Using our paradigm, we cannot fully exclude that general attention orienting caused object detection and looking time differences. We tried to overcome a spatial cueing effect by letting the cue disappear and implemented a fixation cross, that infants had to fixate before latency and looking time was measured. Thus, we eliminated the possibility that the cue drew the attention spatially to one side while the cue was still present. Infants had to remember the social cue for at least one second before they were allowed to show a reaction. In addition, we used a long SOA, to measure a more reflective VPT and let infants evaluate the outcome with a VoE measure. Control conditions used in other studies, such as non-social cues (Pätzold & Liszkowski, 2019), control gesture (Southgate, Johnson, El Karoui, & Csibra, 2010), foils (Bertenthal et al., 2014) or mechanical claws (Daum & Gredebäck, 2011), revealed no or lower effects than found with pointing or grasping actions. Thus, it is very likely that we measure referential understanding caused by specific gestures. However, future studies using our new paradigm should implement further control conditions.

2.6.1 Conclusion

With the current study, we closed the gap between usual cueing paradigms, measuring attention direction to visible objects, which is already present in very young infants, and more complex tasks, measuring referential expectation mostly in interaction-based measures, suitable only for older infants. We here systematically compared different cues and narrowed cue properties for referential expectation. Our findings do not confirm the theoretical assumption that spontaneous, efficient VPT is visible from early on (Surtees & Apperly, 2012). It is more in line with theories suggesting that VPT gets spontaneous in a variety of situations later in development, through experience and interaction (Moll & Tomasello, 2007). Thus, infants show VPT in communicative action, like pointing gestures, around their first birthday. Only later, at least from 3 years on, they show it in non-ostensive and non-relevant actions, similar to VPT in adults. An initial absence of spontaneous representing of a still perspective at 14 months, in light of the presence of spontaneous representing at 36 months, strongly support of a continuous social-constructive view of a developing social understanding.

Chapter 3

Do 8-month-olds show referential expectation for communicative and intentional actions?[1]

RESEARCH HIGHLIGHTS

- 8-month-olds lead their attention faster to the cued object when communicatively pointing, but not when unsuccessfully reaching behind a barrier
- With 8 months of age infants do not show a stable referential expectation indicated by a longer dwell-time in incongruent trials
- 8-month-olds show a cueing effect and referential expectation when successfully reaching behind a barrier, but this may be confounded by exogenous cueing
- Infants do not seem to have visual-perspective-taking abilities from early on

Acknowledgements: We thank all families who participated in these studies. We thank Maike Fuchs for her help in creating the stimuli and running some of the experiments. We are grateful to Marie-Louise Braun, Ishita Noori Haider, and Leonie Kinsky for their assistance with data collection.

[1]This study is part of a paper currently under revision entitled "Inferring hidden objects from still and communicative onlookers at 8-, 14- and 36-months of age" (Jartó & Liszkowski, under revision) in the Journal of Experimental Child Psychology.

3.1 Abstract

It is discussed whether infants have visual-perspective-taking (VPT) abilities from birth or whether they develop this ability only after 9 months of age, when they start to engage in joint attention. Thus, this study examined if 8-month-olds show referential expectation when someone is pointing or reaching behind an occluder. The stimuli set-up was the same as in Study 1. 8-month-olds showed a cueing effect in the pointing condition and when the person was fully reaching behind the occluder, but not when the person stopped half way. Referential expectation indicated by a longer dwell-time to the empty side in incongruent trials was only present in the full reach condition, which could be confounded by exogenous attention direction. We did not find the effect when the person was stopping the reach half way and did not touch the occluder. Thus, we could not find a stable indication that infants can infer the goal and the perspective of an agents by the age of 8 months. Together with results from Study 1, this findings support a constructivist view, assuming VPT abilities only at the end of the first year as a developmental outcome.

Keywords: visual-perspective-taking, referential expectation, eye-tracking, point understanding, instrumental action understanding

3.2 Introduction

We now know that 14-month-olds show visual-perspective-taking (VPT) indicated by a referential object expectation in communicative situations but not in non-communicative situations (see previous Chapter 2 - Study 1). Following up on this positive effect for communicative gestures, the question remains how early in development VPT may be present, again, in contrast to simpler spatial attention direction. While infants are known to follow others' gestures from early on (Daum & Gredebäck, 2011; Rohlfing et al., 2012), different theoretical assumptions exist on the emergence of VPT and to date, empirical evidence is mixed. From a theoretical point of view, Tomasello (2019) would suggest that before 9-12 months of age, before the so called 9 months revolution, infants' joint engagement skills are not fully triadic referential and do not involve understanding others' perspectives. However, Csibra (2003) assumes a referential system elicited by communicative cues from early on. In addition, Luo and Baillargeon (2010) assume an innate mentalistic action interpretation system, in the way that infants consider others' perspectives and mental representations when interpreting their actions.

Empirical evidence for VPT before the 9 months revolution (Tomasello, 2019) provides positive and negative results. For communicative cues, like pointing gestures, the majority of studies finds that infants show indication for VPT by understanding that the other person refers to and represents something when pointing somewhere, not before the end of the first year (e.g. Behne et al., 2012; Pätzold & Liszkowski, 2019). However, our design is based on a study, suggesting that already 8-month-olds show indication for referential expectation when someone looks communicatively behind a barrier (Csibra & Volein, 2008). For other familiar actions, like reaching actions, infants are found to take into account an agent's perspective even earlier (6 months; Luo & Johnson, 2009) and can anticipate hidden goals (Applin & Kibbe, 2019). However, others find that only after 10 months infants anticipate the goal referent in unfulfilled reaching actions (Brandone et al., 2014; Cannon & Woodward, 2012). Thus, to date, empirical evidence does not provide a clear picture at what age infants start to take into account others' perspectives.

With the current study, we address this debate and want to investigate whether we can find indication for VPT at 8 months of age, indicated by inferring the presence of a referent within a person's line of sight, even when the participant cannot see the object. We used the same cueing paradigm as in Study 1. As central cue, a person was either pointing or reaching behind one of two barriers. In half of the trials an object revealed on the cued side (congruent trials), in the other half on the non-cued side (incongruent trials). Again, we measured a cueing effect, indicated by a faster latency to the object in congruent trials than in incongruent trials. In addition, we tested if we can find referential object expectation, indicated by a violation of their expectation (VoE). If 8-month-olds take into account others' perspectives and expect an object in the cued direction in communicative and instrumental action, we expect to find a longer looking time on the empty side in incongruent trials than in congruent trials.

3.3 Experiment 1

In Experiment 1, we tested whether 8-month-olds understand that others are referring to and mentally represent something when someone points, to trace the emergence of communication-induced referential object expectations. In a similar design (where ours is based on) Csibra and Volein (2008) found a VoE indicated by a longer looking time when a person communicatively looked behind one of two barriers and the space behind the cued barrier revealed to be empty. In this study, there was no significant difference found between 8 and 12 months old infants. However, 8-month-olds were not tested separately. When testing only 8-month-olds, Pätzold and Liszkowski (2019) did not find referential expectation when using pupil dilation as VoE measure. Thus, if 8-month-olds already understand that others refer to and mentally represent something when they point in a communicative way, we expect an earlier object detection on congruent trials and a longer looking time to the empty side in incongruent trials.

3.3.1 Methods

3.3.1.1 Participants

Parents were contacted via birth-register in a metropolis in Western Europe. For power analysis we used the effect size $\eta_p^2 = .30$ of Csibra and Volein's (2008) study (as in Study 1). With a p-value of $p = .05$ it required a sample size $N = 18$ infants to have an 95% chance of detecting significant effect in a repeated measurement analysis.

The final sample consisted of $N = 22$ infants (12 females) with a mean age of 8 months and 14 days which ranged from 8;2 to 8;29. Additionally, six infants were invited but excluded because calibration was not possible ($n = 1$) or not accurate enough ($n = 5$). According to exclusion criteria (see data analysis and reduction) sample size differs in each analysis.

3.3.1.2 Set-up and Procedure

Set-Up and procedure was the same as in Experiment 1 of Study 1.

3.3.1.3 Stimulus and Design

Setting of the video remained the same as in the pointing condition of Experiment 1 of Study 1. Except that following Tobii-Studio advice, rendering was set to 25fps (instead of 15fps), so that every time sequence happened 120 ms earlier. To account for possibly longer processing time overall, the outcome phase was extended for 1000 ms to 4800 ms. Presentation of the four congruent and four incongruent trials started with two familiarization trials. For a further description see "Stimulus and Design" of Experiment 1 Study 1.

3.3.1.4 Data Analysis and Reduction

Dependent variables, AOIs, data reduction and exclusion criteria remained the same as in Study 1 of Study 1. For watching the cue phase less than 1000 ms, $n = 1$ trial of 1 infant was excluded. For looking at the fixation cross for less than 60 ms, $n =$ 22 trials of 6 infants were excluded.

A t-test was calculated to compare the dependent variables between congruent and incongruent videos and reported one-tailed according to our directed hypothesis. In addition, binomial test were calculated to assess the effects on an individual level.

3.3.2 Results

3.3.2.1 Cueing Phase

Infants followed the cue in 11.1% (SD = 21.1) of the trials and looked to the not cued occluder in 3.3% (SD = 9.4) of the trials. These values showed a close to significant difference, $t(22) = 1.72$, $p = .050$ (one-tailed), $d_z = 0.47$. Infants did not disengage from the person in 85.6% (SD = 24.2) of the trials.

3.3.2.2 Latency

The second column of Figure 3.2 displays the mean latency of the current experiment. In congruent trials 5% of the variable latency were winsorized. Infants were significantly faster to detect the object in congruent than incongruent trials, $t(17) = 4.00$, $p < .001$ (one-tailed), $d_z = 0.85$. Fourteen out of 18 infants looked faster to the congruent object than to the incongruent object, binomial test, $p = .006$ (one-tailed). The significant pattern of results remained the same when analyzing the time window without the extended 1000 ms in the test phase.

3.3.2.3 Dwell-time

The second column of Figure 3.2 displays the mean dwell-time of the current experiment. In congruent and incongruent trials 5% of the variable duration were winsorized. There was no difference in the dwell-time on the empty location between the congruent and incongruent trials, $t(19) = 0.31$, $p = .379$ (one-tailed), $d_z = 0.06$. Ten out of 20 infants looked longer to the empty side in the incongruent videos than in the congruent videos which is not significantly above chance ($p = .500$, one-tailed). Results remained the same when analyzing the time window without the extended 1000 ms in the test phase (see Appendix A.1.2 for results).

3.3.3 Discussion

We could find a cueing effect to hidden objects for 8-month-olds with a communicative pointing cue. This is in line with evidence that already 4.5-month-olds direct their attention to the pointed at object faster than to an incongruently presented object (Rohlfing et al., 2012). However, this ability seems to be very context sensitive and not very stable at this young age, as for example infants do not show it when the hand is not moving or disappears before the object reveals (Daum et al., 2013; Rohlfing et al., 2012). Thus, the naturalistic and long presentation of our cue may be key components for finding an effect at this young age. For referential expectation, we could not find a stable indication. This adds to the assumption, that although point following is possible, infants do not understand the communicative and referential act of a pointing gesture (Tomasello et al., 2005). In addition, our findings fit to the empirical evidence that 12- but not 8-month-olds show communicative induced object expectation (Pätzold & Liszkowski, 2019).

It is possible that 8-month-olds' are still less familiar with the pointing cue, especially since they do not point themselves at that age. Infants do, however, reach for objects at that age and begin to understand reaching as object-directed (Brandone et al., 2014; Hamlin, Hallinan, & Woodward, 2008). In a final experiment, we therefore substituted the pointing cue with a reaching cue to test whether infants would expect a reaching action to be related to occluded objects, perhaps as indication of an early understanding of the reacher's perspective.

3.4 Experiment 2

Infants are found to understand others' reaching action from very early on. Already 6-month-olds seem to interpret actions as goal directed (Woodward, 1998) and show a cueing effect when presenting a grasping hand (Daum & Gredebäck, 2011; Wronski & Daum, 2014). However, when infants do not see the fulfilled actions they seem to be able to anticipate the goal only with 10 months of age (Brandone et al., 2014). Thus, the question arises if we can find referential object expectation for reaching actions, which is a new approach to test if infants know about the mental representation and the goal of an agent.

In the current Experiment 2, we used again the same general paradigm and hypotheses to test whether 8-month-olds expect an occluded object when someone is reaching non-communicatively behind a barrier. In Experiment 2a, infants watched a complete reach with the extended arm ending behind the occluder. In Experiment 2b, infants watched a reach analogous to the pointing shape, with the reaching hand stopping where the pointing hand of Experiment 4 had stopped, to equate for the surface features of the cue.

3.4.1 Methods

3.4.1.1 Participants

Daum and Gredebäck (2011) found an cueing effect size of $\eta_p^2 = .45$ for a grasping hand in 7-month-olds. With a p-value of $p = .05$ it required a sample size $N = 11$ infants to have an 95% chance of detecting significant effect in a repeated measurement analysis.

The final sample in Experiment 2a consisted of $N = 25$ infants (14 females) with a mean age of 8 months and 14 days, which ranged from 8;2 to 8;29. Additionally, eight infants were invited but excluded because calibration was not possible ($n = 1$), calibration was not accurate enough ($n = 5$) and because eye-tracker lost eye gaze due to high fussiness and movement ($n = 2$).

In Experiment 2b, the final sample consisted of $N = 27$ infants (12 females) with a mean age of 8 months and 14 days, which ranged from 8;2 to 8;29. Additionally, seven infants were invited but excluded because calibration was not accurate enough ($n = 5$) or because the eye-track lost eye gaze due to high fussiness and movement ($n = 2$). According to exclusion criteria (see data reduction and analysis), sample size differs in each analysis.

3.4.1.2 Set-up and Procedure

Set-Up and procedure remained the same as in Study 1. The sampling rate was adjusted to 120 Hz.

3.4.1.3 Stimulus and Design

Setting and size of the video stimuli were the same as in all other experiments (for a detailed description see Experiment 1 of Study 1). In the current experiment, the person was reaching instead behind one of the to boxes. The person started reaching from her lap. In Experiment 2a, the person grasped behind one of the two occluders until the hand was not visible anymore (see Figure 3.1). In Experiment 2b, the reach stopped at the position of the pointing cue from Experiment 1 (see Figure 3.1). The videos had a framerate of 25fps. Timing and presentation order of the video stimuli were the same as in Experiment 1.

a) b)

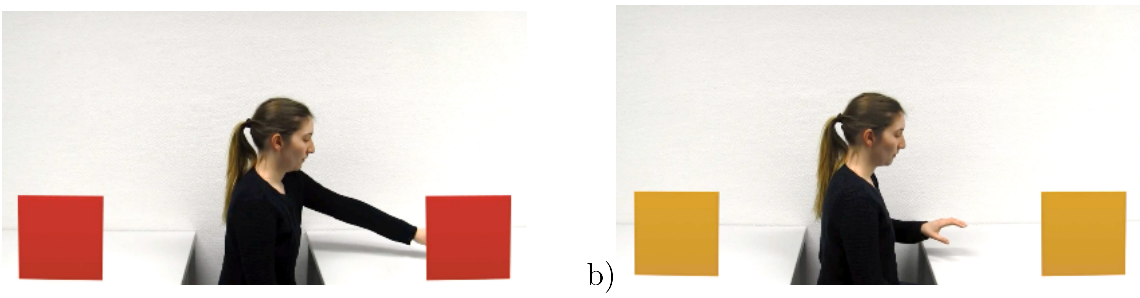

Figure 3.1. Set-up of the cueing phase in a) Experiment 2a and in b) Experiment 2b. In both conditions the persons hand started from her lap and stopped in the position presented in the figure.

3.4.1.4 Data Analysis and Reduction

Dependent variables and AOIs were the same as in Experiment 1. Again, we enabled a Moving Median noise reduction (3 sample window size in Tobii-IVT-filter) to adjust the higher sampling rate (120 Hz) to the smaller one (60 Hz).

In Experiment 2a, for watching the cue phase less than 1000 ms, $n = 21$ trials of eight infants were excluded. For looking at the fixation cross less than 60 ms, $n = 28$ trials of 14 infants were excluded. In Experiment 2b, for watching the cue phase less than 1000 ms, $n = 15$ trials of 11 infants were excluded. For looking at the fixation cross less than 60 ms, $n = 40$ trials of 15 infants were excluded.

3.4.2 Results Experiment 2a

3.4.2.1 Cueing Phase

Infants followed the cue to the cued occluder in 70.5% ($SD = 26.7$) of the trials significantly more often than to the not cued occluder in 3.5% ($SD = 25.0$) of the trials, $t(24) = 10.86$, $p < .001$ (one-tailed), $d_z = 3.57$. Infants did not disengage from the person in 26.0% ($SD = 22$) of the trials.

3.4.2.2 Latency

The second column of Figure 3.2 displays the mean latency of the current Experiment 2a. Dependent t-test revealed that infants were significantly faster to detect the object in congruent than incongruent trials, $t(18) = 3.60$, $p < .001$ (one-tailed), $d_z = 0.83$. Number of infants looking faster to the object side in the congruent videos than in the incongruent videos were 16 out of 19 infants, binomial test was significant ($p = .002$, one-tailed). Results were exact the same when excluding the last 1000 ms from the extended test window.

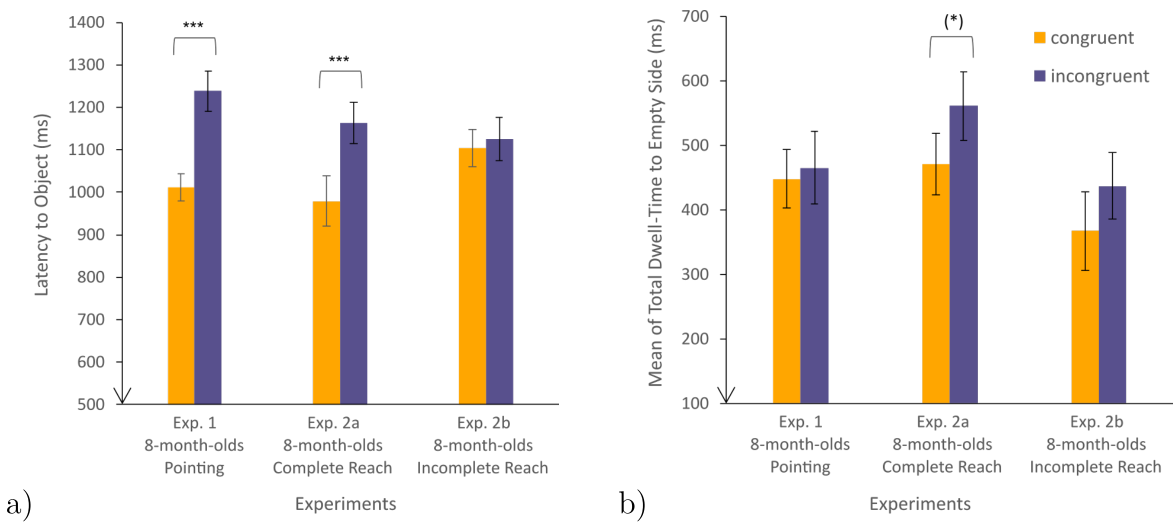

Figure 3.2. Graph a) displays the mean latency (time to first fixation) to the object after cross off-set and b) the mean of total dwell-time to the empty side of the occluder in congruent and incongruent trials in all 3 Experiments. Asterisks display significance of dependent t-tests between congruent and incongruent trials. Asterisks in brackets represent significance only when test-time-window was shortened. P-values: $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$. Error bars represent standard errors.

3.4.2.3 Dwell-time

The second column of Figure 3.2 displays mean dwell-time of the current Experiment 2a. A dependent t-test revealed that infants did not look longer to the empty side in incongruent than congruent trials, $t(19) = 1.18$, $p = .127$ (one-tailed), $d_z = 0.31$. When excluding the last 1000 ms from the extended test window, the comparison was significant, $t(19) = 2.08$, $p = .026$ (one-tailed), $d_z = .499$ ($M_{congruent} = 409.97$, $SD = 205.70$; $M_{incongruent} = 537.88$, $SD = 289.41$). The number of infants looking longer to the empty side in the incongruent videos than in the congruent videos were 13 out of 20 infants (p = .132, one-tailed) for both time-windows.

3.4.3 Results Experiment 2b

3.4.3.1 Cueing Phase

Infants followed the cue to the cued occluder in 12.7% (SD = 17.3) of the trials and looked to the not cued occluder in 9.8% (SD = 21.8) of the trials, revealing no significant difference, $t(26) = 0.50$, $p = .310$ (one-tailed), $d_z = 0.15$. Infants did not disengage from the person in 77.5% (SD = 25.1) of the trials.

3.4.3.2 Latency

The last column of Figure 3.2 displays mean latency of the current Experiment 2b. Dependent t-test revealed that infants were not significantly faster to detect the object in congruent than incongruent trials, $t(17) = 0.28$, $p = .392$ (one-tailed), $d_z = 0.09$. Number of infants looking faster to the object side in the congruent videos than in the incongruent videos were 10 out of 18 infants (p = .408, one-tailed). Results were exact the same when analyzing latency without the extended 1000 ms in the test screen.

3.4.3.3 Dwell-time

The last column of Figure 3.2 displays mean latency of the current Experiment 2b. A dependent t-test revealed that infants did not look longer to the empty side in incongruent than congruent trials, $t(21) = 0.73$, $p = .237$ (one-tailed), $d_z = 0.17$.

Number of infants looking longer to the empty side in the incongruent videos than in the congruent videos were 11 out of 22 infants (p = .500, one-tailed). Non-significant pattern remained the same when analyzing dwell-time without the extended 1000 ms in the test screen (see Appendix A.1.1 for results).

3.4.4 Discussion

Individual reaching for an occluded object instigated object expectations at 8 months of age, as revealed by latency and dwell-time measures, but only if infants saw a full reach. This finding assures internal validity of our paradigm and measures. However, one must note that attention to the cue in 2a was conflated with attention to the location, since the cue did not stop before the location. As such, it may have exogenously cued the location. Indeed, when the reaching cue in Experiment 2b was matched to the distance and kinematics of the pointing cue, it did not cue the congruent side during the cueing phase and it did not enhance the latency to detect the target at that side after the cueing phase. This contrasts with findings for the communicative pointing cue of Experiment 1.

In addition, results of Experiment 2b is in contrast with recent findings, indicating goal anticipation to hidden targets with 6 months of age (Applin & Kibbe, 2019). However, in our paradigm infants never saw the person successfully grasping the object. Thus, representing an object, which has not been seen before, may be a more complex understanding than anticipating to an object seen to be reached at before (Brandone et al., 2014). Further, in the current experiment the action stopped in a very unnatural way, and although they may be able to lead their attention to the cued side when seeing an isolated hand, they may be sensitive to unnatural actions and need natural movement (Wronski & Daum, 2014).

Thus, while Exp. 2a may suggest that infants do seem to understand something about others' directedness of actions, this understanding does not seem to lend itself to understand others' visual perspective.

3.5 General Discussion

In the present study, we tested whether 8-month-olds show VPT in a paradigm suitable for young infants, trying to test for a higher ability than simple spatial attention orienting. We used a cueing paradigm with different central cues – a person either pointing communicatively or reaching behind one of two barriers. The occluder revealed an object either on the congruent or incongruent side. We presented a whole person for a relatively long time, so infants had time to process the very naturalistic cue. The paradigmatic case of the form of VPT we investigated is that a participant cannot see an object, even when following another's line of sight, but can infer the presence of an object given the person's behavioral appearance.

At 8-months of age, we found a faster latency to the cued object than to the not cued object in a communicative pointing gesture and fulfilled reaching action, but not for incomplete reaching actions. Thus, we could replicate previous studies finding a cueing effect for pointing gestures (Rohlfing et al., 2012) and grasping hands (Daum & Gredebäck, 2011) for very young infants. Although we tried to control for spatial attention direction, as infants first had to fixate a cross between the boxes before looking to the object, the found cueing effect may measure a simpler action understanding and may be conflated with spatial attention direction. This is especially the case for the full reach, where the hand touched the occluder. Thus, the cueing effect may be comparable to point following to the next visible object, and not understanding that others refer to something. Similarly, the cueing effect for the grasping action may be more comparable to anticipating the path of a reaching action but not the goal (Cannon & Woodward, 2012).

In addition, for our VoE measure, we only found VPT indicated by referential expectation in the full reach condition. Again, this could be conflated with spatial attention orienting. Not finding referential expectation for the pointing cue and the unsuccessful reaching action, is in line with the theoretical assumption that infants start to learn about others perspectives only after 9 months of age, when they start to engage in joint attentional sequences (Tomasello, 2019). It also fits with empirical evidence that infants take into account what others can see only at the end of the first year (Pätzold & Liszkowski, 2019) and anticipate the goal of unfulfilled reaching actions only from 10 or 11 months on (Brandone et al., 2014; Cannon & Woodward, 2012). The aim of further studies should be to pinpoint the exact time point of emergence.

Our findings are in contrast with the theoretical assumption, that infants have a referential and mentalistic understanding of others' actions from early on (Luo & Baillargeon, 2010). Although, not finding results are not an indication for the absence of an ability, evidence for VPT in early ages is very rare. For communicative gestures, only Csibra and Volein (2008) argue that they found referential expectation for 8-month-olds, even though, they did not test them as separate groups. The majority of tasks show a referential understanding of others' pointing gestures at the end of the first year (Behne et al., 2012; Pätzold & Liszkowski, 2019). Not finding referential expectation for the incomplete reaching actions in our study seems to be in contrast with finding that already 6-month-olds take into account others' perspectives in a reaching action (Luo & Johnson, 2009) or can anticipate the goal object even when it is hidden (Applin & Kibbe, 2019). However, in all these tasks the infants saw a person reaching for an object several times before they watched a test trial. Thus, connecting the hand with the object or the location could be the reason why we find referential expectation in the fulfilled reaching action. This fits with the empirical evidence that only after 10 months of age infants seem to be able to infer the goal of an action, when they have never seen the goal being fulfilled before (e.g. Brandone et al., 2014). Thus, representing an object, which infants have not seen to be reached at before, may be a higher action interpretations ability and a better measure for examining if infants understand the underlying goal and the perspective of a reaching person.

3.5.1 Conclusion

Using a VoE measure to examine referential expectation as indication for VPT is not a common approach. Thus, this measure has to be validated with other measures, which are used to test referential object expectation and VPT. However, finding a cueing effect but not a VoE seems to be an indication, that we capture a higher ability than spatial attention direction. The results of the current study fit with the theoretical approach, that infants develop an understanding of others' perspectives only later in the first year of life and do not show VPT from early on.

Chapter 4

Developmental relations of social-cognitive abilities in the first year of life

RESEARCH HIGHLIGHTS

- A longitudinal approach showed social-cognitive abilities to consolidate from 11 months of age onward.
- Relations between action and communicative understanding support the assumption that infants develop an integrated understanding of others.
- No strong validity was found between interaction-based and eye-tracking tasks in the first year of life.
- Indications for action and referential understanding being separate system and present form early on were not found.
- Results support social constructive views where social-cognitive abilities develop only after 9 months of age through joint attention.

Acknowledgements: We thank all families who participated in this long-term study. We thank Anne Kraus, Öznur Yel, Sibel Ünlü, Senta von Münchow, Björn Lünstäden, Anna Huang, and Rayén Feil for their long-term and intensive participant recruitment and mentoring as well as their assistance with data collection. We want to thank our collaborators from Koç University, and Prof. Dr. Aylin Küntay, Dr. Hilal Sen, Ebru Ger, Sümeyye Koşkulu and Merve Ataman for the valuable feedback and discussions.

4.1 Abstract

Several theories assume that different early social-cognitive abilities are the product of one underlying understanding of others (e.g. Tomasello et al., 2005). A variety of research methods have shown that infants understand a lot about others' actions and communicative gestures in their first year of life already. However, we do not know when these different abilities, measured with different methods, become stable competencies and if, and how, they are related in their emergence. Finding relations between skills and methods would be evidence that we are tapping one understanding. Therefore, we conducted a longitudinal study assessing action understanding and communicative understanding in eye-tracking as well as interaction-based tasks. Monthly between the ages of 8/9 and 12/14 month, we tested three different skills: failed reaching, point following to visible entities, and point following to hidden entities. We found associations between action understanding and communicative understanding in both methods. These were mainly present from 11 months on, when the majority of skills is stable. We find a few relations between methods, mainly from interaction-based tasks to eye-tracking tasks one months later. Our results support the view that one integrated understanding of others develops after the 9-months-revolution and reveals itself first within interaction. Finding stable abilities at the end of the first year and relations between different skills does not fit with the assumption that abilities are present from early on as separate systems.

Keywords: infancy, social-cognitive abilities, action understanding, communicative understanding

4.2 Introduction

A variety of social-cognitive skills develop in infancy. Importantly, infants begin to understand actions as goal-directed (action understanding; Woodward, 1998) and develop a referential understanding of non-verbal gestural communication (communicative understanding; Carpenter et al., 1998; Tomasello, Carpenter, & Liszkowski, 2007). These skills are important precursors to subsequent language development and theory of mind (Brooks & Meltzoff, 2015; Sodian & Kristen-Antonow, 2015). However, crucial questions regarding their emergence and development remain. Conceptionally, some researchers assume skills are present from birth and emerge in behavior early on (Baillargeon et al., 2016), while others argue for a development of abilities at the end of the first year of life (Tomasello, 2019). Additionally, some researchers suppose that action understanding and communicative understanding emerge independently as two separate systems (Csibra, 2003), while others assume that the emergence of these two major skills is based on an integrated understanding of others (Tomasello, 2019). Methodologically, a variety of different paradigms and measures are used that show different time points of emergence. To better understand the origins and development of these key components of social-cognitive and language development, it is important to determine the time points of emergence across different methods, as well as their interrelations.

First, evidence on the time point of emergence of social-cognitive skills is mixed. On the one hand, 6-month-olds are found to evaluate others' actions as goal-directed (Woodward, 1998) or anticipate others' incomplete actions correctly (Applin & Kibbe, 2019). In cueing paradigms, even 4-month-olds detect cued objects earlier than non-cued objects (cueing effect) when presenting others' gazes (Farroni et al., 2003) or pointing gestures (Bertenthal et al., 2014). On the other hand, in other studies, cueing effects are found only at the end of the first year (Daum et al., 2013) and infants understand the referential of pointing gestures with 12 months, but not with 8 months (Pätzold & Liszkowski, 2019). In addition, 10-month-olds are found to anticipate the target object of an unfulfilled reaching action but not 8-month-olds (Brandone et al., 2014). Thus, social-cognitive abilities seem to be very task specific, and it remains unclear by which time point we can find stable competences. Examining stability between measuring time points in a longitudinal design and using different measures enables us to detect the most likely time point of emergence, which is crucial from a theoretical point of view.

Second, different theoretical assumptions exist about the developmental relations of action understanding and communicative understanding in the first year of life. One view suggests that these major social-cognitive skills emerge through an underlying understanding of others as goal-directed and perceiving organisms (Tomasello et al., 2005). In line with this view this view, one would expect relations between skills in action understanding and communicative understanding. Other views argue that action understanding and communicative understanding develop as separate and independent systems (teleological vs. referential system; Csibra, 2003). In line with this view, one would expect skills for action understanding and communicative understanding to be initially unrelated (e.g. Csibra, 2003). Another possibility would be that both abilities are only related to a general level of cognitive functioning (e.g. Grossmann, 2015; Heyes, 2016). However, for all theoretical views empirical evidence is rare.

As evidence for one integrated understanding Carpenter et al. (1998) showed that different social-cognitive abilities, like joint engagement, attention following and imitation are related in infants' emergence around their first birthday. Although these abilities manifest at slightly different time points, there are correlations between the ages of emergence of the different skills and the majority of the infants showed the same developmental order. In addition, there were no relations to non-social object-related abilities, which speaks for a specific development of social-cognitive abilities. Also, others found relations between very early triadic interaction (6-7 months) as well as action understanding at 10 months (Brandone et al., 2019) or relations between communicative abilities and action understanding in the beginning of the second year of life (Colonnesi et al., 2008). However, there is a lack of systematic evidence in the second half of the first year of life, where most of these basic social-cognitive abilities emerge.

Evidence for separate systems of action and communication understanding has been derived on the basis of studies suggesting dissociation between these abilities (for a review see Csibra, 2003). For example, great apes can attribute goals to actions, but do not seem to understand the referential nature of gaze. In addition, children with autism seem to have problems with the referential system (e.g. no sensitivity to direct eye-gaze), but not with the teleological system. Another indication for separate systems is that Yoon et al. (2008) found that 9-month-olds remember different aspects of an object when someone is pointing to it as opposed to reaching for it. However, in a recent study these results could not be replicated (Silverstein et al., 2019). Others argue that the lack of referential communication in apes or in children with autism is due to a motivational deficit in sharing perspectives and not due to a lack of understanding (e.g. Tomasello et al., 2005). Thus, empirical evidence for the emergence of these early social-cognitive abilities is mixed and incomplete. The absence of relations between different skills can of course not confirm their independence, but positive evidence would disfavor the theoretical approach that action understanding and referential understanding emerge as seperate, independent systems.

From a methodological point of view it is important to take into account the situation in which social-cognitive abilities are tested. Typically, infants are either involved in social interaction and have to react within this interaction or they are detached and observe a video/theater scene where their gaze behavior is measured. Some argue that infants learn through second-person interaction and improve their understanding through social interaction (Moore & Barresi, 2017), therefore understanding should be easier and perhaps emerge first in interaction-based tasks. Others argue that with the application of eye gaze measures it is possible to capture social-cognitive skills easier and therefore perhaps earlier than with interaction-based tests because they do not require motor or interaction skills (Baillargeon et al., 2016; Krogh-Jespersen & Woodward, 2016). Only few studies have systematically compared cognitive skills and the paradigms with which they are assessed. These studies suggest concurrent relations between interaction-based and looking-time measures (e.g. Brune & Woodward, 2007; Krogh-Jespersen et al., 2015). However, a systematic and longitudinal approach is necessary to capture concurrent and directional relations o further assess the appropriateness and the internal validity of different methods used in infant research.

In the current study, we took a longitudinal approach employing different paradigms to investigate the development of action understanding and communicative understanding. In a symmetrical design we employed eye-tracking and interaction-based tasks across both abilities. For action understanding, we used a "failed reaching" paradigm requiring infants to anticipate the goal of an action without having seen the goal. With the eye-tracking task, we measured predictive looks to the action goal (Brandone et al., 2014). In interaction-based task, we measured whether infants would hand objects to an experimenter who needs them or point to them for the experimenter (Liszkowski, Carpenter, & Tomasello, 2008; Warneken & Tomasello, 2007). For communicative understanding, we measured two different levels of referential understanding. A simpler level required infants to look at an indicated, visible object (point following). A more advanced level required infants to expect a hidden object at an indicated location (referential expectation). We measured infants' point following skills to visible objects in video stimuli and life interaction (Carpenter et al., 1998; Mundy et al., 2007). For referential expectation, we measured whether an indicated hiding location that did not contain a toy would yield a Violation-of-Expectation as revealed by a longer looking-time (Csibra, 2008). In an interactive task we measured infants' searching behavior for a hidden object in an indicated location (Behne et al., 2012). We also measured infants' general disengagement skills in an eye-tracking based overlap task (Elsabbagh et al., 2013), requiring them to first focus on a central stimulus and then orient toward an appearing peripheral stimulus. We used this as a control measure at the beginning of the study, testing whether early general attention orienting skills are related to social-cognitive abilities later in development.

We collected data monthly between the ages of 8 and 14 months. For each task, we began collecting at slightly earlier ages than the literature suggests in order to find competences on a group level to capture variance and pinpoint the individual emergence. We examined the skills monthly until 12 or 14 months, the age at which respective skills are assumed to be robustly present.

For each task, we analyzed at which month group-level performance exceeded chance patterns. If this reflected the emergence of stable competence, we expected positive intra-task correlations across months. From the time point at which we found stable competences within a skill, we examined relations between skills. If an underlying understanding is responsible for the emergence of different social-cognitive skills, we would expect correlations between action understanding and communicative understanding. We expect relations between methods within each skill if there is internal validity between them.

4.3 Methods

The current study was part of a larger international project on the socio-cultural and social-cognitive development of infants during the first and second year of life. Data collection took place in Hamburg, funded by the BMBF (Bundesministerium für Forschung und Bildung), and at the Koç University in Istanbul, Turkey, funded by the Tübitak (Turkish Science and Research Agency). The focus of the current paper is on the development of social-cognitive abilities. Thus, we present only tasks measuring social-cognitive abilities and only data collected in Germany, to be able to illustrate and analyze it in detail. Measures of social interaction are reported in a different paper and an additional paper is planned in which both lines are integrated in an overall broader analysis, including tasks from different socio-cultural environments.

4.3.1 Participants

To represent the diverse population of Hamburg, the sample (N = 47) included a group of German families (n = 33, female = 15) and a group of families with a Turkish migration background (n = 14, female = 6). All German as well as half of the Turkish-German samples were collected using an existing data base of families who had given prior consent to participate in developmental studies. To increase the Turkish-German sample, recruitment was extended to cultural institutions, canvassing in selected neighborhoods, attending cultural festivals, a fFacebook campaign etc.

Families were invited to the laboratory monthly between the infants' age of eight and 14 months, and additionally at 18 months of age. We allowed infants' age to range to be within seven days before or after they attained the respective monthly age. From 10 months on some parents did not attend some session (e.g. due to illness or holidays) and four families stopped participating in the study (one when the infant was 11 months old and three when the infant was 12 months old). Infants missing per months amounted to $N_{8-months} = 0$, $N_{9-months} = 0$, $N_{10-months} = 6$, $N_{11-months} = 4$, $N_{12-months} = 5$, $N_{14-months} = 8$.

4.3.2 Set-up and Procedure

At the first appointment, parents were informed about the procedure and data privacy. After a ten-minute warm up with the infant, parent and infant participated in a five-minute observation situation that is not part of the current study.

Afterwards, they were asked to go to another room for eye-tracking. A Tobii x120 eye-tracker (Tobii Technology, Stockholm, Sweden) was positioned below a 1920x1200px resolution screen (52 x 34 cm) on which stimuli were presented. The refresh rate of the screen was 60 Hz with a response time of 13.5 ms. The stimulus presentation screen and the computer on which Tobii Studio was running were connected with a DVI-D (18+1) cable. The hardware components of the computer had the recommended specifications (Intel Core 2.60GHz CPU, 16GB RAM, Nvidia Graphic Card, SSD 500GB Hard Disk). Another screen, not visible for the infant, was connected to the Tobii Studio Computer to monitor calibration and stimulus presentation. The multi-screen set-up was in the extended mode. The presentation screen was surrounded by a black canvas. The canvas was about 2.5 meters high, 1.5 meters in width and had triptych-like wings to the left and right side to minimize distraction for the infant. Above the screen, a webcam monitored the behavior of the infant during the session. Infants sat in a safety car chair in front of the screen at a distance of about 62-65 cm. Parents were seated next to the infant and were told not to interact with the infant or direct the infant's attention to the monitor during the session. In case infants refused to sit in the safety chair, they were allowed to sit on parents' lap on a swivel chair. The parents wore opaque glasses during calibration and video presentation so the parents' gaze would not be recorded by the equipment. A five-point calibration with rotating balls and infant-friendly music was used. The eye-tracking session lasted around 3-5 minutes, with three tasks presented at each time point.

After the eye-tracking session the parents and their infants where sent to another test room where the interaction-based tasks were conducted. The room's dimensions were was 3.40m by 3.80m long. All walls were white or covered with white cloths to reduce distraction. Infants were sitting on their parent's lap opposite the experimenter. An 80cm by 80cm desk was between them. Four cameras were positioned on tripods around the table, close to the walls to record the infant and experimenter simultaneously. Two cameras filmed the infant from the front (left and right), the other two filmed experimenter from the front (left and right).

4.3.3 Measures

We measured three different skills (action understanding, point following, and referential expectation) using two different methods (eye-tracking and interaction-based measures), leading to six different tasks. Additionally, we conducted a non-social disengagement task with eye-tracking as a control variable for general attention capacity.

4.3.3.1 Eye-Tracking Tasks

Eye-tracking data were sampled with 120 Hz. Due to technical changes in the beginning and in the end of the study, for four infants at seven appointments the eye-tracker was sampling with 60 Hz. We considered this in the data analysis. For all eye-tracking tasks, we defined areas of interest (AOIs) with a frame of 1 - 2°, dependent of the surrounding AOIs. This is recommended for messy infant data (Dalrymple, Manner, Harmelink, Teska, & Elison, 2018). In addition, we used the Tobii IVT fixation filter with the following settings: The data were filled (interpolation) for smaller than 75 ms time windows. A three-sample median filter was used for the 120 Hz sampling rate but not for 60 Hz in order to standardize the different sampling rates. A fixation had to have a velocity threshold below 30°/s measured in a 20 ms time window and a minimum duration of 60 ms. To identify the fixation in the correct area of interest (AOI), 67 ms had to be detected within a 100 ms time window. The mean position of the eyes were calculated if the left and right eye were detected and had a good validity (indicated by 0 in Tobii Studio). If the eye-tracker only detected one eye, the position of this eye was used in case also only one was visible during calibration.

Action Understanding. Stimuli of the eye-tracking task were similar to the failed reaching task in Brandone et al. (2014). The video stimuli were edited in Adobe After Effects CS6 and converted to 15fps WMV1 codec videos. In the videos of the size of

30.7x24.6°(1280x1024px) a person (18.6x7.5°) reached over a white barrier (6.5x1.5°) for a red ball (2.2x2.2°). Both items were placed on a light gray table (20.0x7.8°). The distance between the person and the ball was 14.2°. (for a chronological illustration see Figure 4.1). We presented four trials in form of four identical videos with small attention grabbing videos of three seconds in between.

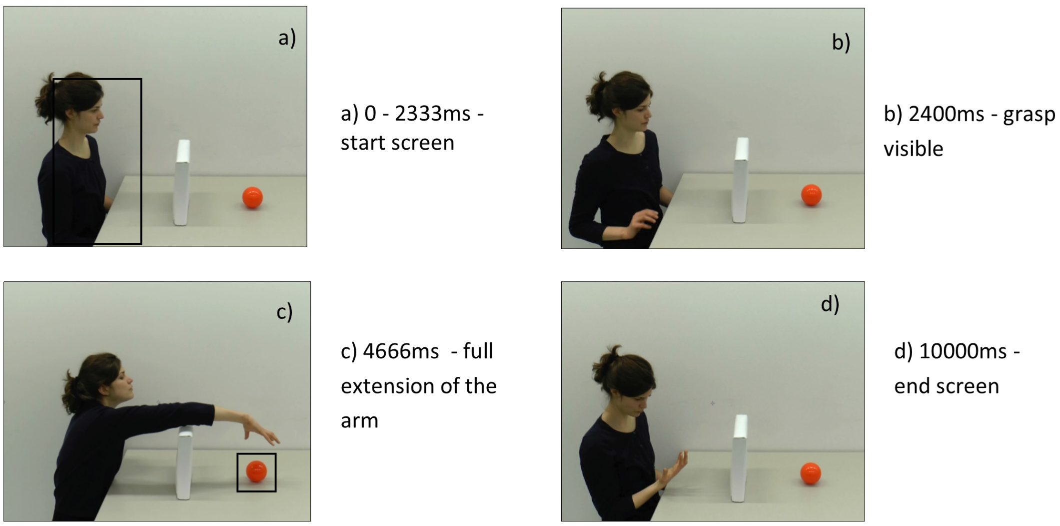

Figure 4.1. Schematic display of the video stimuli used in the "Failed Reaching" task. The black frames in a) represent the AOI of the person and in c) the AOI of the ball.

We defined three AOIs: around the person's front, around the ball and around the whole video (see size and position of the AOIs in Figure 4.1). To exclude the option that a fixation on the ball was an inaccurate fixation on the hand, we defined the ball AOI very close around the ball (0.75°). Between the upper end of the ball AOI and the hand in the fully extended positon was a space of 1°according to the accuracy of the eye-tracking system. All infants who had an accuracy of >1°were excluded from analysis.

Trials were excluded when infants watched less than 50% in the sequence from 2600 ms (start of latency measures) until 7130 ms (end of latency measure). This was the case in 1.5 – 9.8% of the trials across sessions and infants. Furthermore, trials were excluded when infants did not look at the person for at least 100 ms in the starting scene until 2600 ms (0 – 2.7% of the trials across infants and sessions). As infants need around 200 ms to initiate a saccade (Gredebäck, Johnson, & von Hofsten, 2010), the latency measure started 200 ms after the grasp was visible (at 2600 ms). Infants had 2066 ms time to initiate an anticipatory saccade until the full extension of the arm. They had the same time to initiate a reactive saccade from 4666 ms until 6733 ms. Latency was measured until the first fixation to the ball. Dependent variables were the anticipatory looks, reactive looks, and no looks to the ball, all as ratio of valid trials. As additional information, we measured mean latency, starting from the cut-off criterion. Negative values represent anticipatory looks and positive reactive looks.

Some infants had to be excluded due to fussiness ($N_{8-months} = 2$, $N_{9-months} = 3$, $N_{10-months} = 2$, $N_{11-months} = 2$, $N_{12-months} = 4$) or calibration problems ($N_{8-months} = 5$, $N_{9-months} = 6$, $N_{10-months} = 4$, $N_{11-months} = 8$, $N_{12-months} = 6$). Additionally, we decided to present the tasks to 11-month-olds after the data collection had already begun. Thus, at 11 months, we had not presented the tasks to six infants.

Point Following. In this eye-tracking task, we measured infants' point following behavior using video stimuli of the size of 1280x1024px. Videos were edited in Adobe After Effects CS6 and converted to 25fps WMV1 codec videos. We used four pairs of unfamiliar objects (4.5°x3.5°) and matched them according to color and salience.

Both were presented as target and distractor between infants. We presented them for 5360 ms without a person to give infants enough time to process them, while a curtain hid the actor. At 4000 ms an expressive “Ah!” was audible from the back. Then, the curtain (with a different color for each trial) went up for 680 ms and revealed a person (12.4°x9°) sitting behind a table. The person looked straight ahead to establish eye-contact. Then, the actor pointed to one of the objects for 6880 ms by simultaneously looking to the object and making one gaze alternation to the infant. The pointing gesture was visible from 7000 ms on until the end of the video at 13 s and 800 ms. The distance between the person’s finger and the closest corner of the object was 7°. The distance between the objects was 18.6°. We presented four videos, each representing a trial where the person was pointing either to the left or the right object. The presentation order was pseudo randomized with never pointing to the same side for more than two consecutive trials (for a selected scene of the pointing gesture see Figure 4.2).

We defined five AOIs: one around each object (target and distractor) and three around the person (head, left hand, right hand). For size and position of the AOIs see Figure 4.2. For a valid trial infants had to look at the AOI of the person for at least 1000 ms while the pointing gesture was presented. This was the case in 1-5% of the trials across infants and sessions. The dependent variables were infants’ ratio of first looks to the target, looks to the distractor, and no look to the objects. Therefore, the latency to target and distractor were measured, starting 200 ms after the pointing gesture was recognizable (at 7200 ms) until the first fixation to the object. Infants had

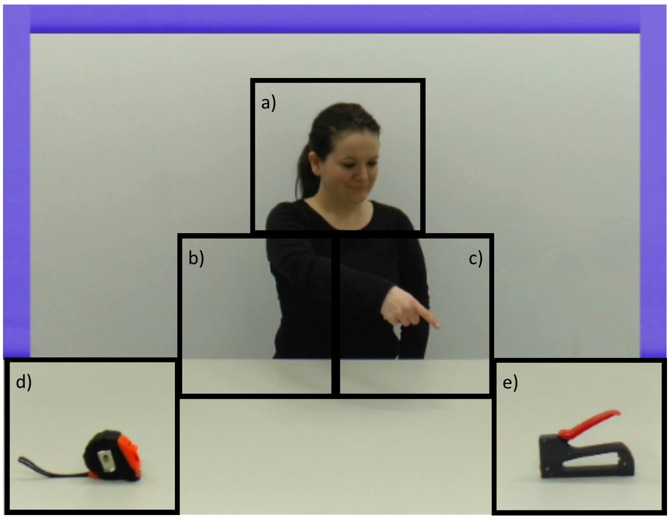

Figure 4.2. Selected scene of the pointing gesture in the point following task for eye-tracking. Black frames represent the AOIs. A – C summarized to the AOI for the person. D and E were either target or distractor.

time to look to the object until the end of the video. Trials in which the latency to the target was shorter than to the distractor were coded as 1. If infants looked faster to the distractor or did not look at the objects at all, trials were coded as 0. The sum of first looks to the target was relativized on valid trials.

We had to exclude infants due to fussiness ($N_{9-months} = 2$, $N_{10-months} = 1$, $N_{11-months} = 2$, $N_{12-months} = 3$) or calibration problems ($N_{9-months} = 1$, $N_{10-months} = 3$, $N_{11-months} = 4$, $N_{12-months} = 5$). At the first three appointments of the study, the task was not ready ($N_{9-months} = 2$, $N_{10-months} = 1$).

Referential Expectation. With this eye-tracking task, we measured infants' referential expectation when someone points to a hidden entity. For this, a person pointed behind one of two boxes which revealed an object either on the congruent or the incongruent side. The video stimuli were the same as in the communicative condition of Study 1. For a detailed description of sizes, timing, AOIs and presentation order see Stimulus and Design of Experiment 1 in Study 1. A difference of the current experiment was that, following the advice of Tobii-Studio, rendering was set to 25fps (instead of 15fps), so that every sequence happened 120 ms earlier than in Experiment 1 of Study 1. In addition, to account for possibly longer processing time overall, the outcome phase was extended for 1 s to 4800 ms.

Trials were excluded when infants looked at the video for less than 1000 ms in the cueing phase (1.89 – 11.02% of the trials across participants and appointments) and less than 60 ms at the fixation cross during its presence (6.54 – 22.36% of trials across participants and appointments). The dependent variable was a difference score of total dwell-time (dwell-time to the empty side of the occluder in incongruent trials minus dwell-time to the empty side in congruent trials). The dwell-time was calculated by the sum of gaze registrations while the object was visible and converted the gaze registrations into milliseconds by multiplying them with the duration of one frame (1 frame = 1 s/120 Hz = 8.333 ms).

Non-Social Disengagement Task. Stimuli of this eye-tracking tasks were pictures of animated animals with the size of 3x3°on a black background screen. A central stimulus expanded and contracted and was accompanied by different sounds to attract the infants' attention to the center. After two seconds, movement and sound stopped and a similar salient picture of an animal was presented as a peripheral stimulus. Both stimuli were presented for five seconds, followed by a 1000 ms inter-trial interval of a black screen. The peripheral stimuli were presented pseudo-randomized at the right or left side of the central stimulus (not more than two times on the same side) with an eccentricity of 10°. This distance was selected for infants to account for low precision and accuracy. AOIs were set around the central and peripheral stimuli.

Two dependent variables were calculated: latency and correct refixations. For the latency measure the time to first fixation in the peripheral AOI was calculated after the peripheral stimulus was presented. Latency was not calculated for trials in which the infant was looking away from the screen for more than 500 consecutive ms after the latency measure started (i.e. after 2000 ms of the trial) or did not look at the central stimulus for 100 ms immediately before the peripheral stimulus appeared (i. e. 1500-2000 ms of the trial). For the correct refixation measure we summed up those trials in which a fixation on the peripheral stimuli happened and divided it by valid trials. Infants had to have at least two valid trials to be included in the analysis.

4.3.3.2 Interaction-based Tasks

Action Understanding. In this interaction-based task, two different tests were conducted, each involving two trials. One test was an out-of-reach helping task adapted from Warneken and Tomasello (2007) in which items were out of the experimenter's reach and infants' helping behavior was tested at 10, 12 and 14 months of age. The other one was a hidden task, where items were hidden for the experimenter and infants' informing behavior was tested at 12 and 14 months. We used different items each month to decrease transition and memory effects (wooden blocks at 10 months, paper balls at 12 months and erasers at 14 months of age). For the hidden trials two differently colored barriers were positioned on the table (see Figure 4.3).

At the start of the helping test, Experimenter 1 (E1) put all items on the table and presented them to the infant. Then E1 went out of the room to get a box. During their absence, another experimenter (E2) entered the test room and complained about the dirty desk. While cleaning the desk and talking to infant and parent, E2 positioned the targets inconspicuously as displayed in Figure 4.3.

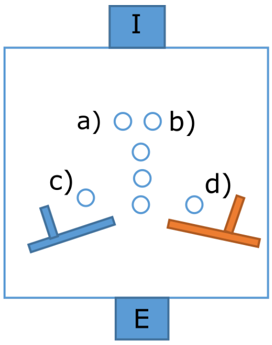

Figure 4.3. This figure displays the set-up of the interaction-based action understanding task, including out of reach and hidden trials. a) and b) are items for the out-of-reach task and they were out of the experimenter's (E) reach but in the infants reach (I). c) and d) are the items for the hidden trials and were out of the experimenters sight but visible for the infant. At 10 months, barriers and item c) and d) were not used.

After cleaning, E2 left the room and E1 reentered with the box, wondering about the new positions of the items. E1 sat down, holding the box between itself and the table and collected the three items within reaching distance, starting with the closest and commenting that the items have to be put into the box. These three items not marked in Figure 4.3 were used as familiarization trials for the out-of-reach task. For the last two test items, E1's hand stopped close to the target. For 10 s, E1 just looked to the items by moving body and arm to try to reach them. If the infant did not hand the object over, in the following 10 s E1 said: "I can't reach it. I can't reach it." while still acting as in the previous 10 s. For the last 10 s, E1 alternated the gaze between the infant and the target, calling the infant by its name and repeating the phrase from the previous 10 s. E1 always directed the reaching gesture toward the target, even when the targets were not on the table anymore (e.g. in the infants hands). E1 took the object only when the infant released the object. The trial ended when the infant handed the object over or 30 s had elapsed. If the infant did not hand the object over, the second trial started in the same way as the first one.

After the two out-of-reach trials, two hidden trials were presented. Therefore, E1 counted the objects in the box, looked surprised, put her hands palmed up next to her torso and searched for other items around the box or under the table. Again, E1 did this for 10 s without looking at the infant, then, for 10 s saying "There are some missing still! Where are they?" ("Da fehlen noch welche! Wo sind die denn?") and another 10 s while repeating the same phrase with gaze alternations between the infant and the table. This procedure was done for both hidden items separately, counting as two trials. A trial ended when the infant pointed to the object or 30 s had elapsed. If the infant did not help, the second trial started in the same way as the first one.

For out-of-reach trials, the behavior was coded as correct helping behavior if the infant handed over or slid the object intentionally to E1. Behavior was coded as intermediate helping behavior if the infant pointed to the objects or held the item in its hand and presented it to E1 but did not release the object. No helping behavior was coded if the infant did not reach for the items at all or held them in their hands not trying to pass the object to the experimenter. For hidden trials, behavior was coded as correct helping behavior if the infant pointed to the hidden object either using the index finger or the whole hand. Intermediate helping behavior was coded if the infant looked to the object and E1, handed the item over, slid the item towards the experimenter, or reached for the items. All other behavior was coded as no helping. 6% of all the trials in both tasks were coded by a second coder. Kappa for out of reach task was $\kappa = .80$ and for hidden trials $\kappa = .88$.

At the 12-months session, two infants had to be excluded because of fussiness. In all other sessions, no infant had to be excluded. Some trials were coded as error trials when infants were not able to reach the item because they threw it away before the trial started or the parents put the items away. At 10 months, 11% of the trials were coded as error trials, at 12 months 8.9%, and at 14 months 5.8%.

Point Following. In the interaction-based point following task, we measured infants' point following behavior in four trials. On each side of the infant, at a distance of 180 cm, two 30x20 cm pictures depicting animals were attached to the wall at a height of 150cm. Two pictures were positioned in the infant's visual field (front pictures), while the other two pictures were positioned slightly behind the infants visual field (back pictures). For the front pictures, infants had to turn their head 45°from looking straight to E1. For the back pictures, they turned 100°. E1 called the infant's name and made eye-contact. Then, E1 looked and pointed to one of the four pictures combined with an expressive "There!" ("Da!"), lasting one second. A gaze alternation followed while pointing to the target for another three seconds. After the point gesture was retracted, E1 waited another five seconds to let the infant react. E1 pointed to the front pictures first and then to the back pictures. The pointing side was pseudo-randomized. Infants were randomly assigned to one of two orders starting either with the left side or the right side and every infant had the same order over all appointments.

The infant's first look was coded, indicated by a stop of the eyes for at least one second. Possible specifications were target, distractor same side, distractor other side, and no choice. In some of the back-picture trials, the infant's first look stopped at the front picture but then they looked to the back target immediately afterwards. Thus, we added a specification: back picture after distractor same side. 20% of the videos were coded by a second coder showing a Kappa of $\kappa = .83$ and matches ranging from 75 – 100%.

One trial had to be excluded because the parent interfered and pointed for their infant. Two infants were excluded at the 9-months appointment due to recording problems and agitation. One infant was excluded at 10 months because the infant was too restless.

Referential Expectation. In this task interaction-based task, we measured infants' understanding of pointing gestures to hidden targets. We used a wooden board (80cm long by 20cm wide) with felt underneath to silently slide it on the table. On each end of the board, we mounted a cardboard box with sponge rubber on the bottom, so placing an object there was soundless. The box was covered by a small cloth to hide the object. The left and right wings of the box formed a triangle with the top oriented to the infants so they could easily see inside the box and grab the object after lifting the cloth.

We presented two familiarization trials and six test trials. In the first familiarization trial, E1 put a small toy in one box and slid the board toward the infant. In the second familiarization trial, E1 did the same with the other box but covered the object with a cloth beforehand. If the infant did not take the object by himself, E1 encouraged him to take the toy, by either showing the object or uncovering the box. We presented an additional third familiarization trial in case the infant did not uncover the box by themselves.

During the test trials, E1 covered each box with a cloth, presented a small object and hid it under one of the two boxes, without the infant knowing which one. E1 called the infant's name to get their attention, established eye-contact, and pointed to the box containing the object. E1 made one gaze alternation and started to slide the board toward the infant while pointing and looking to the box (5 s). When E1 stopped the board in front of the infant, E1 did another gaze alternation (5 s) and pointed and looked to the box for another five seconds. After that, E1 stopped the pointing gesture and the infant had another 10 s to react. If the infant did not take the toy, E1 put it in front of the infant and slid the board back to herself. E1 covered the boxes with different colored cloths and the next trial started after E1 took the previous toy from the infant.

5% of the videos were coded by a second researcher. Coding included which cloth the infant took away first (correct, incorrect, both, none), K = .891, matching 96.30 – 100% and, if the infant was expecting and searching for an object on the indicated side (search yes/no), $\kappa = .74$, matching 93.18 – 100%.

Two infants were excluded at the 9-months appointment, one due to recording problems and one due to agitation. At the appointment at 10 months, one infant was excluded to agitation as well.

4.3.3.3 Analysis

We present our analyses in two main sections. In section one, as a first step, we analyzed each task separately. In section two, we looked at relations between tasks. Section one: for each task, we looked at (a) task performance at each time point; (b) intra-individual stability across time points; and (c) longitudinal change of performance across time points. In addition, we related each task to general disengagement and participant group.

For (a) task performance, we tested dependent variables against each other with dependent t-tests or against chance with one sample t-test (see specific analysis in each task). In addition, we calculated individual competencies by presenting the number of infants showing the relevant skill in at least 50% of the trials.

For (b) intra-individual stability, we calculated Pearson correlations between months. We used bootstrapping with 1000 samples and reported a 95% percentile confidence interval to verify our results also for not normally distributed variables.

For (c) longitudinal change, we tested for mean differences with an omnibus ANOVA and conducted separate dependent t-tests between months (these have more power than the omnibus ANOVA and reveal the exact time point of a significant change). Furthermore, we calculated linear growth models (LGM) for all tasks separately to test for a significant linear increase or decrease of the interested variable. A significant mean of the intercept would mean infants start on average significantly from zero. A significant and positive mean of the slope would signify that the skill improves over time. A significant variance of the slope would mean that infants' developmental patterns are heterogeneous and significantly different from each other. A significant variance of the intercept would mean that infants start on different levels. If the model fit was not good for an overall linear growth model, a piecewise linear growth model was calculated. A good model fit was indicated by a not significant Chi^2, RMSEA $\leq$ .06 (moderate fit $\leq$ .08), $CFI \geq$ .95 (moderate fit $\geq$ .90), TLI $\geq$ .95 (moderate fit $\geq$ .90).

Additionally, we examined relations between social-cognitive tasks and the non-social disengagement task. When associations were found, we did some control analyses when calculating relations between different social-cognitive tasks. In addition, we tested whether we could find skill differences between different participant groups (German and German-Turkish migrant families) for descriptive reasons. The sample size was too small for analyses wihtin each group.

As our main step in section two, we analyzed relations between tasks. Here, we looked at synchronous and predictive relations (a) between skills (action understanding, point following, and referential expectation) within each method and (b) between methods (eye-tracking and interaction-based tasks) within each skill. For (a), we looked at relations between (i) action understanding and point following, (ii) action understanding and referential expectation, and (iii) point following and referential expectation within eye-tracking tasks and within interaction-based tasks separately to see whether the tasks tap related abilities within each method. For (b), we related (i) action understanding, (ii) point following, and (iii) referential expectation between eye-tracking and interaction-based tasks to test whether the two methods tap the same ability.

For all relation-analysis, we calculated concurrent and predictive correlations between skills, starting with the month of onset for each skill. We used the advantage of the longitudinal design and defined a starting point of an emerging skill when task performance was significant and skills were stable, indicated by significant relations between months. We used these indications to ensure enough variability as well as a stable age of onset. We calculated the Bonferroni alpha correction by multiplying the number of time points from the starting points of each skill between two tasks before dividing the p-value by the product. We did this because we expect meaningful relations only after the starting point within each skill when abilities start to be stable and meaningful variable. We tested one-tailed, so we used a p-value of $p = .100$, as we expected relations always in a specific direction: that is a better performance in one task relates to a better performance in the other task. We calculated Pearson correlations with 1000 bootstrapped samples and reported a 95% percentile confidence interval to verify our findings also for not normally distributed variables.

When we found a significant correlation between tasks, we did some control analyses. First, if there was stability within one skill, we controlled for competencies in previous months to capture the initial point of relations. Second, we controlled for general disengagement skills when we found relations to one of the social-cognitive tasks. Both control analyses were done with partial correlation.

All analysis were calculated in IBM SPSS 25. Only LGM was calculated in MPlus Version 6. -since at least one variable was not normally distributed, MLR estimator was used, because it is robust against skewed data.

4.4 Results

4.4.1 Separate Task Analysis

4.4.1.1 Action Understanding - Eye-Tracking

Task performance. Dependent t-tests between anticipatory and reactive looks were only close to significance only at 12 months of age ($t(31) = 1.72$, $p = .095$, $d_z = 0.56$). In all the preceding months, we could not find a significant difference (8 months: $t(39) = -1.16$, $p = .252$, $d_z = 0.28$; 9 months: $t(37) = -0.86$, $p = .393$, $d_z = 0.21$; 10 months: $t(34) = -1.30$, $p = .203$, $d_z = 0.34$; 11 months: $t(26) = 1.32$, $p = .200$, $d_z = 0.42$). The means are displayed in Figure 4.4.

On the individual level, until 10 months significantly more infants showed anticipatory looks in less than 50% of the trials and the binomial test was significant (8 months: 70%, $p = .017$; 9 months: 74%, $p = .005$; 10 months: 69%, $p = .041$). From 11 months on, the number of infants who showed $\geq$ 50% anticipatory looks was equally distributed to infants showing < 50% anticipatory looks and binomial test was not significant (11 months: 48%, $p = 1.00$; 12 months: 59%, $p = .377$).

One-sample t-tests for mean latency (mean time to first fixation on the ball) was significantly different from zero at 8 months in a positive direction (reactive looks) and no significant difference in other months (8 months: $t(32) = 2.23$ $p = .033$, $d = 0.39$; 9 months: $t(29) = 0.64$, $p = .642$, $d = 0.09$; 10 months: $t(28) = 0.24$, $p = .813$, $d = 0.04$; 11 months: $t(22) = -1.17$, $p = .256$, $d = 0.24$; 12 months: $t(30) = -1.31$, $p = .201$, $d = 0.23$). Means were descriptively negative form 11 months on, indicating anticipatory looks (Figure 4.5).

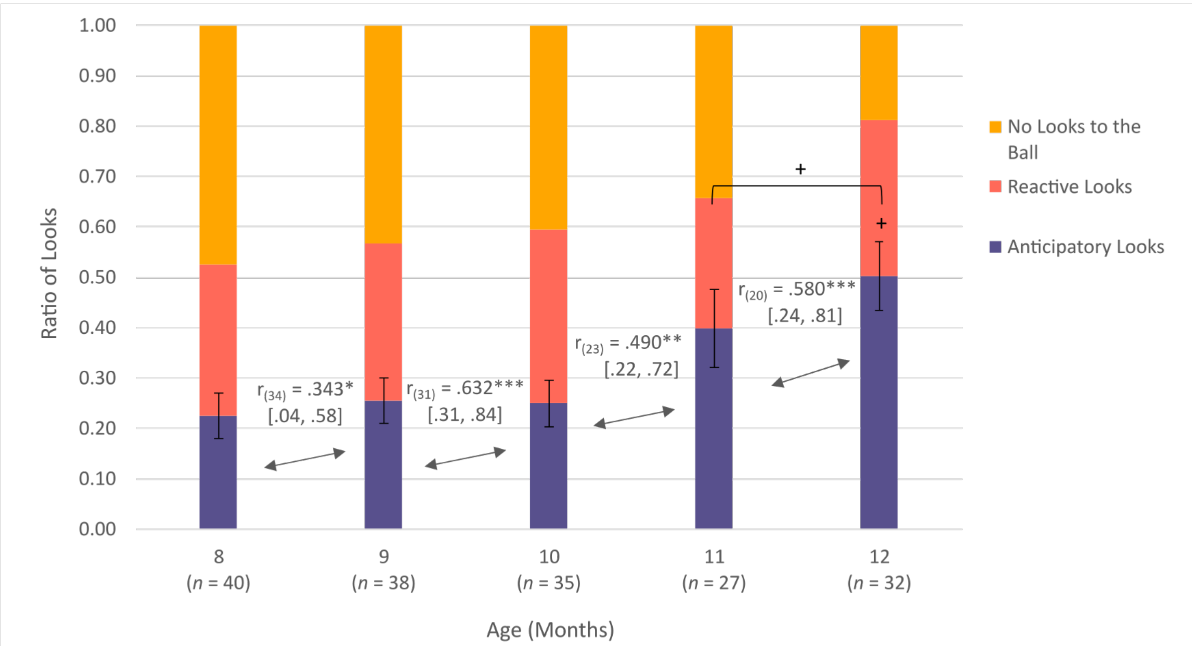

Figure 4.4. Graph displays ratio of anticipatory, reactive and no looks to the ball. Pearson correlation coefficients are calculated between anticipatory looks. Bootstrap 95% CI is presented in brackets. Asterisks in the bars display significance of separate dependent t-test between anticipatory and reactive looks. Asterisks between bars are based on separate dependent t-tests between months. Error bars display mean standard error +/- for anticipatory looks. $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$

Intra-individual stability. Concerning the ratio of anticipatory looks we found strong significant correlations between months from 8 months on (see correlation coefficients in Figure 4.4). Concerning mean latency, we found significant correlations between 11 and 12 months (see correlation coefficients in Figure 4.5).

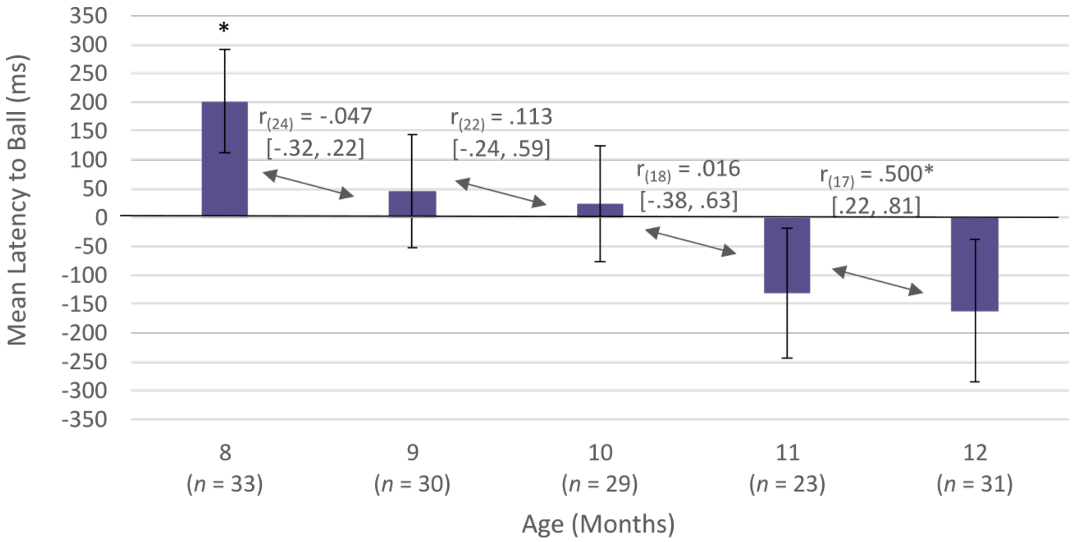

Figure 4.5. Graph displays the mean latency to the ball. Zero represents the cut off criteria for anticipatory looks (full extension of the arm). Negative values display anticipatory latency, and positive values reactive looks. Correlation coefficients are based on Pearson correlation and Bootstrap 95% CI is presented in brackets. Asterisk display single sample t-tests against zero. Error bars display mean standard error +/- for mean latency. P-values: $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$

Longitudinal change. Repeated measurement ANOVA showed a significant difference of anticipatory looks between measuring time points, $F(4, 52) = 4.12$, $p = .006$, $\eta_p^2 = .241$. T-test between months revealed only a marginal significance between 11 and 12 months ($t(19) = -1.91$, $p = .072$). All others were not significant (8 vs. 9 months: $t(33) = 0.00$, $p = 1.00$; 9 vs. 10 months: $t(30) = -0.24$, $p = .811$; 10 vs. 11 months: $t(22) = -1.34$, $p = .196$). Measuring mean latency in a repeated measurement ANOVA showed no significant difference between time points $F(1.80, 16.12) = 2.44$, $p = .123$, $\eta_p^2 = .213$.

The linear growth model, showed a linear growth over time for ratio of anticipatory looks from 10 months on (M_{slope} = .060, $p < .001$) but not between 8 and 10 months (M_{slope} = .005, p = .826). A piecewise linear growth curve model showed a good fit ($Chi^2_{(6)}$ = 2.754, p = .839, p = 14; RMSEA < .000, CI [0.000-0.110]; CFI = 1.000; TLI = 1.160; SRMR = .062). The relations between intercept and slopes were not significant ($r_{8-10months}$ = -.005, p = .614; $r_{10-12months}$=.002, p = .710), showing that there was no influence of the initial level on the increase. The mean intercept was significant ($M_{intercept}$ = .204, $p < .001$), indicating a ratio of anticipatory looks significantly different from zero at the initial level at 8 months. A significant variance of intercepts ($V_{intercept}$ = 0.033; p = .019) indicated a significant inter-individual difference in the starting level. The variance for the slopes from 8 to 10 months was not significant (V_{slope} = .012, p = .184), and close to significance from 10 to 12 months (V_{slope2}=.004, p = .077), indicating a non-significant difference of growth between infants.

In summary, from 11 months on we find stability in the mean latency to fixate on the goal of the reaching action as well as a significant increase. Thus, from 11 months on, infants become faster in anticipating others' reaching actions. In addition, we found stability in anticipatory looks from 8 months on. Thus, action anticipation seems to be present from early on but task performance on group level is only present near the end of the first year.

4.4.1.2 Point Following - Eye-Tracking

Task performance. T-tests between the ratio of first look to target and first look to distractor were significant for 9, 11 and 12 months of age (9 months : $t(41) = 2.58$, $p = .014$, $d_z = 0.50$; 11 months: $t(36) = 3.07$, $p = .004$, $d_z = 0.74$; 12 months: $t(33) = 4.86$, $p < .001$, $d_z = 1.31$), but not for 10 months (10 months: $t(36) = 1.10$, $p = .279$, $d_z = 0.30$). Means and SE are displayed in Figure 4.6.

On the individual level, more infants showed $< 50\%$ first looks to the target at 9 and 10 months and the binomial test was significant (9 months: 69%, $p = .020$; 10 months: 68%, $p = .047$). At 11 and 12 months, the number of infants who followed the pointing gesture correctly in $\geq 50\%$ was equally distributed with infants who followed the pointing gesture $< 50\%$ of the trials and the binomial test was not significant (11 months: 41%, $p = .324$; 12 months: 62%, $p = .229$).

Intra-individual stability. We found high correlations between 11 and 12 months, and not between others. See correlation coefficient in Figure 4.6.

Longitudinal change. A repeated measurement ANOVA for the ratio of first looks to target showed a close to significant difference between all measuring time points, $F(3, 63) = 2.55$, $p = .064$, $\eta_p^2 = .108$. Separate dependent t-tests between months were significant between 11 and 12 months ($t(27) = -2.48$, $p = .019$) but not between 9 and 10 months ($t(34) = 0.43$, $p = .673$) or between 10 and 11 months ($t(29) = -1.03$, $p = .310$).

A linear growth model showed a poor model fit ($\text{Chi}^2_{(5)} = 7.007$, $p = .220$;

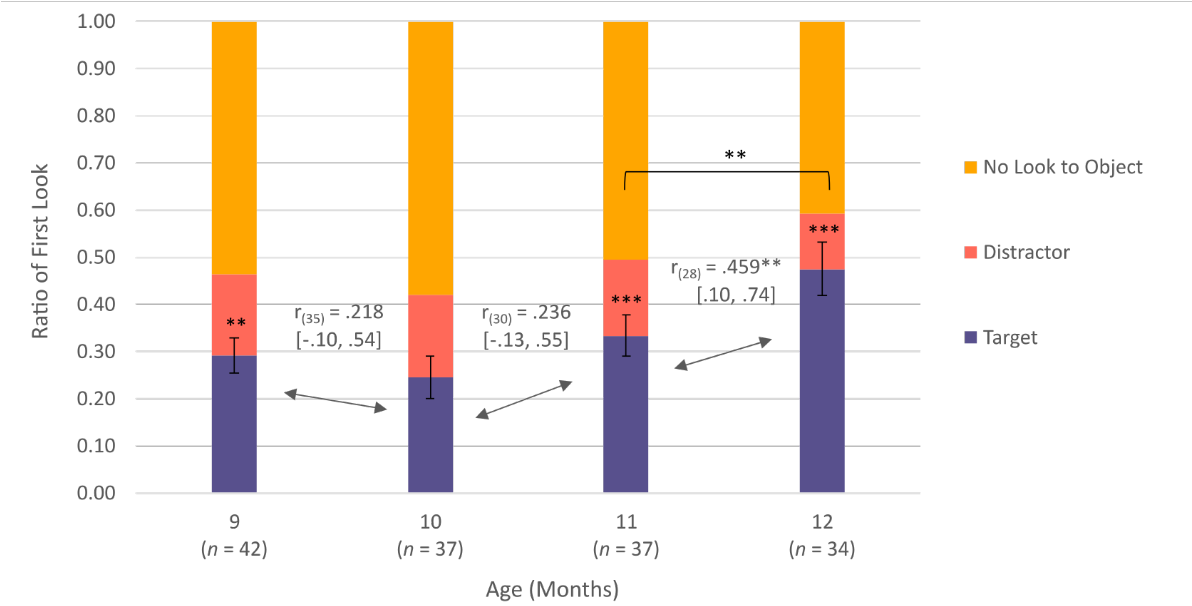

Figure 4.6. Figure displays the mean number of first looks to the target, distractor and no look to either object per valid trial. Correlation coefficients are based on Pearson correlation and Bootstrap 95% CI is presented in brackets. Separate dependent t-tests were calculated between the ratio of first looks to the target versus distractor. Error bars display standard error +/- of mean number of first look to target per valid trial. P-values: $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$

RMSEA = .096; CFI = .712; TLI = .654) because the value of correct point following was slightly higher with 9 months compared with 10 months (see mean and SE in Figure 4.6). A piecewise linear growth model showed a better fit ($Chi^2_{(1)} = 0.279$, $p = .596$, $fp = 8$; RMSEA < .000, CI [0.000-0.322]; CFI = 1.000; TLI = 1.250; SRMR = .026). There we found a significant linear growth from 10 to 12 months ($M_{slope} = .109$, $p < .001$) but not in the months before. The mean staring level from 10 months on was close to significant ($M_{intercept} = .120$, $p = .062$). Intercept and slope were not significantly different between infants ($V_{intercept} = .006$, $p = .906$; $V_{slope} = .005$, $p = .714$) and did not correlated (r < .000, $p = .999$).

We found a reliable point-following skill in the eye-tracking task from 11 months on, indicated by task performance, stability between months and a significant growth from 10 months on. Additionally, we found task performance at 9 months of age but this skill was not related to other months and we did not find it with 10 months. This early effect could be an indication for another skill, like a simple attention-orienting ability, while a profound seem to start only from 11 months on. This will be discussed in subsequent sections.

4.4.1.3 Referential Expectation - Eye-Tracking

Task performance. Dwell-time difference score was positive and significantly different from zero at 11 months ($t(31) = 2.18$, $p = .037$, $d = 0.38$) and 12 months ($t(33) = 2.68$, $p = .012$, $d = 0.46$), indicating a longer dwell-time in incongruent trails. A single sample t-test was close to significance with 9 months ($t(36) = 1.97$, $p = .056$, $d = 0.32$) but not significant at 8 ($t(39) = 0.74$, $p = .466$, $d = 0.12$), and 10 months ($t(35) = 1.06$, $p = .296$, $d = 0.18$). See means and SE of dwell-time difference in Figure 4.7.

On the individual level, for all ahes more infants looked longer to the empty side in incongruent trials than in congruent trials. However, at no month was the binomial test significant (8 months: 65%, $p = .081$; 9 months: (65%, $p = .099$; 10 months: 56%, $p = .618$; 11 months: 66%, $p = .110$; 12 months: 62%, $p = .229$).

Intra-individual stability. There was only one one-tailed significant relation for dwell-time difference score between 9 and 10 months. All other correlations were not significant. See correlation coefficient in Figure 4.7.

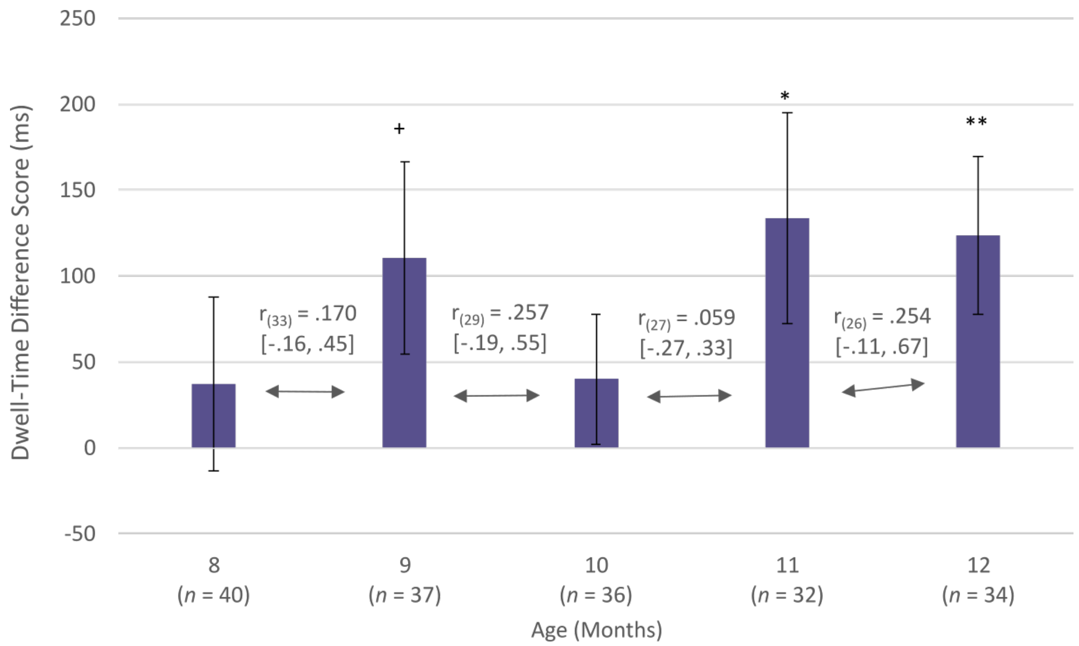

Figure 4.7. Graph displays dwell-time difference score. Positive values show a longer dwell-time to the empty box in incongruent trials than in congruent trials. Error bars display standard error +/-. Correlation coefficients are based on a Pearson correlation and values in brackets represent Bootstrap 95% CI. P-values: $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$.

Longitudinal change. A repeated measurement ANOVA for dwell-time difference score showed no significant difference between months, $F(4, 48) = 0.28$, $p = .888$, $\eta_p^2 = 108$.

For the dwell-time difference score a linear growth model showed a good fit ($\text{Chi}^2_{(10)} = 4.947$, $p = .895$, fp = 10; RMSEA < .000, CI [0.000-0.073]; CFI = 1.000; TLI = 1.000; SRMR = 0.121). The correlation between intercept and slope was not significant (r = -9239, $p = .121$). The mean intercept was not significant (r = 19.305, $p = .653$) while the mean slope was significant using a one-tailed approach (r = 28.012, $p = .098$). The variances of intercept and slope were not significant ($V_{intercept} = 24386$, $p = .137$, $V_{slope} = 3854$, $p = .123$). The linear line in the graph did not fit perfectly.

In summary, at 11 and 12 months of age, we were able to find significant dwell-time difference scores but we did not find stability between months. We also found a significant tendency at 9 months. On the individual level, from 8 months on, more infants looked longer in incongruent than congruent videos, which is one-tailed significant at 8 and 9 months. Thus, some ability seems to be existent in early months. However, valid referential expectation with high dwell-time differences seem to start with 11 months, although they are not yet stable.

4.4.1.4 Action Understanding – Interaction-based

Task performance. A dependent t-test was significant for 10 months ($t(40)$ = -4.64, $p < .001$, $d_z = 1.38$), revealing significantly more no helping behavior than helping behavior. For the out of reach tasks, the t-test approached significance ($t(38) = 2.02$, $p = .051$, $d_z = 0.61$) at 12 months and was significant at 14 months ($t(36) = 3.81$, $p < .001$, $d_z = 1.23$), showing significantly more helping behavior than no helping behavior. Infants did not show significantly more helping behavior in the hidden trials at 12 months ($t(37)$ = -0.37, $p = .711$, $d_z = 0.12$) or 14 months ($t(37) = 0.37$, $p = .711$, $d_z = 0.12$). See means and standard deviations in Figure 4.8.

On the individual level, we found a similar pattern. At 10 months of age, significantly more infants showed no helping behavior (71%) than at least one helping behavior (29%). The binomial test was significant ($p = .012$). However, from 12 months on significantly more infants showed at least one helping behavior in the out of reach tasks than no helping behavior (12 month: 74% helped, $p < .001$; 81% helped,

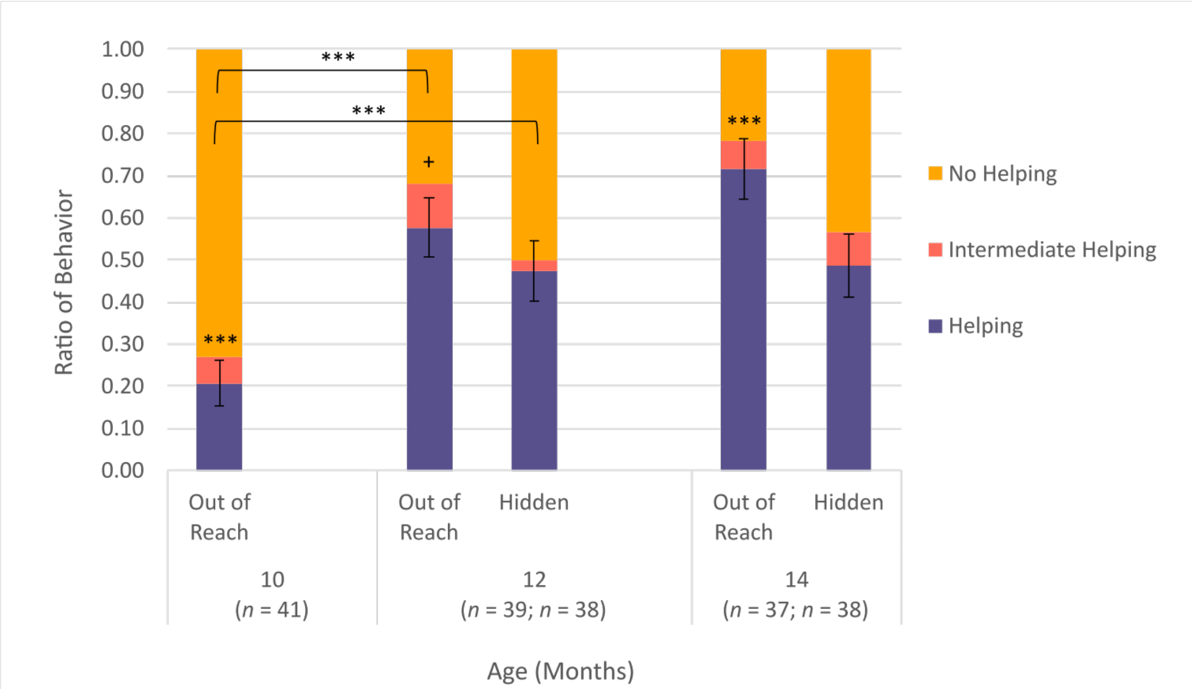

Figure 4.8. The graph displays the ratio of helping, intermediate helping and no helping behavior for action understanding in the interaction-based task. Asterisks in the bars represent t-test results between helping behavior and no helping behavior. Asterisks between bars represent t-test results for helping behavior between months. P-values: $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$.

$p < .001$). We could not produce these results for the hidden trials (12 months: 61% helped, $p = .256$; 14 months: 58% helped, $p = .418$).

Intra-individual stability. We found significant correlations between 10 and 12 months for the out of reach and hidden tasks, and no significant relations between 12 and 14 months. the correlation coefficients are listed in Table 4.3.

Longitudinal change. A repeated measurement ANOVA between sessions showed a significant main effect, $F(2, 56) = 9.32$, $p < .001$, $\eta_p^2 = .250$. In the out of reach tasks, separate dependent t-test between months were significant between 10 and 12 ($t(33) = $ -4.83, $p < .001$) but not between 12 and 14 months ($t(32) = $ -1.38, $p = .176$).

For the hidden trials, we found a significant difference from 10 to 12 months ($t(32) = -3.14$, $p = .004$) but not from 12 to 14 months ($t(32) = -.70$, $p = .496$).

In summary, for an easy way of helping (handing objects which are out of the other's reach), we found task performance at 12 months alreadyand a stable skill from 10 months on. In addition, infants who helped at 10 months were also the ones who started to point to hidden targets at 12 months. However, pointing out hidden targets seemed to be harder than handing the object over, as we do not find task performance at 12 or 14 months. Further, we did not find a stable helping behavior from 12 to 14 months in either task. This could be due to a change in the motivation to help, transferring from joy to be in interaction to prosocial motivation (Dahl & Brownell, 2019).

4.4.1.5 Point Following – Interaction-based

Task performance. For point following to front pictures, we found a significant difference between ration of looks to target and distractor from 9 months on (9 months: $t(44) = 3.03$, $p = .004$, $d_z =$; 10 months: $t(39) = 5.27$, $p < .001$, $d_z = 1.02$; 11 months: $t(42) = 8.12$, $p < .001$, $d_z = 1.46$; 12 months: $t(41) = 9.77$, $p < .001$, $d_z = 2.00$). At 8 months, the t-test was not significant ($t(46) = 0.61$, $p = .547$, $d_z = 0.09$). Means and SE are displayed in Figure 4.9 and reported in detail in Appendix Table A.1. On the individual level, 10 months was the first time point when the majority of infants followed the pointing gesture in $\geq 50\%$ of the trials and the binomial test was significant (10 months: 73%, $p = .006$; 11 months: 81%, $p < .001$; 12 months: 86%, $p < .001$).

At 8 and 9 months, the number of infants following the pointing gesture in $\geq$ 50% of the trials was either smaller than or equally distributed with infants who followed in < 50% of the trials (8 months: 30%, p = .008; 9 months: 53%, p = .766).

For the two trials with back pictures[1] (out of the infants visual view) we found a significant difference only at 12 months ($t(41) = 2.95$, $p = .005$, $d_z = 0.76$) and not in other months (8 months: $t(46)$ = -2.29, $p = .027$, $d_z = 0.49$; 9 months: $t(44)$ = -1.16, $p = .253$, $d_z = 0.26$; 10 months: $t(39) = 0.56$, $p = .578$, $d_z = 0.15$; 11 months: $t(42) = 1.48$, $p = .147$, $d_z = 0.39$). Means and SE are displayed in Figure 4.9 and reported in detail in Appendix Table A.1. On the individual level, at 11 and 12 months of age, the number of infants who looked to the target in $\geq$ 50% of the trials was equally distributed with infants who looked to the target in < 50% the trials and the binomial test was not significant (11 months: 51%, p = 1.00; 12 months: 43%, p = .441). Between 8 and 10 months of age, more infants looked to the target in < 50% of the trials and the binomial test was significant (8 months: 87%, p < .001; 9 months: 77%, p = .007; 10 months: 65%, p = .081).

Mean and SE of an overall point following skill can be found in Figure 4.9. T-tests are reported in Appendix Table A.2.

Intra-individual stability. For point following, we found highly significant correlations between months in all trials, from 10 months onwards (see correlation coefficients in Figure 4.9) and additionally between 8 and 9 months. For the trials with the front pictures only, we found a similar pattern. See the correlation coefficients in Table

[1]The dependent variable "point following to back targets" also included cases where infants stopped at the front object on the same side and then oriented to the back target.

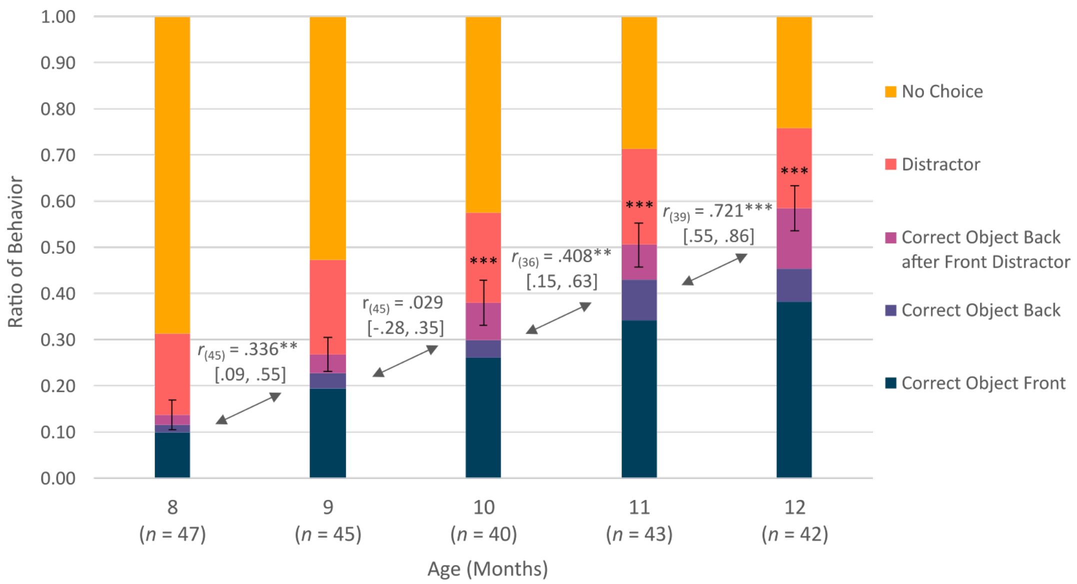

Figure 4.9. The graph displays the point following behavior in the interaction-based task. Correlation coefficients are based on Pearson correlation between all correct point following behaviors. Values in the brackets represent Bootstrap 95% CI. Error bars display standard errors for all correct point following behaviors, including following to front, back, and back after shortly stopping at the front distractor at the same side. Asterisk display significance for t-tests between all correct point following behavior and looking to the distractor. P-values: $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$.

4.3. For point following to the back pictures, we found a highly significant correlation only between 11 and 12 months. See correlation coefficients in Table 4.3.

Longitudinal change. A repeated measurement ANOVA showed a significant main effect for time $F(4, 128) = 20.61$, $p < .001$, $\eta_p^2 = .392$. A t-test between months for the ratio of correct point following behavior in all trials was significant between 8 and 9 months (t = -3.56, $p < .001$) and 10 and 11 months (t = -2.41, $p = .021$), but not between 9 and 10 months (-1.59, $p = .120$) or 11 and 12 months (t = -1.56, $p = .127$).

We found a good fit for a linear latent growth curve model for point following in all trials ($Chi^2_{(10)} = 10.323$, $p = .413$; RMSEA = .026, CI[.000, .162]; CFI = .994;

TLI = .994; SRMR = .120). The mean intercept was significant ($M_{intercept}$=.148, $p <$ .001) and the mean of the slope was positive and significant (M_{slope}=.112, $p < .001$). There were no significant variances for intercept and slope ($V_{intercept}$ = .020; p = .084; V_{slope} = .002, p = .077). We found no relation between intercept and slope (r=.156, p = .716).

Thus, point following to the next visible target seemed to be a stable skill from 10 months on. Point following to back targets seemed to be stable from 11 months on but was only significant on a group level from 12 months on.

4.4.1.6 Referential Expectation – Interaction-based

Task performance. Calculating dependent t-test between ration of correct searching behavior vs. ratio of incorrect searching behavior revealed significant differences from 10 months on. The infants searched more often under the correct cloth than under the incorrect cloth (9 months: $t(44)$ = -1.85, p = .071, d_z = 0.42; 10 months: $t(39)$ = 2.16, p = .037, d_z = 0.51; 11 months: $t(42)$ = 3.309, p = .002, d_z = 0.90; 12 months: $t(41)$ = 3.690, $p < .001$, d_z = 1.01). Means and SE are displayed in Figure 4.10.

On the individual level, the number of infants who searched on the correct side in 50% of the trials or less was significantly higher and the binomial test was significant at 9 months (91%, $p < .001$) and 10 months (73%, p = .006). From 11 months on, they were equally distributed with infants who searched on the correct side in more than 50% of the trials. The binomial test was not significant (11 months: 37%, p = .126; 12 months: 38%, p = .164).

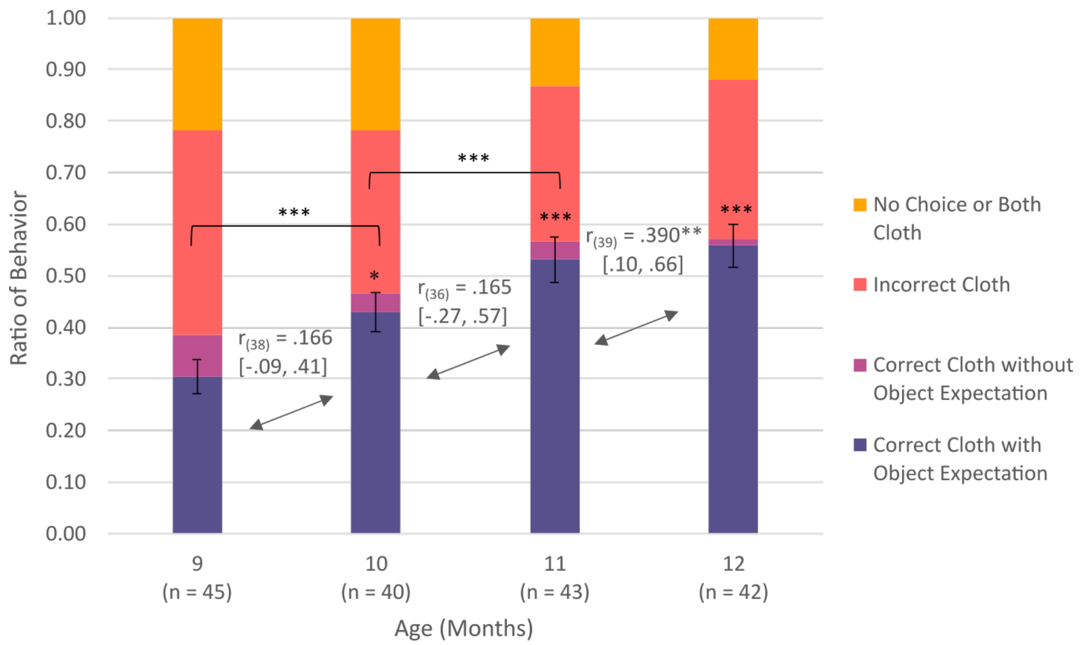

Figure 4.10. The graph displays the sum of behavior in the interaction-based task measuring referential expectation. Correlation coefficients are based on Pearson correlation and values in the brackets display Bootstrap 95% CI. Dependent t-tests were calculated between choosing the correct cloth with expectation of an object against choosing the incorrect cloth. P-values: $+p < .10$, $*p < .05$, $**p < .025$, $***p < .010$.

Intra-individual stability. We found significant relations for correct searching behavior between 11 and 12 months but not between other months. See correlation coefficients in Figure 4.10.

Longitudinal change. A 2 (behavior: correct search, incorrect cloth) by 4 (age groups) repeated measurement MANOVA showed a significant main effect for time $F(3, 93) = 4.01$, $p = .010$, $\eta_p^2 = .115$. We found a significant main effect for behavior, $F(1, 31) = 8.65$, $p = .006$, $\eta_p^2 = .218$, showing more correct than incorrect searching behavior. There was also a significant interaction, $F(3, 93) = 9.08$, $p < .001$, $\eta_p^2 = .227$, showing an increase in correct search behavior over time and a decrease of taking the incorrect cloth. Separate dependent t-test showed a significant increase in correct searching behavior from 9 to 10 months ($t(37)$ = -3.26, p = .002) and 10 to 11 months ($t(35)$ = -2.57, p = .015) but not between 11 and 12 months ($t(38)$ = -0.452, p = .654).

Linear latent growth model for correct searching behavior showed a good model fit ($Chi^2_{(5)}$ = 2.811, p = .729; RMSEA = 0.000; CFI = 1.000; TLI = 1.732). Mean intercept was significant from zero at the initial level at 9 months ($M_{intercept}$=.083, $p < .001$), and the slope was significant as well as positive (M_{slope}=.325, $p < .001$). The intercept and slope were not correlated (r = -.002, p = .625). There were no significant interindividual differences in starting level ($V_{intercept}$ = .007; p = .464) or increase (V_{slope} = .005, p = .071).

Referential expectation measured in an interaction-based task seems to start as a stable skill from 11 months on.

4.4.1.7 Relations with Non-social Disengagement Task

Descriptive of the dependent variable "latency to the peripheral target" was 566 ms (SD = 138, N = 31) and "ratio of refixations to the peripheral target" was 0.86 (SD = 0.22, N = 32).

For eye-tracking tasks, we found no significant relation between both disengagement measures and social-cognitive tasks.

For interaction-based tasks, we found no relations between the different helping tasks and the disengagement task. For point following to the front pictures, we found concurrent correlations to the ratio of refixation in the disengagement task at 8 months ($r(32)$ = .36, p = .046, Bootstrap 95% CI [.22, .53]) and a relation between latency of disengagement and point following at 11 months ($r(30) = .39$, $p = .034$, Bootstrap 95% CI [.15, .61]). For point following to the back targets, we found a significant correlation to mean latency in the disengagement task at 11 months ($r(29) = .434$, $p = .019$, Bootstrap 95% CI [.18, .66]) and 12 months ($r(28) = .52$, $p = .004$, Bootstrap 95% CI [.19, .73]). For referential expectation, there was a strong correlation between mean latency in the disengagement tasks at 8 months and referential expectation at 11 months, again in the wrong direction ($r(29) = .49$, $p = .007$, Bootstrap 95% CI [.15, .77]).

4.4.1.8 Differences between Samples

For eye-tracking tasks, we did not find any differences between German and Turkish-German infants.

For interaction-based tasks, we found one significant difference in the out-of-reach task at 12 months. Turkish-German infants ($N = 10$, M = .30, SD = .42) helped significantly less than German infants ($N = 29$, M = .67, SD = .41), $t(37) = -2.47$, $p = .018$. In point following to the front targets, we found a significant difference at 11 months, $t(41) = -2.76$, $p = .009$. Turkish-German infants followed less (M(14) = .23, SD .18) than German infants (M(29) = .40, SD = 18). We found a similar picture for point following to back pictures at 12 months, $t(40) = 3.15$, $p = .003$). Again Turkish-German infants followed the pointing gesture less (M(12) = .06, SD = .11) than German infants (M(30) = .26, SD = .20). We did not find a difference for referential expectation in the interaction-based task.

4.4.1.9 Summary – Separate Tasks Analysis

In the majority of tasks, we find significant task performance and stability from 11 months on. Therefore, from this time point on, skills start to manifest as stable abilities. However, they do not yet seem to be developed fully, since only half of the infants show the skill at this age.

Although abilities start to arise at 11 months in most of the tasks, some exceptions are present. For action understanding in the eye-tracking task, we find stability for ratio of anticipatory looks from 8 months on, and a significant task performance at 12 months only. However, mean latency is only stable between 11 and 12 months.

For point following in eye-tracking we find an additional task-performance at 9 months, which is not related to other months and not visible at 10 months. Similarly, we found task performance at 9 months for referential expectation in the eye-tracking task. Thus, there seems to be a point following skill at 9 months, which is not related to later point following skills, and may be some preliminary attention direction ability. We will follow up on this when examining the role of this ability on other skills.

For referential expectation in the eye-tracking task, we do not find any stability at all. This means that cueing paradigms do not seem to measure a stable skill at this age. However, in line with the general picture, the dwell-time difference score is significantly different from zero at 11 and 12 months.

We did not measure action understanding in the interaction-based task at 11 months, due to restrictions in the number of tasks during one appointment. However, we were able to find stability between 10 and 12 months - although without significant task performance at 10 months. Thus, in the first year of life helping seems to emerge as a stable skill.

Point following to pictures within the infant's visual field starts to be stable at 10 months already, and task performance is visible at 9 months, similarly as in the eye-tracking task. Point following to pictures out of the infant's visual field follows the general pattern of task performance and stability from 11 months on. This is the case for referential expectation in interaction-based task as well, where we additionally find a significant task performance without stability at 10 months. Therefore, a referential understanding of others pointing gestures by identifying the referent even when it is out of someone's visual field or hidden seems to appear at the end of the first year, while point following to visible objects starts one to two months earlier.

4.4.2 Relations between Skills within each Method

4.4.2.1 Action Understanding and Point Following – Eye-tracking

All correlation coefficients are listed in Table 4.1.

Looking at concurrent relations, the Pearson correlation indicated a high positive relation between skills at 12 months. This relation remained when controlling for the ratio of anticipatory looks in the failed reaching task in the preceding month ($r_{p(17)}$ = .52, p = .022, Bootstrap 95% CI [.20, .80]) as well as the ratio of correct point following in the preceding months ($r_{p(23)}$ = .47, p = .018, Bootstrap 95% CI [.20, .74]). There was also a one-tailed concurrent correlation at 11 months that increased in strength when controlling for the previous month in action understanding,

$r_{p(19)} = .487$, $p = .025$, Bootstrap 95% CI [-.12, .81]. However, bootstrapping results did not verify this relation.

We mainly found predictive correlations from the ratio of anticipatory looks in the failed reaching task to the ratio of first looks to the target in the point following task. One was from action understanding at 11 months to point following at 12 months. However, this relation did not extend when controlling for the competence in the previous months (action understanding at 10 months: $r_{p(16)} = .32$, $p = .191$, Bootstrap 95% CI [-.12, .72]; point following at 11 months: $r_{p(17)} = .33$, $p = .172$, Bootstrap 95% CI [-.06, .74]). Moreover, there were only weak and one-tailed significant relations from early action understanding (8 and 10 months) to later point following (12 months). In the other direction, there was only one one-tailed significant correlation from point following at 11 months to action understanding at 12 months.

Using mean latency in the failed reaching tasks, instead of the ratio of anticipatory looks, we could validate the high concurrent correlation to ratio of correct point following at 12 months ($r(31) = -.61$, $p < .001$, Bootstrap 95% CI [-.76, -.42]). The relation remained significant when controlling for competence in the previous months, however, the remaining n was partly very small (action understanding at 11 months: $r_{p(14)} = -.51$, $p = .045$, Bootstrap 95% CI [-.76, -.30]; point following at 11 months: $r_{p(22)} = -.53$, $p = .007$, Bootstrap 95% CI [-.76, -.29]). Furthermore, there was one predictive correlation from action understanding at 10 months to point following at 12 months ($r(22) = -.46$, $p = .030$, Bootstrap 95% CI [-.66, -.22]). There were no other significant correlations (see correlation coefficients in Appendix Table A.3.

We controlled for a general disengagement skill at 8 months, resulting in indications for relations to social-cognitive tasks (see correlations in the section Relations with Non-social Disengagement Task). The concurrent relations at 12 months remained when controlling for ratio of refixation (anticipatory looks: $r(21) = .53$, $p = .009$, Bootstrap 95% CI [.22, .81]; latency in the failed reaching tasks: $r(20) = -.55$, $p = .008$, Bootstrap 95% CI [-.74, -.37]). The predictive relation from the ratio of anticipatory looks in the failed reaching task at 11 months to the ratio of correct point following at 12 months did not hold when controlling for mean latency in the disengagement task ($r_{p(13)} = .36$, $p = .182$, Bootstrap 95% CI [-.25, .81]), however, the remaining $n = 14$ was comparatively small. All other predictive relations from action understanding to point following at 12 months remained when controlling for the ratio of refixations (anticipatory looks in the failed reaching task: 8 months: $r_{p(19)} = .47$, $p = .032$, Bootstrap 95% CI [.12, .70]; 10 months: $r_{p(17)} = .59$, $p = .008$, Bootstrap 95% CI [.29, .78].

In summary, we found concurrent and bidirectional relations when task performance in both skills was stable, at 11 and 12 months. This means that infants who can anticipate others' reaching actions also follow others' pointing gestures correctly. We also found some indications for predictive correlations from action understanding (8-10 months) to later point following (12 months). We found this relation before we found task performance in action understanding, but could already see a stability in anticipatory looks to the ball at early ages.

Table 4.1

Correlation coefficients for action understanding and communicative understanding in eye-tracking tasks.

	action understanding - eye-tracking				
	8 months	9 months	10 months	11 months	12 months
point following eye-tracking					
9 months	-.12 [-.39, .16]	-.19 [-.43, .06]	.04 [-.24, .35]	.23 [-.18, .63]	.07 [-.28, .39]
10 months	.02 [-.37, .39]	-.26 [-.51, .07]	.02 [-.30, .44]	.09 [-.34, .45]	.14 [-.31, .53]
11 months	.06 [-.23, .43]	.17 [-.18, .54]	.17 [-.20, .57]	.35+ [-.22, .69]	.35+ [-.01, .63]
12 months	.37+ [.05, .63]	.26 [-.06, .55]	.38+ [.08, .67]	.49** [.17, .80]	.55*** [.27, .76]
referential expectation eye-tracking					
8 months	.11 [-.19, .43]	-.05 [-.39, .23]	-.07 [-.45, .32]	-.16 [-.55, .23]	-.24 [-.56, .18]
9 months	-.07 [-.35, .27]	-.14 [-.44, .14]	.00 [-.40, .33]	-.12 [-.51, .43]	-.09 [-.44, .23]
10 months	.03 [-.29, .40]	.02 [-.35, .34]	.05 [-.28, .40]	.07 [-.29, .40]	.20 [-.10, .52]
11 months	.06 [-.34, .52]	-.12 [-.42, .37]	-.07 [-.41, .41]	.05 [-.38, .53]	-.02 [-.36, .40]
12 months	.13 [-.41, .50]	.29 [-.18, .59]	.53*** [.06, .75]	.54** [.28, .75]	.04 [-.27, .39]

Note. $+p < .1$, *$p < .05$. **$p < .025$, ***$p < .01$. Bootstrapped 95% CIs reported in brackets. Bonferroni adjusted alpha level: $p < .01$; $p = 0.1/(5*2)$; Median N = 29 [20; 36].

4.4.2.2 Action Understanding and Referential Expectation – Eye-tracking

All correlation coefficients can be found in Table 4.1.

There were no concurrent relations between the ratio of anticipatory looks in the failed reaching task and the duration-effect indication referential expectation.

We found predictive correlations from early action understanding (10 & 11 months) to later referential expectation (12 months). The correlation from ratio of anticipatory looks in the failed reaching task at 10 months remained when controlling for previous action anticipation skill at 9 months ($r_{p(23)}$ = .47, p = .017, Bootstrap 95% CI [.04, .87]). The predictive relation from action anticipation at 11 months was no longer significant when controlling for previous action anticipation at 10 months ($r_{p(17)}$ = .38, p = .112, Bootstrap 95% CI [-.08, .74]).

Regarding mean latency in the failed reaching task, we did not find any correlations to referential expectation. For correlation coefficients, see Appendix Table A.3.

When controlling for mean latency in the disengagement task, the predictive relation from ratio of anticipatory looks at 11 months to referential expectation at 12 months remained significant ($r_{p(15)}$ = .62, p = .009, Bootstrap 95% CI [.30, .99]).

Thus, the ability to anticipate others reaching actions at 10 and 11 months seems to be partly relevant for referential expectation in the eye-tracking task at 12 months.

4.4.2.3 Point Following and Referential Expectation – Eye-tracking

All correlation coefficients are listed in Table 4.2.

We found one concurrent relation at 9 months correlations between point following and referential expectation in the eye-tracking tasks. However, at this time point we did not find stable skills in either task.

All other predictive correlations were present before we found stable skills and relations were negative, which we did not expect.

Table 4.2

Correlation coefficients for communicative understanding in eye-tracking tasks.

	referential expectation – eye-tracking				
point following eye-tracking	8 months	9 months	10 months	11 months	12 months
9 months	-.25 [-.49, .00]	.36* [.02, .61]	.22 [-.08, .46]	-.09 [-.59, .34]	-.26 [-.54, .13]
10 months	.08 [-.31, .40]	.18 [-.28, .52]	.15 [-.21, .41]	-.34+ [-.60, .00]	-.41* [-.61, -.19]
11 months	.05 [-.30, .37]	-.29 [-.53, .02]	.01 [-.34, .40]	-.08 [-.43, .31]	-.08 [-.42, .34]
12 months	.00 [-.37, .33]	-.01 [-.39, .34]	-.01 [-.36, .36]	-.05 [-.45, .45]	.00 [-.24, .24]

Note. $+p < .1$, $*p < .05$. $**p < .025$, $***p < .01$ Bootstrapped 95% CIs reported in brackets. Bonferroni adjusted alpha level: $p < .017$; $p = 0.1/(2*3)$. Median N = 34 [29; 39].

In summary, following others' pointing gestures to visible objects in a video setting with communicative aspects (e.g. eye-contact) seem to measure different competences than violation-of-expectation in cueing paradigms although both skills started to become stable from 11 months on. In both tasks, we found an indication of task performance at 9 months an they are correlated to each other. These early and related effects may be an indication for spatial cueing effects before infants understand the referential nature of a pointing gesture. This issue will be discussed further in section 4.5.

4.4.2.4 Action Understanding and Point Following - Interaction-based

The correlation coefficients are listed in Table 4.3.

For point following to front pictures, we found only very weak concurrent and predictive relations.

For point following to the back targets we found a high concurrent correlation to the out of reach tasks at 12 months. This relation remained significant when controlling for point following skill in the previous months (11 months: $r_{p(33)} = .53$, $p = .001$, Bootstrap 95% CI [.29, .74]) as well as when controlling for the out-of-reach task in the previous month (10 months: $r_{p(31)} = .37$, $p = .037$, Bootstrap 95% CI [.02, .69]). There was only one weak predictive relation from point following at 12 months to out-of-reach tasks at 14 months.

When controlling for mean latency in the disengagement task, the concurrent relation between the out of reach tasks and point following to the back pictures at 12 months remained significant ($r_{p(24)} = .41$, $p = .038$, Bootstrap 95% CI [.01, .74]).

This yields the results that infants who search for a referent that is specifically pointed at outside their visual field also understand others' goals of non-communicative actions and helped their interaction partner at 12 months. Leading the infants' attention to the next visible object (point following to front targets) was only partly related to later helping behavior.

4.4.2.5 Action Understanding and Referential Expectation – Interaction-based

All correlation coefficients are listed in Table 4.3.

No relation was found between referential expectation and action understanding in interaction-based tasks when both skills were present as stable abilities.

4.4.2.6 Point Following and Referential Expectation - Interaction-based

Find all correlation coefficients displayed in Table 4.3.

For point following to the front pictures, we found no concurrent correlations. Some weak predictive correlations were present only before both tasks revealed stable skills.

When focusing on point following to the back pictures we found a high concurrent correlation to referential expectation at 11 months. This relation remained significant when controlling for latency in the disengagement task ($r_{p(26)} = .48$, $p = .010$, Bootstrap 95% CI [.10, .75]). We did not find predictive relations.

Therefore, with point following to back pictures we seem to measure a related ability as with point following to hidden entities, when they both begin to be stable competencies.

Table 4.3

Correlation coefficients for action understanding and communicative understanding in interaction-based tasks

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
action understanding															
1. 10 months - oor															
2. 12 months - oor	.32+ [.02, .56]														
3. 14 months - oor	.18 [-.12, .43]	-.26 [-.59, .14]													
4. 12 months - hid	.45** [.19, .69]	.59** [.35, .14]	-.01 [-.38, .06]												
5. 14 months - hid	.00 [-.32, .33]	.11 [-.25, .47]	.38* [.06, .67]	.01 [-.33, .37]											
point following - front pictures															
6. 8 months	.29+ [-.04, .59]	.28+ [-.03, .54]	.14 [-.14, .35]	.17 [-.18, .50]	.35* [.07, .57]										
7. 9 months	.11 [-.25, .46]	.21 [-.10, .55]	.04 [-.32, .37]	.25 [-.07, .55]	.13 [-.23, .46]	.36* [.11, .60]									
8. 10 months	.28+ [-.05, .56]	.19 [-.16, .56]	.14 [-.27, .52]	.37* [.06, .62]	.14 [-.23, .48]	.35* [.07, .58]	.12 [-.22, .43]								
9. 11 months	.16 [-.12, .38]	.41* [.08, .74]	.38* [-.01, .71]	.35* [.05, .58]	.30+ [-.01, .59]	.36* [.16, .53]	.16 [-.12, .43]	.44** [.11, .72]							
10. 12 months	.24 [-.03, .44]	.24 [-.10, .57]	.35* [.00, .67]	.18 [-.15, .48]	.29+ [-.04, .58]	.28+ [.07, .45]	.02 [-.30, .31]	.14 [-.23, .49]	.61** [.35, .81]						
point following - back pictures															
11. 8 months	.25 [-.22, .60]	.14 [-.24, .43]	.03 [-.33, .24]	.02 [-.30, .38]	.27 [-.07, .50]	.32* [-.01, .63]	.01 [-.25, .30]	.36* [.16, .54]	.17 [-.19, .38]	.22 [.11, .35]					
12. 9 months	-.05 [-.33, .27]	.15 [-.14, .44]	-.13 [-.49, .21]	.11 [-.24, .43]	.10 [-.24, .43]	.20 [-.23, .47]	.11 [-.10, .50]	.05 [-.18, .38]	.17 [-.27, .36]	.02 [-.07, .38]	.09 [-.20, .42]				
13. 10 months	.19 [-.18, .59]	.23 [-.10, .49]	-.04 [-.42, .35]	.15 [-.20, .51]	-.15 [-.43, .19]	.12 [-.18, .52]	-.01 [-.31, .37]	.31+ [.05, .57]	.27 [.00, .48]	.21 [-.04, .40]	.01 [-.21, .41]	-.15 [-.37, .11]			
14. 11 months	.25 [-.12, .59]	-.02 [-.39, .32]	.13 [-.22, .42]	.00 [-.37, .36]	.14 [-.22, 46]	.38* [.00, .66]	.25 [-.05, .55]	.30+ [.00, .57]	.26+ [.01, .49]	.25 [-.03, .46]	.40** [.06, .66]	.30 [-.02, .59]	.08 [-.23, .47]		
15. 12 months	.20 [-.15, .50]	.44** [.13, .71]	-.09 [-.42, .27]	.29+ [-.06, .58]	.08 [-.24, .39]	.53** [.26, .75]	.26 [-.05, .55]	.46** [.19, .69]	.58** [.38, .73]	.37* [.05, .62]	.16 [-.09, .45]	.28 [-.02, .55]	.19 [-.11, .50]	.46** [.17, 70]	
referential expectation															
16. 9 months	-.02 [-.31, .33]	.30+ [-.03, .59]	-.11 [-.46, .27]	.12 [-.19, .49]	.23 [-.10, .54]	.13 [-.15, .41]	.20 [-.12, .52]	.31+ [.04, .58]	.21 [-.06, .47]	.29+ [.01, .51]	.21 [-.06, .47]	-.08 [-.34, .19]	.16 [-.30, .54]	.10 [-.25, .42]	.35* [.10, .56]
17. 10 months	-.02 [-.35, .37]	-.02 [-.38, .37]	.04 [-.31, .40]	.05 [-.34, .42]	.08 [-.27, .46]	-.09 [-.39, .19]	.33* [.04, .59]	.28+ [-.05, .59]	.10 [-.21, .43]	-.15 [-.49, .23]	.19 [-.13, .46]	.14 [-.15, .41]	.06 [-.27, .37]	.06 [-.32, .41]	-.09 [-.40, .28]
18. 11 months	.16 [-.12, .47]	.06 [-.29, .39]	.16 [-.15, .41]	-.07 [-.48, .32]	.08 [-.28, .40]	.30+ [-.05, .58]	-.02 [-.34, .31]	.19 [-.14, .50]	.10 [-.20, .40]	.05 [-.30, .36]	.31* [-.06, .55]	.10 [-.28, .48]	.29+ [.04, .54]	.49** [.18, .73]	.08 [-.26, .40]
19. 12 months	.18 [-.20, .51]	.18 [-.16, .46]	.14 [-.15, .41]	.05 [-.29, .37]	.15 [-.19, .44]	.37* [-.02, .69]	.15 [-.14, .42]	.11 [-.21, .41]	.15 [-.13, .44]	.08 [-16, .33]	.39* [.01, .62]	.01 [-.32, .33]	.32+ [.11, .54]	.24 [-.13, .56]	-.08 [-.37, .20]

Note. $+p < .1$, $*p < .05$. $**p < .025$, $***p < .01$ Bootstrapped 95% CIs reported in brackets. Bonferroni adjusted alpha level: for point following to front pictures $p < .017$; $p = 0.1/(2*3)$, for all others $p < .025$; $0.1/(2*2)$. Median N = 38 [32; 47]

4.4.2.7 Summary – Relation between Skills

For an overview see Figure 4.11. In eye-tracking tasks, point following and action understanding are linked concurrently at 11 and 12 months but we also find indications for directional relations from action understanding to point following. Similarly, for referential expectation, correlations are predominantly predictive from action understanding (10 and 11 months) to referential expectation at 12 months. Therefore, developing action understanding seems to be similarly relevant for both levels of communicative understanding. The two levels of communicative understanding are not related to each other at this age. However, we find concurrent relations at 9 months, when there is task performance in both tasks without stability within each task. This may support the assumption that there is an early attention orienting skill, that is different from the later stable referential understanding.

In interaction-based tasks, we find some weak predictive relations from point following to front target to later helping behavior. Concurrent relations are between point following to back targets and action understanding at 12 months. In contrast, we could not find any links between action understanding and referential expectation. Different from eye-tracking tasks, we found relations between both communicative skills in interaction-based tasks. Concurrent associations are high at 11 months, between the infant's point following to back targets and referential expectation when pointing to hidden targets. There are only few relations from point following to visible objects to point following to objects out of the infants' visual field, and no relations to referential expectation when pointing to hidden objects. These findings suggest that early point following skills seem to play only a small role for developing a referential expectation, which is in line with findings from Brune and Woodward (2007).

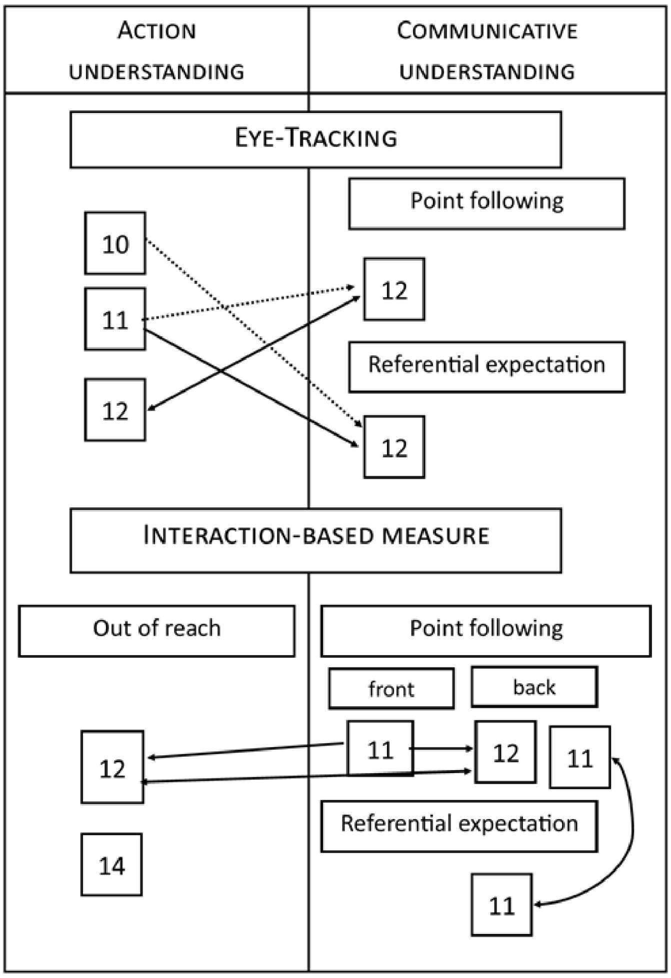

Figure 4.11. Schematic overview of significant relations between skills within methods. Full arrows represent significant relations after Bonferroni alpha correction. Dotted arrows represent close to significant relations after onferroni correction. The numbers in squares represent the age of the infants in months.

4.4.3 Relations between Methods within each Skill

4.4.3.1 Action Understanding

Find the correlation coefficients for anticipatory looks in the failed reaching task and helping behavior in interaction-based task in Table 4.4.

Table 4.4

Correlation coefficients for action understanding between methods

	Action Understanding – Eye-tracking				
	8 months	9 months	10 months	11 months	12 months
Action Understanding Interaction-based					
10 months	.02	-.18	-.16	.48**	-.19
	[-.27, .33]	[-.50, .20]	[-.41, .14]	[.16, .76]	[-.51, .15]
12 months - out of reach	-.10	-.14	-.09	.23	.05
	[-.44, .25]	[-.51, .20]	[-.43, .36]	[-.22, .64]	[-.34, .44]
14 months - out of reach	-.06	-.10	-.17	.09	-.06
	[-.45, .31]	[-.47, .19]	[-.55, .20]	[-,32, .46]	[-.45, .37]
12 months - hidden	-.15	-.17	-.19	.37	-.12
	[-.48, .21]	[-.46, .18]	[-.49, .18]	[-.05, .72]	[-.48, .29]
14 months - hidden	-.08	-.04	-.33	.15	-.06
	[-.40, .32]	[-.43, .29]	[-.64, .02]	[-.27, .52]	[-.44, .34]

Note. $+p < .1$, $*p < .05$. $**p < .025$, $***p < .01$. Bootstrapped 95% CIs reported in brackets. Bonferroni adjusted alpha level: $p < .017$; $p = 0.1/(2*3)$. Median N = 30 [21; 35].

We found no concurrent correlations. There was one predictive correlation from helping at 10 months to the ratio of anticipatory looks in the eye-tracking task one month later. This relation remained significant when controlling for the previous months in the eye-tracking task ($r_{p(20)} = .47$, $p = .029$, Bootstrap 95% CI [.09, .75]).

However, we could not validate this correlation by using the latency measure in the failed reaching task. Find correlation coefficients for latency in the eye-tracking tasks displayed in Appendix Table A.4.

However, this result has to be interpreted with caution since at 10 months we did not find tasks performance in the helping tasks, although there was already stability to the next appointment at 12 months, where already half of the infants helped.

In summary, action anticipation speed in the eye-tracking task does not seem to be relevant for action understanding in interaction-based tasks, however, there is a tendency the other way around. However, the relation is from 10 months on where only very few infants show helping behavior.

4.4.3.2 Point Following

The correlation coefficients are listed in Table 4.5.

For point following to front targets in interaction-based task, there was a one-tailed significant concurrent correlation at 11 months. We found very weak predictive correlations from point following in interaction at 10 months to point following in eye-tracking at 11 months.

Point following to back targets was concurrently correlated to point following in eye-tracking at 11 months. When controlling for mean latency in the disengagement task for point following to back pictures at 11 months, the relation remained significant $r_{p(22)} = .47$, $p = .019$, Bootstrap 95% CI [.01, .80].

Therefore, point following to front targets in the interaction-based task does not seem to be relevant for point following in the eye-tracking task or the other way around. Concerning point following to back targets the interaction-based task seems to measure similar abilities as point following in eye-tracking videos. When both skills emerge at 11 months, we found high concurrent relations.

Table 4.5

Correlation coefficients for point following between methods

Point Following Interaction-based		Point Following - Eye-tracking 9 months	10 months	11 months	12 months
Front Pictures					
	8 months	-.04	-.10	.35*	.15
		[-.28, .24]	[-.40, .24]	[-.07, .64]	[-.20, .47]
	9 months	.15	-.02	.12	.10
		[-.12, .45]	[-.38, .33]	[-.25, .48]	[-.29, .45]
	10 months	.14	-.07	.35*	.31+
		[-.20, .47]	[-.39, .26]	[-.05, .66]	[-.04, .59]
	11 months	.07	.02	.08	.07
		[-.22, .35]	[-.32, .38]	[-.29, .41]	[-.31, .48]
	12 months	.01	-.12	.16	-.10
		[-.35, .35]	[-.50, .25]	[-.18, .45]	[-.45, .27]
Back Pictures					
	8 months	.24	-.07	.27	-.02
		[-.16, .54]	[-.31, .21]	[-.15, .57]	[-.23, .18]
	9 months	.15	-.08	.09	.31+
		[-.21, .48]	[-.41, .28]	[-.30, .45]	[-.07, .64]
	10 months	-.07	-.03	.29	.17
		[-.34, .28]	[-.37, .27]	[-.01, .59]	[-.12,.48]
	11 months	-.04	.02	.39**	.14
		[-.31, .25]	[-.33, .42]	[.02, .66]	[-.22, .48]
	12 months	-.09	-.08	.21	.16
		[-.36, .23]	[-.47, .31]	[-.20, .56]	[-.23, .54]

Note. $+p < .1$, *$p < .05$. **$p < .025$, ***$p < .01$. Bootstrapped 95% CIs reported in brackets. Bonferroni adjusted alpha level: $p < .017$; $p = 0.1/(2*3)$. Median N = 35 [30; 42].

4.4.3.3 Referential Expectation

The correlation coefficients are listed in Table 4.6.

For referential expectation measured with different methods, we found one concurrent relation. However, this was at 10 months where both skills were not yet stable. We found one predictive correlation from the 10-month's skill in the interaction-based task to the 11-month's skill in the eye-tracking task.

In summary, we find only very weak relations between methods measuring referential expectation. This could be due to different task demands and differences in

Table 4.6

Correlation coefficients for referential expectation between methods

	Referential Expectation – Eye-tracking				
	8 months	9 months	10 months	11 months	12 months
Referential Expectation Interaction-based					
9 months	.12	.10	-.07	.13	-.11
	[-.16, .37]	[-.22, .40]	[-.37, .26]	[-.37, .53]	[-.32, .20]
10 months	-.39**	-.01	.33+	.42**	.01
	[-.68, -.12]	[-.32, .33]	[.03, .57]	[-.03, .68]	[-.31, .42]
11 months	.12	-.25	-.01	.19	.09
	[-.14, .38]	[-.54, .10]	[-.41, .33]	[-.20, .59]	[-.16, .44]
12 months	.34*	.05	.24	.16	-.12
	[-.032, .617]	[-.23, .36]	[-.17, .55]	[-.21, .52]	[-.46, .32]

Note. $+p < .1$, $*p < .05$. $**p < .025$, $***p < .01$. Bootstrapped 95% CIs reported in brackets. Bonferroni adjusted alpha level: $p < .017$; 0.1/(2*3). Median N = 34 [29; 39].

communicative embeddedness, which seems to be relevant for communicative understanding in the time window of emergence.

4.4.3.4 Summary – Relations between Methods

We found only very few and not very stable links between methods but there is a similar tendency for all abilities. Directional links are exclusively from skills in interaction-based tasks and skills in eye-tracking tasks one months later. This is especially present for point following to visible targets, becoming stable from 10 months on already. All other directional relations are to be interpreted with caution as we find them before they are stable skills in interaction-based tasks. Concurrent relations are only visible for point following, where task demand is very similar. Similarly, infants first start to show a more sophisticated communicative understanding within interaction (following to objects outside their visual field) and then start to follow to visible objects in less communicative video scenes. We did not find concurrent relations for the other skills.

However, we did not measure action understanding in the interaction-based task for all measurement time points. We may have missed concurrent relations at 11 months but we also did not find significant relations at 12 months or for referential expectation. Thus, we only find partial validity between different methods at this age. Different task demands and differences in the communicativeness of the situation seem to influence whether we capture early social-cognitive abilities or not.

4.5 Discussion

With this study, we addressed two main questions: a conceptual and a methodological one. Regarding the conceptual question, we tested empirically when two major early social-cognitive skills, namely action understanding and communicative understanding, emerge to be stable abilities and whether they are related in their development. Regarding the methodological question, we examined whether we tap similar abilities when infants are tested in different settings, either involved in interaction or observing videos stimuli. Therefore, we captured action understanding and communicative understanding longitudinally with interaction-based as well as eye-tracking methods in the second half of the infants' first year. Finding developmental relations between skills, we see as evidence that they emerge based on one related understanding. Relations between methods would account for internal validity within one skill.

Regarding the emergence of competencies, as a general picture, we find stability within skills between 11 and 12 months of age. Thus, solid action understanding and communicative understanding seem to start from 11 months on. This time point of emergence speaks for the theoretical assumption that infants develop an understanding of others' perspectives and underlying goals after 9 months of age once they begin to be able to jointly attend to outside entities with others (9-months-revolution; Tomasello et al., 2005). Infants seem to need some time to construct a stable understanding of others' actions and communicative gestures as we find stability only at the end of the first year. In addition, the different skills are not fully developed at this time point. Only half of the infants show competences. Our results are contrary to the approach that social-cognitive abilities are present from birth and emerge in behavior early on (e.g. Baillargeon et al., 2016; Csibra, 2003). Before 11 months, we find some abilities only in particular months (communicative understanding in eye-tracking with 9 months but not with 10 months) without stability that lasts into the following months. At this age, infants may be able to direct their attention to the next visible object but do not understand the referential nature of a pointing gesture. These early effects seem to be in line with single performances found in previous cross-sectional designs. However, we could not confirm them as stable competences with a longitudinal analysis. Therefore, early social-cognitive abilities seem to consolidate themselves at the end of the first year, which speaks for the view that they are a developmental outcome in contrast to a starting point of human social cognition.

Regarding the developmental relations, we find synchronous relations between the majority of skills when we find competencies to be stable from 11 months on. At an earlier age, we only find very weak and singular concurrent relations at 9 and 10 months. In addition, predictive relations from earlier to later months are only very weak and hard to interpret because the different skills are not stable by themselves. Thus, when different social-cognitive skills start to be present as stable competencies, they seem to be manifestations of a comprehensive understanding of other human beings. Our findings are in line with the theoretical assumption that one underlying understanding develops during the first year of life and presents itself through different skills (Tomasello et al., 2005). In further studies, different aspects of action understanding should be examined to capture a broader picture of relations between action understanding and communicative understanding. However, finding relations in the first year of life leaves little room for the view that action understanding and communicative understanding emerge as independent, separate systems (Csibra, 2003).

We can partly exclude that general disengagement abilities are the reason for relations between different skills. We did not find relations between general disengagement in non-social tasks and social tasks, in the way that faster disengagement and higher flexibility in shifting one's attention is beneficial for social-cognitive abilities. On the contrary, we found that infants who disengage later in non-social tasks show higher abilities in some social-cognitive tasks at 11 months. This means that higher focused attention could lead to better skills in social cognition (Ruff, Capozzoli, & Saltarelli, 1996). However, this disagrees with several other findings that show that the ability to flexible shift one's attention is beneficial for learning and memory development (for an overview see Oakes & Amso, 2018). In order to examine the role of general attention abilities for social-cognitive development, general attention should be measured longitudinally as well. However, this was not the focus of our study and we had to minimize the number of tasks at each time point, so we only measured it at a sensitive period at the beginning of the study. For the current study, it is important to highlight that relations between social-cognitive abilities remained when controlling for general disengagement abilities.

From a methodological point of view, we do not find an advantage of eye-tracking tasks over interaction-based tasks as previously assumed (Krogh-Jespersen & Woodward, 2016). If at all, we find it to be the other way around. In all task relations, predictive relations stem exclusively from interaction-based tasks at 10 months to eye-tracking tasks one month later. These relations are to be interpreted with caution because abilities did not consolidate themselves as stable abilities at this time point. Nevertheless, they support the assumption that infants learn through interaction (Moore & Barresi, 2017). Thus, infants first show preliminary abilities within social interaction before they show an ability in less communicative observational settings. This difference should be taken into account when deciding for a specific method to measure abilities in a sensitive period.

When abilities start to become stable at 11 months on, we find single synchronous relations between eye-tracking and interaction-based tasks. Thus, there is partial validity between tasks when infants start to show stable abilities. However, we can only confirm this for point following, where the task demand is very similar – looking to the object that is beeing pointed at. We did not find it for action understanding or referential expectation where infant have to search for an object or hand an object to an interaction partner. Interaction-based tasks require motor abilities and social motivational components, which seem to be very different to eye-gaze behavior. One difference to most other looking-time tasks is that in our referential expectation task infants had a limited time (5 s) to show a looking-time difference. In general, looking-time tasks are infant-controlled and end when the infant looks away for some time. Therefore, it may take more processing time for such young infants to show a similar understanding than within interaction. However, as we do not find relations for other eye-gaze measures as well (e.g. anticipatory looking), it has to be taken into account that the methods used, differing in their dependent variable and the test environment, have a crucial influence on abilities being present at this time point.

The current study gives important empirical evidence for different theoretical assumptions regarding the development of early social-cognitive abilities. Action understanding and communicative understanding seem to consolidate in a related way at the end of the first year, revealing itself first within interaction. All these results leave more room for social interactional accounts (e.g. Moore & Barresi, 2017) in the way that infants develop an understanding of others through interaction. Examining the precursors of these abilities might therefore be a good starting point for further research. Our results disfavor nativist views, which assume that social-cognitive abilities are the starting point and part of separate systems.

Chapter 5

General Discussion

The aim of this thesis was to examine the development of early social cognitive abilities. Different theories exist about the emergence of early social cognition. Empirical evidence is mixed and systematic comparisons or longitudinal approaches are rare. The current thesis focused on three major questions to test different theoretical approaches empirically. The first one: do infants need communicative situations to show early social cognitive abilities? The second question asks, when do different social cognitive abilities start to be stable? With the third question i wanted to know, whether these different abilities emerge separately or whether they develop as an integrated understanding?

5.1 Summary of Findings

In Study 1, I tested which situations elicit social cognitive abilities, in this case visual-perspective-taking (VPT). This first research question was motivated by the assumption that if infants develop their understanding of others through interaction, as social constructivist views would assume (e.g. Moore & Barresi, 2017; Tomasello et al., 2005), social cognitive abilities should first be visible in very interactive and communicative situations. In contrast, if infants have some innate implicit belief tracking or psychological reasoning abilities, as some nativist views would suggest (e.g. Apperly, 2013; Baillargeon et al., 2016; Luo & Baillargeon, 2010), abilities should be present in a variety of situations independent of the degree of communication. Therefore, I conducted a cueing paradigm, where a person was either communicatively pointing behind one of two boxes or simply directed to one side, similar to the adult VPT tasks (e.g. Samson et al., 2010). In the communicative condition, 14-month-olds detected the hidden object earlier in the congruent than in the incongruent trials. In addition to this cueing effect, 14-month-olds looked longer to the empty box in incongruent than in congruent trials, indicating violation of expectation. Thus, they seem to know that the person sees something and refers to something, and therefore formed a referential object expectation. However, in the non-communicative setting 14-month-olds did not display VPT abilities. Incorporating facial expression, to make the cue more salient and clearer that the person refers to something did not help 14-month-olds. At 36 months of age, infants were able to take others' perspectives in non-communicative settings as well. Thus, the findings of Study 1 support social constructivist views claiming that abilities develop and are revealed first in communicative interaction (e.g. Moore & Barresi, 2017; Siposova & Carpenter, 2019; Tomasello, 2019). A spontaneous VPT ability, present in less communicative situations as well, seems to develop over time and was not found to be present from early on as two-system theory suggests (Apperly, 2013).

In Studies 2 and 3, I addressed the second research question and examined the exact time point of emergence of social cognitive abilities. In Study 2, I followed up on the positive effect of VPT in the communicative setting with 14-month-olds by investigating younger infants. I found some point following ability in 8-month-olds, indicated by a faster latency to the object in congruent trials. In addition, there were some preliminary looking-time effects when the person reached successfully behind a barrier. However, in the pointing and failed reaching condition I did not find a reliable VPT taking ability indicated by a looking-time difference in 8-month-olds. In addition, in Study 3, I tested action understanding and communicative understanding in a longitudinal design and I found them to be stable from 11 months on. Before this age, infants showed some point following abilities, however, there was no evidence that they understand the referential nature of a pointing gesture. Thus, the findings regarding the age of emergence fit with social constructivist views, which propose that infants construct an understanding of others' perspectives and goals after 9 months of age, when they are able to engage in joint attention (e.g. Liszkowski, 2018; Tomasello et al., 2005). The current findings disfavor the nativist views suggesting innate social cognitive abilities, which should be present at least to some extend from early on (Baillargeon et al., 2016; Csibra, 2003; Csibra & Volein, 2008; Luo & Johnson, 2009).

In Study 3, I also examined research question 3, if and how different social cognitive abilities are related in their development. While some nativist views argue that communicative and action understanding are based on two separate systems (Csibra, 2003), social constructivist views assume that an integrated understanding of others causes different social cognitive skills (Tomasello et al., 2005). In a longitudinal design, I measured action understanding and communicative understanding monthly between 8 and 12 months of age. From 11 months on, when the majority of skills started to be stable, I found correlations between different social cognitive skills, such as action and communicative understanding. To exclude the possibility that general attention improvement mediates these relations, I examined disengagement abilities in nonsocial tasks in the beginning of the study at 8 months of age. Even after controlling for general disengagement abilities, relations between social cognitive skills remained. Thus, finding relations in the first year of life is in line with the assumption that infants develop an integrated understanding of others (Tomasello et al., 2005). They fit less with the theoretical assumption that early social cognitive abilities are based on two different action-interpretation systems (Csibra, 2003).

5.2 Integration into Theories

Overall, the results of the current thesis are more in line with the social constructivist views and less with the nativist views. In the following section, I will present arguments for this postulation in more detail, focusing again on my three research questions, which are based on divergent assumptions between different theories. I will integrate the findings of the current studies into the theoretical approaches presented in the introduction.

5.2.1 Age of Emergence

First, social cognitive abilities were found to be stable from 11 months of age onward, and not earlier. Between 8 and 10 months, only some preliminary, not stable competences especially in directing attention were present. This fits with social constructivist views, assuming that infants construct an understanding of others perspectives only after 9 months of age (Tomasello et al., 2005). By 9 months of age, infants start to engage in joint attention. Through joint engagement, they experience others' attention together with their own and form and construct an understanding of others' perspectives and goals (Siposova & Carpenter, 2019). Thus, these early social cognitive abilities seem to be already a developmental outcome and not a starting point of human social cognition, as nativist accounts claim (e.g. Baillargeon et al., 2016).

That abilities start to be stable from 11 months on, does not mean that infants do not understand anything about others before this age. By 8 months of age, infants showed some attention direction abilities when a person pointed communicatively or reached successfully behind a barrier but they did not show reliable VPT abilities. By 9 months of age, some communicative abilities were found in eye-tracking tasks but without relations to later months. Thus, this ability was not stable. In addition, in interaction-based measures, infants were able to follow others' pointing gestures to the next visible object from 10 months onward. These findings, again, support social constructivist views in the way that infants have to follow others' attentional focus to engage in joint attentional episodes. Thus, these preliminary but not stable competences may be first attempts to engage in triadic interaction.

Shortly after infants showed preliminary attention-direction abilities, they started to understand others' pointing gestures in a referential manner. In interaction-based measures, point following to objects within the infants' visual field was related to their later abilities to identify the referent even when it was out of the infants' visual field. Thus, this gradual development is an indication that infants construct a referential understanding of others' perspectives and improve over time. However, the emergence of early attention direction ability and joint engagement skills remains an open and highly discussed question. Some social constructivist views also claim a maturational component of some early abilities or make a special motivation to share psychological states responsible for the start of joint engagement (e.g. Tomasello, 2019), while others suggest social reinforcement being an important factor from early on (see for an overview Del Bianco et al., 2019).

Finding preliminary communicative abilities in earlier months may also be seen as an indication for innate social cognitive systems. Csibra (2003) assumes that the referential system is activated by the use of different communicative cues, such as eye contact, infant directed speech and contingent interaction. Most of these cues were present in the communicative tasks of the current studies. Thus, it may be that a referential action understanding system was activated. In addition, these preliminary abilities may be an indication for an early emerging implicit belief tracking (Apperly, 2013) or a psychological reasoning system (Baillargeon et al., 2016). However, there was no stability in these competences, which does not fit with the assumption of innate abilities. If a social understanding system is present from early on and elicited at one time point, this ability should be present under the same circumstances again one month later. There is no argument for systems being present at specific months but not in the following months.

For action understanding, some preliminary abilities were found as well. Although action understanding in the eye-tracking task was significant on group level only at 12 months, goal anticipation was stable from 8 months on. A significantly different intercept in the growth curve analysis indicated that some infants, but not the majority, showed goal anticipation already at 8 months and did this in a stable manner. However, the increase was the same for all infants independent of the starting point, indicating a gradual increase of action anticipation in eye-tracking tasks. Thus, some infants were able to understand others' actions slightly earlier than others. This fits with social constructivist views, as some infants may have a beneficial environment. The question of the driving force of such individual differences should be examined in further research. However, nativist views would have trouble to explain individual differences between infants.

5.2.2 Situational Influences

The findings of the current studies indicate that abilities are first present within communicative situations. 14-month-olds needed communicative pointing gestures to show VPT and did not show it when a person was only sitting and directed attention to one side. This is in line with previous findings, where either point following to visible objects or visual perspective taking in an interactive task was measured: Abilities were found when pointing was combined with speech (Daum et al., 2013) or when infants interacted in joint attention (Moll et al., 2007). The aim of Study 1 was to measure a higher ability than simple point following but captured with a looking-time task. This is assumed to be easier than motor abilities which are needed within interaction. Thus, also in looking-time measures infants need communication to take others' perspectives and form a referential object expectation. In addition, results of Study 3 indicated that abilities were visible slightly earlier in interaction-based measures than in eye-tracking tasks, where infants had to interpret actions presented on video stimuli. When abilities consolidated in interaction, they were also visible in eye-tracking tasks one month later. This is in line with the social constructivist views, assuming abilities learned through and revealed within interaction, which is communicative per se (Moore & Barresi, 2017).

For younger infants, abilities were not found in the current thesis, even when communicative features were implemented. Nativists might still argue that the communication in our tasks was not salient enough, because necessary cues, such as direct gaze, infant directed speech and contingent interaction were not presented at one time (e.g. Csibra, 2003). This may be the case in the eye-tracking tasks of the current studies but not in the interaction-based tasks. In addition, it had been shown that other attention grabbing actions are also sufficient to elicit gaze following and for example direct gaze is not necessary: 6-month-olds also followed the gaze when a person was shaking its head before looking at an object (Szufnarowska, Rohlfing, Fawcett, & Gredebäck, 2014) or even when doing nothing (Gredebäck, Astor, & Fawcett, 2018).

Pätzold and Liszkowski (2019) argue that 12 month-olds form an object representation when it is communicatively relevant. In their study, communicatively pointing behind a box without direct eye contact with the infant was sufficient to elicit object expectation (Pätzold & Liszkowski, 2019). Thus, not finding a stable indication for referential expectation in earlier months, although a variety of different situations was used, it speaks against an innate referential system as postulated by Csibra (2003).

Others may argue that situations were not salient enough or infants did not have enough time to process the information (e.g. Baillargeon et al., 2016). Often theater scenes are used to measure infants' abilities, as this set up is closer to infants' reality than video scenes. In the studies of the current thesis, looking-time was measured while presenting video stimuli. Thus, it may be that for young infants video scenes are less relevant than theater scenes. However, videos depicted natural action of real persons. In addition, others found abilities for 6-month-olds when presenting video stimuli (e.g. Cannon & Woodward, 2012; Kim & Song, 2015). Thus, there is little empirical evidence supporting this assumption.

Another argument may be that infants did not have enough time to process and understand the situations (e.g. Baillargeon et al., 2016). In looking-time or habituation tasks, infants normally have several seconds and minutes to observe the situations and the measures stop when the infants do not pay attention anymore. The tasks of the current thesis were not controlled by infants and presentation time was limited. Thus, it may be argued, that if infants had more time to process, effects and abilities would have been found in younger ages as well. However, the majority of tasks implemented in the current thesis were already used in other infant studies, where abilities were found in younger infants. For instance, 6-months olds anticipated the goal object in a reaching action within a 3 to 6-second time window (Applin & Kibbe, 2019; Kim & Song, 2015) and 6-9 month olds needed approximately 2.3 milliseconds to follow another's gaze direction (Byers-Heinlein et al., 2020). Thus, the time window of 4 - 6 s used in the current studies for the majority of the tasks, long processing times do not seem to be necessary. Thus, this argument does not seem to be valid for not finding abilities in younger ages.

5.2.3 Relations between different Skills

Finding relations between the development of action understanding and communicative understanding again is an argument for the social constructivist views. At the end of the first year, when the majority of abilities became stable and present in at least half of the infants in Study 3, these different skills were related. Thus, when infants started to understand others' pointing gestures as referential, they were able to anticipate the goal of unfulfilled actions, and the other way around. Finding concurrent relations especially between these higher abilities fit with the assumption that infants' understanding of others transforms after 9-months of age (Tomasello et al., 2005). It can be seen as indication for an integration of the two evolutionary adapted pathways claimed by Tomasello (2019), one for action understanding similar to apes and one with a human-specific motivation to share psychological states.

Directional relations from one skill to the other were bidirectional but very rare. There was one relation indicating that early point following abilities to objects within the infants' visual field with 11 months was related to action understanding in interaction-based measure one month later. This fits with recent findings of Brandone et al. (2019) who found that infants' joint attentional skills are related to later action anticipation in unfulfilled reaching actions. Since attention following is a basic skill for joint engagement, the found relations between point following and action understanding supports the suggestion that infants develop an understanding of others' perspectives and goals through joint attentional skills. In support, Rüther and Liszkowski (under revision) found that parent pointing and infants' own pointing ability predicts infants' referential understanding of others pointing gestures. Other directional relations were from early action understanding to later communicative understanding in eye-tracking tasks, which fits with the assumption that infants need a basic understanding of others as goal directed acting agents to engage in joint attention (Tomasello et al., 2005). However, the directional relations from action understanding to point following were very low, which may be due to measuring a higher action understanding ability with the failed reaching tasks and not the basic action understanding which Tomasello et al. (2005) refer to.

Overall, concurrent relations between action understanding and communicative understanding support that a multifaceted interaction with pointing to, reaching for and handing objects built the basis of social understanding (Liszkowski, 2018; Moore & Barresi, 2017) and infants' own communicative abilities (Rüther & Liszkowski, under revision). Special experience in second-person interaction (Moore & Barresi, 2017) and a human specific motivation to share psychological states (Tomasello et al., 2005) may drive this interplay between different social cognitive skills.

Csibra (2003) argues that communicative and action understanding are separate systems in the first year of life, which are activated by different cues. In contrast to a specific human motivation (Tomasello et al., 2005), he highlights a specific ability to understand others' communicative attempts within the referential system, separated from a teleological system for interpreting instrumental actions. He assumed the two systems to merge only in the second year of life to a mentalistic action interpretation system. However, in the current thesis, relations were found already in the first year of life, which do not fit with the assumptions that the systems are independent from each other. Thus, results of Study 3 can be seen as an argument that social cognitive abilities start to be mentalistic already in the first year of life.

An integrated understanding of others may also be an argument for an inborn psychological reasoning system assumed by Baillargeon et al. (2016). They argue that infants understand others perspectives and actions in an integrated way and could show that 6-month-olds already interpret others' actions based on their visual access (Luo & Johnson, 2009). In Study 3, an integrated understanding was only found at the end of the first year. At this age, infants already had a long time to experience and learn from their surroundings. Thus, although a variety of studies with 12-month-olds are taken as indication for innate abilities, it is problematic to use these findings as argument for innate abilities, as infants had gained substantial social interaction and experience until this time.

5.3 Methodological Implications

One additional focus of this thesis was the methodological validation of different tasks and measures. A variety of methods is commonly used in infant research, like neuropsychological, behavioral or eye-tracking methods. Studies of the current thesis focused on infants' behavior in interaction-based experiments and infants' gaze behavior in eye-tracking tasks. Although tasks in Study 3 were matched very closely between the two kinds of measures, relations between abilities within interaction and infant gaze behavior were very rare. Thus, infants seem to be very sensitive to the designs and the test environment when measuring abilities during such a sensitive period when abilities just emerge. In addition, we did not find abilities earlier in looking-time measures as suggested by some researchers (e.g. Baillargeon et al., 2016). If at all, the few correlations revealed a link in the other direction. Thus, an interactive test environment seems to be beneficial to show early social cognitive abilities, which again supports social constructivist views.

Point following in interaction-based measures, for instance, started one month earlier than point following in video stimuli, although both set-ups were closely matched. In both set-ups, objects were placed within the infant's visual field, situations started with direct gaze, gaze alternation took place in between, and the time to show a response was identical. In addition, the starting points were related. Infants who were able to follow others' pointing gesture within interaction at 10 months were able to follow a pointing gesture present in video stimuli at 11 months. Accordingly, point following in video scenes seems to be more difficult. This is supported by finding concurrent relations between point following in video scenes and point following to objects out of the infants' visual field in the interaction-based measure. Thus, when infants are able to follow a pointing gesture in video scenes, they seem to understand the referential nature of others' pointing gestures. It has to be kept in mind, that in the eye-tracking set up of Study 3, the objects were positioned spatially far away from the pointing gesture. This is different in the majority of tasks, where objects are placed very close to the actor's head or hand (e.g. Senju & Csibra, 2008), which may make the referent more salient even for very young infants. Thus, object position and communicative relevance should be considered when measuring point following.

To measure referential expectation, I implemented tasks, where infants had to understand pointing gestures to hidden entities and either search for the referent object or show indication for object expectation. Again, relations between tasks were rare. Infants who searched for an object by pulling away a cloth at 10 months also showed object expectation indicated by a longer looking-time in an eye-tracking task one month later. However, at 10 months only very few infants showed referential understanding in interaction-based measures and the ability was not yet stable. Thus, this relation has to be interpreted with caution. Even though both abilities started at 11 months and both tasks measured point understanding to hidden objects, these tasks may not measure similar abilities, as we did not find concurrent relations from 11 months on. There may be various reasons for missing correlations, however, identifying them is not within the scope of this thesis. For instance, maybe some infants are better when interacting while others are better in observing situations, at the time when referential understanding just starts to develop. When the ability is consolidated later in development, relations may be more evident.

For action understanding, again, there was only one relation between interaction-based and eye-tracking task, yet before the majority of infants showed an ability in interaction-based measures. However, these preliminary abilities within interaction were stable to the next measuring time point at 12 months. Thus, infants who were able to understand the goal in a failed reaching task and reacted appropriately from early on also anticipated the goal of a failed reaching action in an eye-tracking task. This can be seen as an indication that infants learn to understand others' goals within interaction and later show this ability even in less interactive tasks. Thus, I could not confirm the assumption that the speed of goal anticipation is a precursor for appropriate social interaction (Krogh-Jespersen et al., 2015). However, within interaction, infants saw the person reaching for the object successfully before they had to anticipate the goal. Thus, this methodological difference may have made it easier for infants to anticipate the goal in the interaction-based task. In addition, action understanding within interaction was not measured at every time point, as no previous studies have tested this skill at this young age and some tasks had to be left out due to high tasks demands at one time point. Thus, the methodological validation for action understanding has to be interpreted with caution. However, it adds important information on action understanding and goal anticipation at this young age.

Overall, the results did not reveal stable validity between tasks. Test environment and task set-up seem to be very influential on finding effects, especially in sensitive periods, when infants just start to develop abilities. Thus, several aspects have to be considered when deciding which experimental method to use. For instance, the position of the objects, the processing time, and the information infants receive have to be taken into account. In sum, for measuring social cognitive abilities, highly communicative and interactional situations seem to be beneficial. When using goal anticipation speed or latency as the dependent variable, there was no benefit for observational situations.

5.4 Limitations and Outlook

The current thesis provides important empirical evidence for the development of early social cognitive abilities. Several aspects of different skills were measured with a variety of tasks to examine their emergence and developmental relations. However, the thesis is only a snapshot of early social cognitive development. Thus, it may be argued that some potentially important aspects were left out. I will present some of these aspects, discuss them and provide an outlook for further research.

For measuring action understanding, it may be criticized that very similar tasks were used in the current studies, which may only measure one specific aspect of action understanding. In all tasks, infants had to anticipate the goal without seeing a fulfilled reaching action before. There was only one exception when measuring action anticipation in the interaction-based task, where infants saw the experimenter grasping some objects before being unable to reach the two test items. In general, in most of the studies measuring action understanding at young ages, infants see a person reaching for an object several times and then their expectation to reach for the same object is measured as an indication for understanding others' preferences and goal directedness. Thus, it may be the case that action understanding captured with these kind of tasks measures a different or a simpler ability. Accordingly, interpreting successful reaching actions may emerge earlier in non-communicative situations as well, and may reveal different relations to communicative understanding than found in the current studies. However, there is some evidence that a related understanding is captured when infants have to interpret fulfilled and not fulfilled actions, even though anticipating unsuccessful reaching actions emerges slightly later than interpreting successful ones (Brandone et al., 2014, 2019). In the current thesis, the failed reaching task was used to exclude lower level processes as infants had to anticipate the goal without ever seeing the goal being fulfilled. In addition, the aim was to match both tasks for eye-tracking and interaction-based tasks as closely as possible. Thus, further research could include other aspects of action understanding, like interpreting successful actions, to capture a broader picture but the current thesis already presents information about an important aspect of action understanding.

With the design of Study 3, it cannot be fully ruled out that general attention abilities do not play a role in the development of early social cognitive abilities. The ability to disengage from nonsocial cues in an overlap task was not measured longitudinally but only at one time point in a sensitive period at 8 months of age. At this time point, infants are good at directing their attention but there is still variation. Although there were some relations between general attention and social cognitive skills, relations between social cognitive skills remained when controlling for general attention direction. Thus, it is unlikely that infants' general attention abilities are related to an increase of social cognitive understanding. However, it may be that the developmental growth of general attention is related to the growth of social cognitive abilities. If there are relations, the direction of this relation would be interesting. Do infants develop social cognitive skills on the basis of general attention improvement or do infants improve their general attention abilities through social interaction (Yu & Smith, 2016). Thus, future research should examine influences on the development of social cognitive abilities and include general attention abilities.

An issue, which is present in many infant studies and may be the main cause of replication crisis, is a small sample size. In the cross sectional studies, I did a power calculation and for the communicative cueing paradigm, I had a variety of reference studies (e.g. Csibra & Volein, 2008; Daum et al., 2013). Based on the high effects found in these studies, I was able to test an accurate sample size. For non-communicative tasks, no reference studies existed. Thus, it may be that the sample size of Study 3 was too small to find an effect in the non-communicative task. However, on individual level we did not find a significant difference as well. Even when there is an effect, which is very small, it supports the assumption, that showing abilities is easier in communicative settings. Bayesian testing would be helpful especially to test how likely it is that there is no effect when the effect is not visible. However, without a reference study this was not possible.

For the longitudinal design of Study 3, 47 participants were measured at close time points and this was a moderate sample size, considering that it is not easy to find families who agree to participate in such a long and time-consuming study. However, especially in eye-tracking tasks, data loss was very high and for correlations only a few infants remained. Thus, testing hypothesis between different theories was only possible with correlational analysis, although theory testing with latent models would have been very interesting. It would be especially interesting to test how much variance of a common social cognitive understanding is explained by different skills. Latent growth models gave an indication for individual differences at the starting point. Looking into this in more detail and examining possible influences on these differences should be the aim of further research. In addition, this thesis could serve as a basis to focus on important developmental periods, e.g. 11 and 12 months, where collecting data from more participants would be easier and theory testing with structural equation models would be possible.

As a final outlook, I want to highlight the importance of examining precursors and influences on early social cognitive skills in future research. In the current thesis, only few relations were found from early to later social cognitive abilities. Thus, beside individual abilities, some other factors play a role for the development of social cognitive abilities in the first year of life. As social cognitive abilities were earlier present in interaction-based measures, it seems likely that the social environment is an important factor for developing abilities. There is already evidence that parent communicative interaction with their infants is related to the starting point of infants own pointing behavior (Ger et al., 2018; Rüther & Liszkowski, under review) and their understanding of other communicative acts (Rüther & Liszkowski, under revision). However, the infant's own ability to use a pointing gesture was predictive for their understanding of others' pointing gestures; and especially the parents' reaction to infants' behavior (Brandone et al., 2019; Ger et al., 2018; Rüther, 2019) was a crucial component for the development of social cognitive abilities. Thus, it seems to be the interplay between infants' abilities and their environment, which influences the development of early social cognitive abilities. Future research should, on the one hand, examine which abilities within the infant drive the development of early social cognitive abilities and whether they are based on maturational components or are dependent on interaction. On the other hand, a focus should lie on how different interaction and environments relate to social cognitive abilities. Examining this interplay before social cognitive abilities develop, allows gaining a deeper understanding of the origins of human social cognition and helps to implement helpful intervention in a sensitive period.

5.5 Concluding Remarks

The current thesis provides important empirical evidence for the emergence of early social cognitive abilities, supporting social constructivist views. Different social cognitive abilities emerge as stable and related competences at the end of the first year. First revealing themselves within communicative situations and developing gradually fits with the assumption that infants construct an understanding of others' perspectives and goals through communicative interaction. In addition, slightly different starting points between infants indicate influences on the emergence of social cognitive abilities, which disfavors theories assuming innate systems to be activated at specific time points in all infants. The aim of the current thesis was to examine the development of crucial precursors for later developing human specific social cognition and language. The findings permit the conclusion that communicative and action understanding are not the starting point of social cognition but a developmental outcome. Future research should examine the influences on the emergence of these important early social cognitive abilities.

Appendix A

Appendices

A.1 Appendix - Study 2

A.1.1 Dwell-time Results for the Shortened Outcome Phase in Experiment 1

Dwell-time results of Exp. 1 without the extended 1000ms in the test phase There was no difference in the dwell-time on the empty location between the congruent and incongruent trials, t(19) = 0.47, p = .323 (one-tailed), dz = 0.08 ($M_{congruent}$= 405.63, SD = 259.10; $M_{incongruent}$= 427.71, SD = 278.10).

A.1.2 Dwell-time Results for the Shortened Outcome Phase in Experiment 2b

Dwell-time results of Exp. 2b without the extended 1000ms in the test phase There was no difference in the dwell-time on the empty location between the congruent and incongruent trials, t(21) = 0.55, p = .240 (one-tailed), dz = 0.15 ($M_{congruent}$= 343.09; SD = 373.37; $M_{incongruent}$= 396.50, SD = 333.59).

A.2 Appendix - Study 3

A.2.1 Descriptive Statistics of Point Following in Interaction-based Measure

Table A.1

Descriptive statistics of point following in interaction-based measure

Point Following Interaction-based Measure	Front Target		Back Target		Distractor	
	M	SD	M	SD	M	SD
8-months	0.20	0.17	0.07	0.18	0.18	0.09
9-months	0.39	0.21	0.14	0.21	0.21	0.10
10-months	0.53	0.20	0.24	0.19	0.19	0.10
11-months	0.69	0.20	0.33	0.21	0.21	0.10
12-months	0.77	0.18	0.40	0.17	0.17	0.09

Note. Ratio of point following behavior.

A.2.2 T-test Results of Overall Point Following in Interaction-based Measure

Table A.2

T-test results of overall point following in interaction-based measure

Overall Point Following Interaction-based Measure	Target		Distractor						
	M	SD	M	SD	n	*r*	*t*	*p*	*Cohens d*
8-months	0.14	0.23	0.18	0.21	47	0.34	-1.02	.312	-0.17
9-months	0.27	0.25	0.21	0.24	45	0.17	1.35	.184	0.26
10-months	0.38	0.31	0.38	0.22	40	-0.06	3.06	.004	0.70
11-months	0.51	0.31	0.21	0.21	43	-0.11	4.98	.000	1.13
12-months	0.59	0.32	0.17	0.19	42	-0.18	6.70	.000	1.59

Note. Ratio of point following behavior.

A.2.3 Correlation Coefficient Action Understanding (Mean Latency) and Communicative Understanding in Eye-tracking Tasks

Table A.3

Correlation coefficient action understanding latency and communicative understanding in eye-tracking tasks.

	Action Understanding - Eye-tracking Latency				
	8 months	9 months	10 months	11 months	12 months
Point Following Eye-tracking					
9 months	.16 [-.12, .52]	.11 [-.39, .55]	-.12 [-.45, .35]	-.18 [-.63, .34]	-.05 [-.38, .22]
10 months	.07 [-.31, .38]	.01 [-.38, .46]	-.03 [-.48, .30]	-.20 [-.65, .32]	-.14 [-.50, .19]
11 months	.00 [-.39, .39]	-.16 [-.54, .16]	-.20 [-.45, .07]	-.29 [-.61, .23]	-.26 [-.52, .11]
12 months	-.21 [-.55, .09]	-.15 [-.55, .20]	-.46* [-.66, -.22]	-.30 [-.68, .12]	-.61*** [-.76, -.42]
Referential Expectation Eye-tracking					
8 months	-.23 [-.19, .43]	.20 [-.39, .23]	-.27 [-.45, .32]	-.05 [-.55, .23]	.33 [-.56, .18]
9 months	.04 [-.35, .27]	.01 [-.44, .14]	-.02 [-.40, .33]	.26 [-.51, .43]	-.08 [-.44, .23]
10 months	-.02 [-.29, .40]	-.05 [-.35, .34]	.10 [-.28, .40]	-.12 [-.29, .40]	-.18 [-.10, .52]
11 months	.40 [-.34, .52]	.07 [-.42, .37]	.13 [-.41, .41]	.02 [-.38, .53]	-.04 [-.36, .40]
12 months	.11 [-.41, .50]	-.03 [-.18, .59]	-.09 [.06, .75]	-.36 [.28, .75]	-.14 [-.27, .39]

Note. +p < .1, *p < .05. **p < .025, ***p < .01 ((Bonferroni adjusted alpha level: p < .025; 0.1/(2*2)). Median N = 25 [17; 31]

A.2.4 Correlation Coefficient Action Understanding in Eye-tracking (Mean Latency) and Action Understanding in Interaction-based Task

Table A.4

Correlation coefficient action understanding in eye-tracking (mean latency) and action understanding in interaction-based measure.

	Action Understanding - Eye-tracking Latency				
	8 months	9 months	10 months	11 months	12 months
Action Understanding Interaction-based					
10 months	-.13	-.03	-.19	-.29	-.05
	[-.47, .25]	[-.36, .28]	[-.59, .35]	[-.68, .15]	[-.25, .49]
12 months - out of reach	.18	-.13	-.34	-.06	-.09
	[-.18, .44]	[-.51, .29]	[-.59, -.01]	[-.45, .36]	[-.48, .30]
14 months - out of reach	.13	.29	.19	.01	.15
	[-.26, .41]	[-.28, .67]	[-.13, .67]	[-.44, .50]	[-.41, .67]
12 months - hidden	.24	.17	-.09	-.24	-.06
	[-.05, .51]	[-.29, .62]	[-.38, .24]	[-.68, .27]	[-.41, .34]
14 months - hidden	.19	.18	.46*	-.33	.15
	[-.31, .59]	[-.24, .50]	[.18, .68]	[-.65, .04]	[-.26, .51]

Note. +p < .1, *p < .05. **p < .025, ***p < .01 (Bonferroni adjusted alpha level: p < .025; 0.1/(2*2). Median N = 25 [19; 29]

References

Apperly, I. A. (2013). Can theory of mind grow up? mindreading in adults, and its implications for the development and neuroscience of mindreading. In S. Baron-Cohen, M. Lombardo, & H. Tager-Flusberg (Eds.), *Understanding other minds.* Oxford University Press.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological review*, *116*(4), 953–970. doi: 10.1037/a0016923

Applin, J. B., & Kibbe, M. M. (2019). Six–month–old infants predict agents' goal–directed actions on occluded objects. *Infancy*, *24*(3), 392–410. doi: 10.1111/infa.12282

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual review of psychology*, *67*, 159–186. doi: 10.1146/annurev-psych-010213-115033

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 1278–1289.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *The British journal of developmental psychology*, *30*(3), 359–375. doi: 10.1111/j.2044-835X.2011.02043.x

Bertenthal, B. I., Boyer, T. W., & Harding, S. (2014). When do infants begin to follow a point? *Developmental psychology*, *50*(8), 2036–2048. doi: 10.1037/a0037152

Brandone, A. C., Horwitz, S. R., Aslin, R. N., & Wellman, H. M. (2014). Infants' goal anticipation during failed and successful reaching actions. *Developmental science*, *17*(1), 23–34. doi: 10.1111/desc.12095

Brandone, A. C., Stout, W., & Moty, K. (2019). Triadic interactions support infants' emerging understanding of intentional actions. *Developmental science*, e12880. doi: 10.1111/desc.12880

Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: how infants interpret adult looking behavior. *Developmental psychology*, *38*(6), 958.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(01), 131. doi: 10.1017/S030500090700829X

Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: a longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of experimental child psychology*, *130*, 67–78. doi: 10.1016/j.jecp.2014.09.010

Brune, C. W., & Woodward, A. L. (2007). Social cognition and social responsiveness in 10-month-old infants. *Journal of Cognition and Development*, *8*(2), 133–158. doi: 10.1080/15248370701202331

Bukowski, H., Hietanen, J. K., & Samson, D. (2015). From gaze cueing to perspective taking: Revisiting the claim that we automatically compute where or what other people are looking at. *Visual cognition*, *23*(8), 1020–1042. doi: 10.1080/13506285.2015.1132804

Butler, S. C., Caron, A. J., & Brooks, R. (2000). Infant understanding of the referential nature of looking. *Journal of Cognition and Development*, *1*(4), 359–377.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*(5), 606–637.

Byers-Heinlein, K., Tsui, R. K. Y., van Renswoude, D., Black, A. K., Barr, R., Brown, A., ... Singh, L. (2020). *The development of gaze following in monolingual and bilingual infants: A multi-lab study.* doi: 10.31234/osf.io/sgfhv

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal–based action predictions. *Developmental science*, *15*(2), 292–298.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, i–174.

Cole, G. G., Smith, D. T., & Atkinson, M. A. (2015). Mental state attribution and the gaze cueing effect. *Attention, perception & psychophysics*, *77*(4), 1105–1115. doi: 10.3758/s13414-014-0780-6

Collicott, C., Collins, S., & Moore, C. (2009). The development of gaze following in a third–party context. *Infancy*, *14*(3), 363–376.

Colonnesi, C., Rieffe, C., Koops, W., & Perucchini, P. (2008). Precursors of a theory of mind: A longitudinal study. *British Journal of Developmental Psychology*, *26*(4), 561–577. doi: 10.1348/026151008X285660

Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1431), 447–458. doi: 10.1098/rstb.2002.1235

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, *107*(2), 705–717. doi: 10.1016/j.cognition.2007.08.001

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168. doi: 10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in cognitive sciences*, *13*(4), 148–153. doi: 10.1016/j.tics.2009.01.005

Csibra, G., & Gergely, G. (2013). Teleological understanding of actions. *Navigating the social world: What infants, children, and other species can teach us*, 38–43.

Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason'in infancy. *Cognition*, *72*(3), 237–267.

Csibra, G., & Volein, Á. (2008). Infants can infer the presence of hidden objects from referential gaze information. *British Journal of Developmental Psychology*, *26*(1), 1–11. doi: 10.1348/026151007X185987

Dahl, A., & Brownell, C. A. (2019). The social origins of human prosociality. *Current Directions in Psychological Science*, *28*(3), 274–279. doi: 10.1177/0963721419830386

Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska, E. P., & Elison, J. T. (2018). An examination of recording accuracy and precision from eye tracking data from toddlerhood to adulthood. *Frontiers in psychology*, *9*, 803. doi: 10.3389/fpsyg.2018.00803

Daum, M. M., & Gredebäck, G. (2011). The development of grasping comprehension in infancy: covert shifts of attention caused by referential actions. *Experimental brain research*, *208*(2), 297–307. doi: 10.1007/s00221-010-2479-9

Daum, M. M., Ulber, J., & Gredebäck, G. (2013). The development of pointing perception in infancy: effects of communicative signals on covert shifts of attention. *Developmental psychology*, *49*(10), 1898–1908. doi: 10.1037/a0031111

de Groote, I., Roeyers, H., & Striano, T. (2007). Gaze following as a function of affective expression in 3-, 6- and 9-month-old infants. *Infant behavior & development*, *30*(3), 492–498. doi: 10.1016/j.infbeh.2006.12.009

Del Bianco, T., Falck-Ytter, T., Thorup, E., & Gredebäck, G. (2019). The developmental origins of gaze–following in human infants. *Infancy*, *24*(3), 433–454. doi: 10.1111/infa.12276

Elsabbagh, M., Fernandes, J., Webb, S. J., Dawson, G., Charman, T., & Johnson, M. H. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological psychiatry*, *74*(3), 189–194. doi: 10.1016/j.biopsych.2012.11.030

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of experimental child psychology*, *85*(3), 199–212. doi: 10.1016/S0022-0965(03)00022-5

Ferguson, H. J., Apperly, I., & Cane, J. E. (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *The Quarterly Journal of Experimental Psychology*, *70*(8), 1646–1660.

Fizke, E., Butterfill, S., van de Loo, L., Reindl, E., & Rakoczy, H. (2017). Are there signature limits in early theory of mind? *Journal of experimental child psychology*, *162*, 209–224. doi: 10.1016/j.jecp.2017.05.005

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

Flom, R., Deák, G. O., Phill, C. G., & Pick, A. D. (2004). Nine-month-olds' shared visual attention as a function of gesture and object location. *Infant Behavior and Development*, *27*(2), 181–194. doi: 10.1016/j.infbeh.2003.09.007

Flom, R., & Johnson, S. (2011). The effects of adults' affective expression and direction of visual gaze on 12-month-olds' visual preferences for an object following a 5-minute, 1-day, or 1-month delay. *The British journal of developmental psychology*, *29*(Pt 1), 64–85. doi: 10.1348/026151010X512088

Flom, R., & Pick, A. D. (2005). Experimenter affective expression and gaze following in 7-month-olds. *Infancy*, *7*(2), 207–218. doi: 10.1207/s15327078in0702{\textunderscore}5

Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of experimental psychology. Human perception and performance*, *42*(2), 158–163. doi: 10.1037/xhp0000138

Gardner, M. R., Hull, Z., Taylor, D., & Edmonds, C. J. (2018). 'spontaneous' visual perspective-taking mediated by attention orienting that is voluntary and not reflexive. *Quarterly Journal of Experimental Psychology*, *71*(4), 1020–1029.

Ger, E., Altınok, N., Liszkowski, U., & Küntay, A. C. (2018). Development of infant pointing from 10 to 12 months: The role of relevant caregiver responsiveness. *Infancy*, *23*(5), 708–729. doi: 10.1111/infa.12239

Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological science*, *20*(3), 347–353.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, *138*(6), 1085–1108. doi: 10.1037/a0028044

Gredebäck, G., Astor, K., & Fawcett, C. (2018). Gaze following is not dependent on ostensive cues: A critical test of natural pedagogy. *Child Development*. doi: 10.1111/cdev.13026

Gredebäck, G., & Daum, M. M. (2015). The microstructure of action perception in infancy: Decomposing the temporal structure of social information processing. *Child Development Perspectives*, *9*(2), 79–83. doi: 10.1111/cdep.12109

Gredebäck, G., Johnson, S., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental neuropsychology*, *35*(1), 1–19. doi: 10.1080/87565640903325758

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological bulletin*, *141*(6), 1266–1287. doi: 10.1037/bul0000002

Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as i do: 7-month-old infants selectively reproduce others' goals. *Developmental science*, *11*(4), 487–494. doi: 10.1111/j.1467-7687.2008.00694.x

Hemmerich, W. (2019). *Winsorizing-rechner: Statistikguru.*

Hertenstein, M. J., & Campos, J. J. (2004). The retention effects of an adult's emotional displays on infant behavior. *Child Development*, *75*(2), 595–613.

Heyes, C. (2014a). False belief in infancy: a fresh look. *Developmental science*, *17*(5), 647–659. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1111/desc.12148 doi: 10.1111/desc.12148

Heyes, C. (2014b). Submentalizing: I am not really reading your mind. *Perspectives on psychological science : a journal of the Association for Psychological Science*, *9*(2), 131–143. doi: 10.1177/1745691613518076

Heyes, C. (2016). Born pupils? natural pedagogy and cultural pedagogy. *Perspectives on psychological science : a journal of the Association for Psychological Science*, *11*(2), 280–295. doi: 10.1177/1745691615621276

Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proceedings. Biological sciences*, *282*(1819). doi: 10.1098/rspb.2015.1683

Kim, E. Y., & Song, H.-j. (2015). Six-month-olds actively predict others' goal-directed actions. *Cognitive Development*, *33*, 1–13. doi: 10.1016/j.cogdev.2014.09.003

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science (New York, N.Y.)*, *330*(6012), 1830–1834. doi: 10.1126/science.1190792

Krcmar, M. (2010). Can social meaningfulness and repeat exposure help infants and toddlers overcome the video deficit? *Media Psychology*, *13*(1), 31–53. doi: 10.1080/15213260903562917

Krogh-Jespersen, S., Liberman, Z., & Woodward, A. L. (2015). Think fast! the relationship between goal prediction speed and social competence in infants. *Developmental science*, *18*(5), 815–823. doi: 10.1111/desc.12249

Krogh-Jespersen, S., & Woodward, A. (2016). Infants origins of soical cognition. In L. Balter & C. Tamis-LeMonda (Eds.), *Child psychology.*

Liszkowski, U. (2018). Emergence of shared reference and shared minds in infancy. *Current opinion in psychology*, *23*, 26–29. doi: 10.1016/j.copsyc.2017.11.003

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, *108*(3), 732–739. doi: 10.1016/j.cognition.2008.06.013

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489–512. doi: 10.1016/j.cognition.2006.10.007

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

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental science*, *12*(1), 142–149. doi: 10.1111/j.1467-7687.2008.00741.x

Marno, H., Farroni, T., Vidal Dos Santos, Y., Ekramnia, M., Nespor, M., & Mehler, J. (2015). Can you see what i am talking about? human speech triggers referential expectation in four-month-old infants. *Scientific reports*, *5*, 13594. doi: 10.1038/srep13594

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental psychology*, *31*(5), 838–850. doi: 10.1037/0012-1649.31.5.838

Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental science*, *10*(6), 826–835. doi: 10.1111/j.1467-7687.2007.00615.x

Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in human neuroscience*, *7*, 558. doi: 10.3389/fnhum.2013.00558

Moll, H., & Tomasello, M. (2004). 12–and 18–month–old infants follow gaze to spaces behind barriers. *Developmental science*, *7*(1).

Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental psychology*, *43*(2), 309–317. doi: 10.1037/0012-1649.43.2.309

Moore, C., & Barresi, J. (2017). The role of second-person information in the development of social understanding. *Frontiers in psychology*, *8*, 1667. doi: 10.3389/fpsyg.2017.01667

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*(3), 718–735.

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.

Oakes, L., & Amso, D. (2018). Development of visual attention. In J. T. Wixted & S. S. Stevens (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (Vol. 76, pp. 1–33). New York: John Wiley & Sons, Inc. doi: 10.1002/9781119170174.epcn401

Okamoto-Barth, S., & Kawai, N. (2006). The role of attention in the facilitation effect and another "inhibition of return". *Cognition*, *101*(3), B42–B50.

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

Pätzold, W., & Liszkowski, U. (2019). Pupillometry reveals communication-induced object expectations in 12- but not 8-month-old infants. *Developmental science*, *22*(6), e12832. doi: 10.1111/desc.12832

Reifman, A., & Keyton, K. (2010). Winsorize. In N. J. Salkind (Ed.), *Encyclopedia of research design.* Sage.

Rochat, P., Querido, J. G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental psychology*, *35*(4), 950–957. doi: 10.1037/0012-1649.35.4.950

Rohlfing, K. J., Longo, M. R., & Bertenthal, B. I. (2012). Dynamic pointing triggers shifts of visual attention in young infants. *Developmental science*, *15*(3), 426–435. doi: 10.1111/j.1467-7687.2012.01139.x

Ruff, H. A., Capozzoli, M., & Saltarelli, L. M. (1996). Focused visual attention and distractibility in 10-month-old infants. *Infant Behavior and Development*, *19*(3), 281–293.

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review*, *34*(3), 265–293. doi: 10.1016/j.dr.2014.04.001

Rüther, J. (2019). Socialcultural influences on the development of infants' cultural learning skills in the first two years of life. *Dissertation Thesis.*

Rüther, J., & Liszkowski, U. (under review). Ontogeny of index-finger pointing. *Child Development.*

Rüther, J., & Liszkowski, U. (under revision). Ontogenetic emergence of cognitive reference comprehension. *Cognitive Science*.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of experimental psychology. Human perception and performance*, *36*(5), 1255–1266. doi: 10.1037/a0018729

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: implicit mentalizing or domain-general processing? *Journal of experimental psychology. Human perception and performance*, *40*(3), 929–937. doi: 10.1037/a0035175

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current biology : CB*, *18*(9), 668–671. doi: 10.1016/j.cub.2008.03.059

Silverstein, P., Gliga, T., Westermann, G., & Parise, E. (2019). Probing communication-induced memory biases in preverbal infants: Two replication attempts of yoon, johnson and csibra (2008). *Infant behavior & development*, *55*, 77–87. doi: 10.1016/j.infbeh.2019.03.005

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition*, *189*, 260–274. doi: 10.1016/j.cognition.2019.03.019

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental psychology*, *51*(9), 1190–1200. doi: 10.1037/dev0000039

Sodian, B., Licata, M., Kristen-Antonow, S., Paulus, M., Killen, M., & Woodward, A. (2016). Understanding of goals, beliefs, and desires predicts morally relevant theory of mind: A longitudinal investigation. *Child Development*, *87*(4), 1221–1232. doi: 10.1111/cdev.12533

Sodian, B., Thoermer, C., & Metz, U. (2007). Now i see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental science*, *10*(2), 199–204. doi: 10.1111/j.1467-7687.2007.00580.x

Southgate, V., Johnson, M. H., El Karoui, I., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological science*, *21*(3), 355–359. doi: 10.1177/0956797610362058

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130*(1), 1–10. doi: 10.1016/j.cognition.2013.08.008

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, *10*(1), 89–96. doi: 10.1111/j.1467-7687.2007.00569.x

Surtees, A., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, *83*(2), 452–460. doi: 10.1111/j.1467-8624.2011.01730.x

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97–105. doi: 10.1016/j.cognition.2015.12.010

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? *Scientific reports*, *4*, 5304. doi: 10.1038/srep05304

Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, perception & psychophysics*, *72*(3), 695–705. doi: 10.3758/APP.72.3.695

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(34), 8491–8498. doi: 10.1073/pnas.1804761115

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Belknap Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). In search of the uniquely human. *Behavioral and Brain Sciences*, *28*(5), 721–727.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705–722. doi: 10.1111/j.1467-8624.2007.01025.x

Tomasello, M., & Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental psychology*, *39*(5), 906–912. doi: 10.1037/0012-1649.39.5.906

Uithol, S., & Paulus, M. (2014). What do infants understand of others' action? a theoretical account of early social cognition. *Psychological research*, *78*(5), 609–622. doi: 10.1007/s00426-013-0519-3

von Hofsten, C., Dahlström, E., & Fredriksson, Y. (2005). 12-month-old infants' perception of attention direction in static video images. *Infancy*, *8*(3), 217–231. doi: 10.1207/s15327078in0803{\textunderscore}2

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children*, *23*(3), 34–41.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, *11*(3), 271–294.

Wellman, 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.

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

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, *18*(1), 53–57.

Wronski, C., & Daum, M. M. (2014). Spatial orienting following dynamic cues in infancy: grasping hands versus inanimate objects. *Developmental psychology*, *50*(8), 2020–2029. doi: 10.1037/a0037155

Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*, *105*(36), 13690–13695.

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current biology : CB*, *26*(9), 1235–1240. doi: 10.1016/j.cub.2016.03.026