The Social Attentional Foundations of Infants' Learning from Third-Party Social Interactions

Von der Fakultät für Lebenswissenschaften

der Universität Leipzig

genehmigte

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doctor rerum naturalium

Dr. rer. nat.

vorgelegt von

M. Sc. Psychologie, Kyra Maleen Thiele

geboren am 26. Mai 1989 in Münster

Dekan: Prof. Dr. Marc Schönwiesner
Gutachter: Prof. Dr. Daniel Haun
                  Prof. Dr. Stefanie Höhl

Tag der Verteidigung: 04. November 2021

Human infants rely on social interactions to acquire culturally relevant knowledge about their environment. Aside from active participation ("first-party perspective"), infants encounter social interactions through third-party observation ("third-party perspective"). Despite the absence of own involvement, the mere observation of others' interactions represents an essential source of social learning opportunities. The overarching aim of this dissertation was to deepen our understanding of the foundations of infants' observational learning from third-party interactions. This was achieved by investigating (a) social attentional developments and motivational influences driving infants' attention toward third-party interactions (Study 1 & 2), and (b) factors influencing infants' attention and memory while observing third-party interactions (Study 3).

Study 1 investigated how infants' attentional orienting to third-party interactions develops in parallel with their active social attention behavior. In Experiment 1, 9.5- to 11-month-old infants looked longer than 7- to 8.5-month-olds at videos showing two adults engaging in a face-to-face interaction, when simultaneously presented with a non-interactive back-to-back scene showing the same people acting individually. Moreover, older infants showed higher social engagement (including joint attention) during parent-infant free play. Experiment 2 replicated this age-related increase in both measures and showed that it follows continuous trajectories from 7 to 13 months of age. These findings suggest that infants' attentional orienting to others' social interactions coincides with developments in their social attention behavior during own social interactions.

Study 2 examined the incentive value of social interactions as a proximal driver of infants' attentional orienting to third-party interactions. In a gaze-contingent associative learning task, two geometrical shape cues were repeatedly paired with two kinds of target videos showing either a dyadic face-to-face interaction or a non-interactive back-to-back scene. We found that 13-month-old infants performed faster saccadic latencies and more predictive gaze shifts toward the cued target region during social interaction trials. This suggests that social interaction targets can serve as primary reinforcers in an associative learning task, supporting the view that infants find it intrinsically rewarding to observe others' social interactions.

Study 3 investigated infants' object encoding in the context of observed social interactions. In Experiment 1, 9-month-old infants were presented with four types of videos showing one object and two adults. The scenarios varied regarding the eye contact between the adults (eye contact or no eye contact) and the adults' object-directed gaze (looking toward or away from the object). Infants showed increased object encoding, but only when seeing two adults looking at an object together, following mutual eye contact. We found an identical pattern of results in a matched first-party design during which 9-month-old infants were directly addressed by one single adult on screen (Experiment 2). Together, these findings suggest that the capacity to learn about novel objects by observing third-party interactions emerges in the first postnatal year, and that it may depend on similar factors as infants' learning through direct social interactions at this age.

The findings of all three studies are integrated in a general discussion. In summary, the results of this thesis suggest that, throughout the first year after birth, infants develop abilities and preferences enabling them to approach and efficiently learn from third-party social interactions.

Acknowledgements

I am full of gratitude for the enormous support, help, and guidance of many people. This dissertation would not have been possible without you, and I want to say thank you. First, to my first supervisor, Daniel. Thank you for your constant trust and support, your immense generosity and understanding, and for always giving me the feeling of having my back—especially after the break-in. You never pushed me in any direction but instead provided a safe environment that gave me the time, the space, and the courage to develop my own interests and find my own way. I know now that this is not the norm and I am deeply grateful.

Second, I want to thank Robert, my second supervisor. It may seem ironic, but without DesEff I would probably not have started a PhD or even considered working in science. Thank you for this experience and for being an essential member of my PhD team since my very first steps. I am sincerely thankful for your mentoring advice, your constructive criticism and honesty, your solution-oriented pragmatism, and for helping me structure my thoughts like no one else can do.

I would certainly never have achieved any of this without my unofficial third supervisor, Christine. That you joined my PhD team for my first study and remained on board for the following ones was one of the most fortunate things that happened to me during my PhD. The way you care about others is remarkable and I am thankful for having experienced this firsthand. You were always there for me, without ever expecting anything in return and I cannot thank you enough for that.

The fourth person I would like to thank is Gustaf. The time in your lab has been a very special and formative episode of my PhD for me. From the first day on, you gave me the feeling of being warmly welcomed, making it very easy for me to settle in. Thank you for this great opportunity, for everything you taught me about latencies and learning, for always keeping your door open for advice, your kindness and generosity, your enormous support and responsiveness, and for sharing your infectious enthusiasm for the small details. I would also like to thank the entire Uppsala Child and Babylab team, especially Josh, Anton, Kim, Kahl, Linda, and Jessica, who treated me like a member of their team and made my stay an unforgettable experience. Kim, I feel honored to be a part of GAZMO. It was and still is a pleasure and great fun working with you.

Very special acknowledgements go to my PhD siblings, Anne, Marie, Roman, Theo, Sarah, David, and Sarah, who accompanied me on most of this journey. I am very grateful that I could share so many first times with you—the first teaching sessions, conference travels, posters, and talks. The time with you has been incredibly intense and memorable for me. I will never forget your endless

and unconditional support, your comfort and warmth, your empathy and sensitivity, your patience, your encouragement, your honesty, your advice and supervision, your constant readiness to help, your inspiration, your loyalty, and your selflessness—at work and far beyond. Thank you for celebrating the successes and failures, for patiently listening to my complaints without ever complaining, and for becoming true friends. David, the master of hugging, I did not forget that you checked on me every day for weeks after the break-in. Sarah, my secret twin and longest office mate! You are one of the most thoughtful and caring persons I know. I am so grateful that I have been on this road together with you, from constructing DesEff apparatuses to writing up our theses. Theo, thank you for countless clever ideas and study advice and for reminding me about the important things in life. Sarah, for never giving up on pushing my sociality, for taking my paranoid thoughts as seriously as my rational ones, for suppressing your natural impulse to sing and whistle, and for making our work place feel like home to me. Also, a huge thank you to both of you, Sarah and Sarah, for your constant help with translation issues! Roman, thank you for always and at any time being there for me, for numerous inspiring conversations that strongly influenced my studies, for never saying no to commenting on manuscript drafts, and for making the back-to-back situation almost comfortable. Anne! You truly are a "Fels in der Brandung", no matter what happens. Thank you.

To all my friends and colleagues who acted in the video stimuli of the studies: Sarah, Anne, Sarah, Charlotte, Claudia, Ariane, Amelie, and Anja. Thank you for your patience with my impatience and for practicing and performing the weirdest choreographies. I also owe a huge thank you to all the babies and families who watched the results of these productions in Leipzig and Uppsala! Without your participation, this dissertation would not have been possible.

Thank you to the entire LFE family! Especially to Katja, for coordinating most of my studies, and to Madlen, for never making me feel bad about repeatedly asking the same questions concerning Dienstreiseanträge and Reisekostenabrechnungen. To Sascha, for his great technical support! The constantly growing CCP team! Thank you to all of you! Especially to Anja, for navigating me through the paperwork jungle at the MPI, to Katharina, Petra, and Susanne, for making it possible to test a study in the middle of a pandemic, to Clara and Nori, for supporting me with participant recruitment, and to Steven, for teaching me basically any tech-related skill from video editing to programming. To Nori and Johanna, for supporting me with reliability coding, and again to Nori, for her beautiful artworks. Thank you to Roger, for helping me with statistical advice. To Denis! For many thought-provoking discussions and challenging questions that were always (if I understood them) a great source of inspiration, especially during my final stretch of writing. Roman and Sarah L.P., thank you for taking the time to read this thesis and for your immensely helpful feedback!

My parents and my sister! Monika, Volkhard, and Leonora, I am deeply thankful for your unconditional support, your unlimited understanding and patience, and your boundless and unshakable faith in me. And finally, to Steven—for everything.

Summary

Human infants strongly rely on social interactions to acquire culturally relevant knowledge about their environment (Csibra & Gergely, 2011; Reid & Striano, 2007; Tomasello, 2016). In their daily lives, infants encounter social interactions in multiple ways, including own participation ("first-party perspective") and third-party observation ("third-party perspective"). Despite the absence of own involvement, the mere observation of others' social interactions represents an essential source of social learning opportunities for infants (Paradise & Rogoff, 2009).

During the past ten years, an increasing number of studies have investigated how infants represent social interactions between third parties (e.g., Beier & Spelke, 2012; Papeo, 2020) and how they come to use third-party interactions to learn about their environment (Matheson, Moore, & Akhtar, 2013). The growing interest in the third-party perspective has accompanied a generally increasing awareness that the extent to which infants encounter direct teaching practices and, thus, rely on observational learning, varies significantly across individuals and cultural contexts (Gaskins & Paradise, 2010). To better account for the diversity of early social learning strategies, researchers have emphasized the importance of studying observational forms of learning in addition to direct participatory forms (Rogoff, Paradise, Arauz, Correa-Chávez, & Angelillo, 2003).

In contributing to this research, the overarching aim of this dissertation was to deepen our understanding of the foundational processes and mechanisms involved in infants' observational learning from third-party interactions. This was achieved by taking a broader perspective, which not only focuses on infants' attention and memory during the actual learning situation (Study 3) but also considers social attentional developments and underlying motivational mechanisms enabling infants to detect and approach third-party interactions (Study 1 & 2).

Studies and Results

The empirical part of this dissertation comprises three preregistered experimental studies which address three gaps in the previous third-party interaction literature. First, there has been a lack of studies investigating the influence of infants' intrinsic motivations on their attention and learning in the context of third-party interactions. Within the framework of this thesis, I consider endogenous motivations like infants' intrinsic social motivation as "door openers" to potential social learning opportunities, as they drive infants' attention toward social interactions. The

second shortcoming was the lack of studies investigating infants' learning from third-party interactions within the first year of life. In the context of third-party interactions, object-related learning had been earliest demonstrated in the second year of life. Third, it remained unclear what factors promote infants' learning about novel objects within observed interactions.

All three studies presented in this thesis relied on measures of overt visual attention, recorded via screen-based eye tracking (Studies 1–3) or manual coding of infants' looking behavior during a naturalistic free play interaction (Study 1). The overall study population consisted of 7- to 14.5-month-old typically developing infants growing up in a Western, urban, industrialized environment.

Study 1: How does infants' orienting to third-party social interactions develop during the second half of the first year of life, and how does this development coincide with changes in infants' active social attention behavior?

Previous research had revealed that 14-month-old infants preferentially orient their visual attention to situations in which they can observe two people engaging in a face-to-face interaction (over non-interactive back-to-back scenes, Galazka, Roché, Nyström, & Falck-Ytter, 2014). However, the ontogenetic origins of this attentional preference remained unclear. To address this gap, the first aim of Study 1 was to investigate how infants' attentional orienting toward third-party social interactions develops between 7 and 13 months of age—a period during which infants' social attention behavior in first-party interactions undergoes a significant development (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Clearfield, Osborne, & Mullen, 2008; Striano & Rochat, 1999). The second aim of Study 1 was to compare infants' attentional orienting toward third-party interactions with their social orienting in first-party interactions.

The study consisted of two experiments testing different age ranges. In Experiment 1, we compared two age groups of infants: one group consisting of 7- to 8.5-month-old infants ($N = 20$) and one group consisting of 9.5- to 11-month-olds ($N = 20$). In Experiment 2, we tested infants at a broader and continuous age range between 7 and 13.5 months of age ($N = 50$). All participants came from Leipzig (Germany). To assess infants' attentional orienting toward third-party interactions, we measured their looking times while they were simultaneously presented with two types of videos in a preferential-looking task. One video showed a dyadic face-to-face interaction between two adults (a clapping game, touching their hands, leaning their heads toward one another). The other video showed a mirrored non-interactive scene during which the same actors performed the identical motions individually while standing back-to-back. To assess infants' social attention behavior in first-party interactions, we recorded their interaction behavior during a naturalistic free play session with their caregiver. We coded four infant looking behaviors in relation to their caregiver: looks in the direction of their caregiver, looks at the caregiver's face, eye contact, and joint attention looks.

Our results in Experiment 1 revealed that 9.5- to 11-month-old infants showed a stronger looking time preference than 7- to 8.5-month-olds for videos showing a dyadic face-to-face interaction over videos showing a non-interactive back-to-back scene. Moreover, during active social interaction with their caregiver, infants in the older age group showed more social looking behaviors (especially joint attention) than younger infants (see also Carpenter, Nagell, et al., 1998; Striano & Rochat, 1999). In Experiment 2, we could replicate our findings from Experiment 1 and extend them by showing that the increase in both measures follows continuous developmental trajectories from 7 to 13 months of age. In a merged sample over both experiments ($N = 90$), infants' orienting toward others' interactions was positively correlated with their social attention during own social engagement, but this correlation was mainly driven by the infants' age.

Together, the results from Study 1 show that infants' attentional orienting toward others' interactions coincides with developments in their social attention during own social interactions. This suggests that more general motivational mechanisms may steer infants' attention toward both direct social interaction partners and social interactions between others.

Study 2: Is it intrinsically rewarding for infants to observe third-party interactions?

Another question that remained open from the previous literature is what proximal mechanisms underlie infants' attentional orienting toward third-party social interactions. In direct social interactions, infants show increasing signs of seeking and liking social engagement toward the end of the first year of life, suggesting that they find it intrinsically rewarding to engage in social interactions (Striano & Bertin, 2005; Venezia, Messinger, Thorp, & Mundy, 2004). Study 2 aimed to investigate whether similar reward mechanisms drive infants' attention toward third-party interactions. We used reinforcement learning as an indicator of intrinsic reward value (Berridge & Robinson, 2003).

Thirty-two infants between 13 and 14.5 months from Uppsala (Sweden) participated in the study. In a gaze-contingent associative visual learning task, infants saw two non-social cues (circle or triangle) repeatedly paired with two kinds of target videos: a dyadic face-to-face interaction or a non-interactive back-to-back scene (based on the stimuli created for Study 1). Every infant was presented 12 trials per condition, resulting in 24 trials in total. At the beginning of each trial, one of the two cues appeared in the center of the screen. When the infant looked at it, the cue disappeared, and the associated target videos appeared right or left from the cue. To assess infants' associative learning performance, we measured their saccadic latencies and anticipatory gaze shifts to the correct target region across trials. Based on reward learning principles (Skinner, 1938), we assumed that if it is rewarding for infants to observe third-party interactions, they should learn the association between an arbitrary shape cue and a target video more effectively if the target shows a social interaction as compared to a non-interactive control scene (see also Tummeltshammer, Feldman, & Amso, 2019).

We found that infants showed faster saccadic latencies and more predictive gaze shifts during social interaction trials over control trials, indicating superior learning of cue-target associations guiding them to videos depicting a dyadic face-to-face interaction. These findings suggest that, without additional reinforcement, social interaction targets can selectively power infants' associative learning, indicating that third-party interactions can serve as primary reinforcers in 13-month-old infants.

Study 3: Can infants in the first year of life use third-party interactions to learn about objects? If so, do similar processes contribute to infants' observational learning from third-party interactions as to their referential learning in direct social interactions?

In the context of third-party interactions, object-related learning had been earliest demonstrated at 18 months in previous studies (Akhtar, 2005; Matheson et al., 2013). It remained unclear whether already preverbal infants can learn about objects through merely observing others' interactions and, if so, whether similar factors contribute to this observational learning as to their referential learning through direct social engagement (i.e., an interplay between eye contact and object-directed gaze). To address this question, we used a looking time based object-encoding task.

Across two experiments, we compared 9-month-old infants' object encoding in a third-party observational context (Experiment 1) with a first-party context (Experiment 2). In Experiment 1, $N = 32$ infants between 9 and 10 months from Leipzig (Germany) were presented with four types of videos showing two adults and one object (the "familiarized" object). Based on a 2 × 2 design, the videos varied systematically regarding the eye contact between the two agents (eye contact or no eye contact) and their object-related gaze (looking toward or away from the object). In Experiment 2, another sample of 9-month-old infants ($N = 32$) was tested in a matched first-party design during which they were directly addressed by an adult on screen. As an indicator of object encoding, we used infants' novelty preference, a defining feature of the infant visual recognition memory. After each video, we presented the familiarized object next to a novel object in a preferential-looking phase. We assumed that longer looking times to the novel object would reversely indicate an increased previous encoding of the familiarized object (Rose, Melloy-Carminar, & Bridget, 1982).

In Experiment 1, infants showed an increased novelty preference, but only after observing two adults attending to the familiarized object together, following mutual eye contact. Moreover, our results in Experiment 2 replicated the previous finding, that also in direct interactions infants' object encoding depends on the interplay between eye contact and object-directed gaze. Like in Experiment 1, infants' object encoding was only enhanced when an adult made eye contact with them before shifting their gaze toward a visible object (see also Cleveland & Striano, 2007; Okumura, Kanakogi, Kobayashi, & Itakura, 2020).

Taken together, the results from Study 3 suggest that already 9-month-old infants can learn about the perceptual features of novel objects through merely observing triadic "person-person-object" interactions between others. Moreover, the findings indicate that similar factors may influence early observational learning as object learning in self-experienced triadic interactions at this age.

Discussion

In summary, the empirical work of this thesis adds three substantial contributions to our understanding of infants' attention to and learning from third-party social interactions. First, the presented studies consider infants' intrinsic motivations as proximal drivers toward third-party social interactions. Following social motivation frameworks (Berridge & Robinson, 2003; Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012), the current results provide supporting evidence for two psychological components of social motivation: social orienting (Study 1) and social reward (Study 2). Second, this thesis extends the previous literature by uncovering factors that promote infants' learning about novel objects within observed social interactions. By systematically varying the gaze direction of two adults towards one another and towards a visible object, the findings from Study 3 demonstrate that observed joint attention provides a supportive context for 9-month-old infants' learning about novel objects. Third, this thesis contributes to our understanding of the ontogenetic origins of infants' attention to and learning from others' social interactions. The findings from Study 3 demonstrate that the ability to learn about novel objects through observing third-party interactions starts to emerge in the first year after birth. Moreover, the results from all three studies demonstrate striking similarities between developments in the third- and first-party interaction domain. This raises the possibility that similar factors may underpin infants' attention and learning in the context of observed and self-experienced social interactions.

Given the vast intercultural variability in the degree to which infants are involved in direct face-to-face interactions in their daily lives, the overall findings of this dissertation align with the idea that infants are equipped with a highly efficient toolkit of motivations and abilities, enabling early cultural learning in different environmental contexts.

Zusammenfassung

Menschliche Säuglinge verlassen sich auf soziale Interaktionen um kulturell relevantes Wissen über ihre Umgebung zu erwerben (Csibra & Gergely, 2011; Reid & Striano, 2007; Tomasello, 2016). In ihrer täglichen Umgebung begegnen Säuglinge sozialen Interaktionen auf verschiedenen Wegen. Neben der eigenen Beteiligung („Teilnehmerperspektive") beobachten sie soziale Interaktionen zwischen anderen Personen („Beobachterperspektive"). Trotz der Abwesenheit jeglicher Involviertheit stellt das reine Beobachten sozialer Interaktionen eine wesentliche Quelle sozialer Lernmöglichkeiten dar (Paradise & Rogoff, 2009).

Während der vergangenen zehn Jahre wurde in einer zunehmenden Anzahl an Studien erforscht, wie Säuglinge soziale Interaktionen zwischen Dritten repräsentieren und wahrnehmen (z. B. Beier & Spelke, 2012; Papeo, 2020) und wie sie durch die Beobachtung von sozialen Interaktionen über ihre Umgebung lernen können (z. B. Matheson et al., 2013). Das wachsende Interesse an der Beobachterperspektive ging mit einem zunehmenden Bewusstsein dafür einher, dass das Ausmaß, in dem Säuglinge direkten Lehrpraktiken begegnen (und somit das Ausmaß in dem sie auf Beobachtungslernen angewiesen sind) erheblich zwischen Individuen und kulturellen Kontexten variiert (Gaskins & Paradise, 2010). Um die Vielfalt früher sozialer Lernstrategien stärker zu berücksichtigen, haben Wissenschaftlerinnen und Wissenschaftler die Wichtigkeit betont, beobachtungsbasierte Lernformen zusätzlich zu partizipativen Formen des Lernens zu erforschen (Rogoff et al., 2007).

Um einen Beitrag zu dieser Forschung zu leisten, war es das übergeordnete Ziel dieser Dissertation ein tieferes Verständnis für die grundlegenden Prozesse und Mechanismen zu erlangen, welche es Säuglingen ermöglichen aus der Beobachtung von sozialen Interaktionen zu lernen. Um dies zu erreichen, habe ich in dieser Arbeit einen breiteren Fokus auf die Lernsituation gelegt, der sowohl Gedächtnisprozesse während der Beobachtungssituation selbst betrachtet (Studie 3), als auch Entwicklungen in der sozialen Aufmerksamkeit und Motivation, die es Säuglingen ermöglichen soziale Interaktionen zwischen anderen Menschen zu erkennen und sich ihnen anzunähern (Studie 1 & 2).

Studien und Ergebnisse

Der empirische Teil dieser Dissertation umfasst drei präregistrierte experimentelle Studien, die folgende Lücken in der bisherigen Literatur adressieren. Erstens fehlte es an Studien, die im Kontext beobachteter Interaktionen den Einfluss intrinsischer Motivationen des Säuglings auf dessen Aufmerksamkeit und Lernen berücksichtigen. Im Rahmen dieser Dissertation betrachte ich endogene Motivationen, wie beispielsweise soziale Motivation, als eine Art „Türöffner" zu potentiellen sozialen Lerngelegenheiten, da sie die Aufmerksamkeit des Säuglings in Richtung von möglichen sozialen Interaktionen lenken. Eine zweite Limitation in der bisherigen Forschung war das Fehlen von Studien, die das Lernen aus der Beobachtung sozialer Interaktionen im ersten Lebensjahr untersuchen. Bisher wurde objektbezogenes Lernen im Kontext von beobachteten Interaktionen frühestens im zweiten Lebensjahr nachgewiesen. Drittens blieb unbekannt, welche Faktoren innerhalb einer beobachteten Interaktion das Lernen objektbezogener Informationen im Säuglingsalter fördern.

Alle drei Studien in dieser Dissertation basieren auf Maßen der offenen visuellen Aufmerksamkeit. Zur Messung der visuellen Aufmerksamkeit verwendeten wir bildschirmbasierte Eye-Tracking-Technologien (Studien 1–3), sowie die manuelle Kodierung des Blickverhaltens der Säuglinge während einer naturalistischen sozialen Interaktion (Studie 1). Die Studienpopulation bestand aus sich typisch entwickelnden Säuglingen zwischen 7 und 14.5 Monaten, die in einer westlichen, urbanen, industrialisierten Umgebung aufwuchsen.

Study 1: Wie entwickelt sich in der zweiten Hälfte des ersten Lebensjahres das Orientieren von Aufmerksamkeit in Richtung beobachtbarer sozialer Interaktionen? Wie verhält sich jene Entwicklung in Relation zu Veränderungen im aktiven sozialen Aufmerksamkeitsverhalten während selbst erlebter Interaktionen?

Frühere Studien haben gezeigt, dass 14 Monate alte Säuglinge ihre visuelle Aufmerksamkeit bevorzugt auf Situationen lenken, in denen zwei Personen einander zugewandt miteinander interagieren (in Kontrast zu nicht interaktiven Szenen in denen zwei Personen dieselben Bewegungen Rücken an Rücken ausführen, Galazka et al., 2014). Die ontogenetischen Ursprünge dieser Aufmerksamkeitspräferenz blieben jedoch unklar. Um diese Lücke zu schließen, war es das erste Ziel von Studie 1 zu untersuchen, wie sich das Orientieren von Aufmerksamkeit in Richtung sozialer Interaktionen zwischen 7 und 13 Monaten entwickelt—eine Periode, in der das soziale Aufmerksamkeitsverhalten von Säuglingen in selbst erlebten Interaktionen bedeutsame Entwicklungen durchläuft (Carpenter, Nagell, et al., 1998; Clearfield et al., 2008; Striano & Rochat, 1999). Das zweite Ziel von Studie 1 war es, die Aufmerksamkeitspräferenz für Interaktionen Anderer mit der sozialen Aufmerksamkeit in selbst erlebten Interaktionen zu vergleichen.

Studie 1 umfasste zwei Experimente, die sich hinsichtlich des getesteten Altersbereichs unterschieden. In Experiment 1 wurden zwei Altersgruppen von Säuglingen verglichen: eine

Gruppe bestand aus Säuglingen zwischen 7 und 8.5 Monaten ($N = 20$) und die andere Gruppe bestand aus Säuglingen zwischen 9.5 und 11 Monaten ($N = 20$). In Experiment 2 wurden Säuglinge aus einem breiteren und kontinuierlichen Altersbereich zwischen 7 und 13.5 Monaten getestet ($N = 50$). Alle teilnehmenden Familien kamen aus Leipzig (Deutschland). Um die Aufmerksamkeitspräferenz für soziale Interaktionen Dritter zu erfassen, maßen wir die Blickzeiten der Säuglinge in einem Blickzeit-Präferenz-Test („preferential-looking"), in dem zwei Arten von Videos gleichzeitig präsentiert wurden. Eines der Videos zeigte eine dyadische Interaktion zwischen zwei einander zugewandten Erwachsenen (ein Klatschspiel, Berühren der Hände, aneinander lehnen der Köpfe). Das andere Video zeigte eine gespiegelte, nicht interaktive Szene, in der dieselben Akteure dieselben Bewegungen Rücken an Rücken stehend ausführten. Um das soziale Aufmerksamkeitsverhalten der Säuglinge während selbst erlebter Interaktion zu erfassen, zeichneten wir Videoaufnahmen ihres natürlichen Interaktionsverhaltens während einer freien Spiel Situation mit ihrer primären Bezugsperson auf. Basierend auf diesen Aufnahmen kodierten wir vier Arten von Blicken der Säuglinge in Richtung der Bezugsperson: allgemeine Blicke in Richtung der Person, Blicke in das Gesicht der Person, Blickkontakt und geteilte Aufmerksamkeit auf ein Objekt.

Unsere Ergebnisse in Experiment 1 ergaben, dass im Vergleich zu Säuglingen zwischen 7 und 8.5 Monaten, Säuglinge zwischen 9.5 und 11 Monaten eine stärkere Blickzeitpräferenz für Videos zeigten, in denen zwei Personen einander zugewandt miteinander interagierten. Außerdem zeigten Säuglinge in der älteren Altersgruppe während des freien Spiels mehr soziale Blickverhaltensweisen (insbesondere geteilte Aufmerksamkeit) als jüngere Säuglinge (siehe auch Carpenter, Nagell, et al., 1998; Striano & Rochat, 1999). In Experiment 2 replizierten wir unsere Ergebnisse aus Experiment 1 und erweiterten diese, indem wir zeigten, dass der Anstieg beider Maße von 7 bis 13 Monaten einem kontinuierlichen Entwicklungsverlauf folgt. In einer zusammengeführten Stichprobe beider Experimente ($N = 90$) fanden wir einen positiven Zusammenhang zwischen der Aufmerksamkeitspräferenz für soziale Interaktionen Dritter und der sozialen Aufmerksamkeit während selbst erlebter sozialer Interaktion. Diese Korrelation wurde jedoch hauptsächlich durch das Alter der Säuglinge bestimmt.

Insgesamt zeigen die Ergebnisse aus Studie 1, dass das Orientieren von Aufmerksamkeit in Richtung sozialer Interaktionen Dritter Ähnlichkeiten zu Entwicklungen der sozialen Aufmerksamkeit in selbst erlebten Interaktionen aufweist. Dies deutet darauf hin, dass gemeinsam zugrundeliegende motivationale Mechanismen die Aufmerksamkeit von Säuglingen sowohl in Richtung direkter Interaktionspartner, als auch in Richtung beobachtbarer sozialer Interaktionen lenken.

Study 2: Ist es für Säuglinge intrinsisch belohnend soziale Interaktionen zu beobachten?

Eine weitere Frage die in der bisherigen Literatur offenblieb, ist welche unmittelbaren Mechanismen die Aufmerksamkeit von Säuglingen in Richtung sozialer Interaktionen lenken. Im Kontext selbst erlebter sozialer Interaktionen wurde in früheren Studien festgestellt, dass Säuglinge gegen Ende des ersten Lebensjahres vermehrt Anzeichen dafür zeigen, dass sie Möglichkeiten für soziale Interaktion aufsuchen und mögen. Dies deutet darauf hin, dass sie es intrinsisch belohnend finden mit anderen Menschen zu interagieren (Striano & Bertin, 2005; Venezia et al., 2004). Auf dieser Annahme aufbauend war es das Ziel von Studie 2 zu untersuchen, ob ähnliche Belohnungsmechanismen die Aufmerksamkeit von Säuglingen in Richtung beobachtbarer Interaktionen lenken. Wir nutzten Verstärkungslernen als einen Indikator für den intrinsischen Belohnungswert sozialer Interaktionen (Berridge & Robinson, 2003).

Insgesamt nahmen $N = 32$ Säuglinge im Alter zwischen 13 und 14.5 Monaten aus Uppsala (Schweden) an der Studie teil. In einem Test zu assoziativem visuellem Lernen sahen die teilnehmenden Säuglinge zwei nicht soziale Hinweisreize („Cues") wiederholt gepaart mit zwei Arten von Zielreizen („Zielvideos"): eine dyadische Interaktion oder eine nicht-interaktive Szene. Die Zielvideos basierten auf den für Studie 1 erstellten Video Stimuli. Jeder Säugling sah 12 Trials pro Bedingung, demnach 24 Trials insgesamt. Zu Beginn eines jeden Trials erschien einer der beiden Cues (Kreis oder Dreick) in der Mitte des Bildschirms. Wenn der Blick des Säuglings auf den Cue fiel, verschwand dieser und die mit dem Cue assoziierten Zielvideos erschienen auf der rechten oder linken Seite der zentralen Cue-Position. Um die assoziative Lernleistung der Säuglinge zu beurteilen, maßen wir die sakkadische Latenz und prädiktive Initialisierung ihrer Blickbewegungen zur korrekten Zielregion über mehrere Trials hinweg. Basierend auf den Prinzipien des Belohnungslernens (Skinner, 1938) gingen wir von folgender Annahme aus: Wenn Säuglinge es belohnend finden soziale Interaktionen zu beobachten, dann sollten sie die Assoziation zwischen einem willkürlichen geometrischen Cue und einem Zielvideo effektiver lernen, wenn das Zielvideo eine soziale Interaktion zeigt, als wenn es eine nicht interaktiven Szene zeigt (siehe auch Tummeltshammer et al., 2019).

Wir fanden heraus, dass Säuglinge schnellere sakkadische Latenzen und mehr prädiktive Blickbewegungen ausführten, wenn die Zielvideos eine soziale Interaktion zeigten als wenn sie eine nicht interaktive Szene zeigten. Dies deutet darauf hin, dass Säuglinge Assoziationen besser lernen, die ihnen ermöglichen dyadische Interaktionsszenen zwischen zwei Personen zu beobachten. Darüber hinaus indiziert das Ergebnis, dass beobachtete Interaktionen ohne jegliche zusätzliche Verstärkung assoziatives Lernen selektiv verstärken. Dies wiederum impliziert, dass beobachtete Interaktionen als primäre Verstärker für 13 Monate alte Säuglinge dienen können.

Study 3: Können Säuglinge im ersten Lebensjahr aus der Beobachtung sozialer Interaktionen über neue Objekte lernen? Wenn ja, tragen ähnliche Prozesse zu diesem Beobachtungslernen bei, wie zu referentiellem Lernen während selbst erlebter Interaktionen?

In der bisherigen Forschung wurde objektbezogenes Lernen im Kontext beobachteter sozialer Interaktionen frühestens im Alter von 18 Monaten nachgewiesen (Akhtar, 2005; Matheson et al., 2013). Es blieb daher unklar, ob bereits präverbale Säuglinge durch die bloße Beobachtung sozialer Interaktionen über neue Objekte lernen können und, falls ja, ob ähnliche Faktoren zu diesem Beobachtungslernen beitragen wie zu referentiellem Lernen während selbst erlebter sozialer Interaktionen (d. h. ein Zusammenspiel aus Blickkontakt und objektbezogenem Blick). Um dieser Frage nachzugehen, verwendeten wir ein auf Blickzeiten basierendes Paradigma, welches es ermöglicht die Enkodierung von visuellen Objekteigenschaften zu messen („Objektenkodierung"). In zwei Experimenten verglichen wir die Objektenkodierungsleistung von 9 Monate alten Säuglingen in einem Beobachtungskontext (Experiment 1) mit einem selbst erlebten Kontext (Experiment 2).

In Experiment 1 wurden $N = 32$ Säuglinge zwischen 9 und 10 Monaten aus Leipzig (Deutschland) mit vier Arten von Videos konfrontiert. In jedem der Videos waren jeweils zwei Erwachsene und ein Objekt zu sehen. Basierend auf einem 2 × 2 Design variierten die Videos systematisch hinsichtlich des Blickkontakts zwischen den beiden Akteuren (Blickkontakt oder kein Blickkontakt) und des objektbezogenen Blicks der Akteure (Blick in Richtung des Objekts oder vom Objekt weg). In Experiment 2 wurde eine zusätzliche Stichprobe von 9 bis 10 Monate alten Säuglingen ($N = 32$) in einem angepassten Studiendesign getestet, bei dem sie selbst aus Teilnehmerperspektive von einem Erwachsenen auf dem Bildschirm adressiert wurden. Als Indikator für erfolgreiche Objektenkodierung verwendeten wir die Präferenz für neue Informationen („novelty preference"), ein definierendes Merkmal des visuellen Gedächtnisses von Säuglingen. Hierzu zeigten wir in einer Blickzeit-Präferenz-Phase („preferential-looking") nach jedem Video dasselbe Objekt welches die Säuglinge soeben gesehen hatten neben einem neuen Objekt. Wir nahmen an, dass längere Blickzeiten auf das neue Objekt umgekehrt auf eine erhöhte vorherige Enkodierung des bekannten Objekts hinweisen würden (Rose et al., 1982).

In Experiment 1 zeigten Säuglinge eine Blickzeitpräferenz für neue Objekte, allerdings nur wenn sie zwei Erwachsene beobachtet hatten die nach vorherigem Blickkontakt gemeinsam auf ein Objekt schauten. Darüber hinaus konnten wir in Experiment 2 den Befund aus vorherigen Studien replizieren, dass die Objektenkodierung von Säuglingen auch in aktiv erlebten Interaktionen von dem Zusammenspiel zwischen Blickkontakt und objektbezogenem Blick abhängt. Wie in Experiment 1 zeigten Säuglinge nur dann verstärkte Objektenkodierung, wenn ein sozialer Partner Blickkontakt mit ihnen aufnahm bevor diese Person ihren Blick auf ein sichtbares Objekt richtete (siehe auch Cleveland & Striano, 2007; Okumura et al., 2020).

Insgesamt zeigen die Ergebnisse aus Studie 3, dass Säuglinge bereits im Alter von 9 Monaten durch die bloße Beobachtung triadischer „Person-Person-Objekt" Interaktionen über die visuellen Eigenschaften unbekannter Objekte lernen können. Darüber hinaus weisen unsere Befunde darauf hin, dass ähnliche Faktoren zu frühem Beobachtungslernen beizutragen scheinen wie zu referenziellem Lernen in selbst erlebten triadischen Interaktionen im selben Alter.

Diskussion

Zusammenfassend leisten die empirischen Befunde dieser Dissertation drei wesentliche Beiträge zu unserem Verständnis bezüglich der Aufmerksamkeit und des Lernens von Säuglingen im Kontext beobachteter sozialer Interaktionen. Erstens berücksichtigen die Studien intrinsische Motivationen als treibende Kräfte, die Säuglinge zu sozialen Interaktionen führen. In Übereinstimmung mit Theorien zu sozialer Motivation (Berridge, Robinson, & Aldridge, 2009; Chevallier et al., 2012) erbringen die Ergebnisse der präsentierten Studien einen Nachweis für zwei psychologische Komponenten von sozialer Motivation: soziales Orientieren („social orienting", Studie 1) und soziale Belohnung („social reward", Studie 2). Der zweite Beitrag der hier dargestellten Forschung ist das Aufdecken von Faktoren, die das Lernen während der Beobachtungssituation selbst fördern. Durch das systematische Variieren der Blickrichtung zweier Erwachsener zueinander und in Richtung eines sichtbaren Objekts zeigen die Ergebnisse aus Studie 3, dass das Beobachten von geteilter Aufmerksamkeit einen unterstützenden Kontext für objektbezogenes Lernen bei 9 Monate alten Säuglingen darstellt. Drittens trägt die diese Dissertation zu unserem Verständnis der ontogenetischen Ursprünge der Aufmerksamkeit und des Lernens im Kontext von beobachteten Interaktionen bei. Die Befunde aus Studie 3 zeigen, dass die Fähigkeit durch die Beobachtung sozialer Interaktionen über neue Objekte zu lernen sich im ersten Jahr nach der Geburt entwickelt. Darüber hinaus weisen die Ergebnisse aller drei Studien auffällige Ähnlichkeiten in der Entwicklung von Teilnehmer- und Beobachterperspektive auf. Dies wirft die Möglichkeit auf, dass ähnliche Faktoren der Aufmerksamkeit und dem Lernen in beobachteten und selbst erlebten sozialen Interaktionen zugrunde liegen könnten.

In Anbetracht der interkulturellen Variabilität bezüglich des Ausmaßes, in dem Säuglinge in ihrem Alltag in direkten Interaktionen involviert sind, sind die Gesamtbefunde dieser Dissertation in Einklang mit der Idee, dass Säuglinge mit hocheffizienten Motivationen und Fähigkeiten ausgestattet sind, die frühes kulturelles Lernen in verschiedenen Kontexten ermöglichen.

Contents

1

General Introduction

Human infants face the immense challenge of learning a large amount of culturally relevant knowledge. How do they manage this fundamental task in an environment that constantly confronts them with vast amounts of sensory impressions and novel information? How can they recognize in which content it is worth investing their limited attention and memory resources?

One answer to these questions is that infants strongly rely on social interactions to navigate their attention toward culturally relevant content and filter it from less relevant information in their environment (Csibra & Gergely, 2006; Reid & Striano, 2007; Tomasello, 2001). Social interactions provide a rich source of social cues and, as such, a particularly meaningful context for guiding infants' attention and learning. In their daily lives, infants encounter social interactions in multiple ways. Aside from direct participation ("first-party perspective"), they are surrounded by social interactions between others, such as conversations between their caregivers or playful interactions between their siblings ("third-party perspective"). Despite the absence of own involvement, the mere observation of others' social interactions represents an essential source of social learning opportunities (Paradise & Rogoff, 2009).

In the last decade, an increasing number of studies have investigated how infants represent social interactions between third parties (e.g., Gredebäck & Melinder, 2010; Papeo, 2020) and how they come to use third-party interactions to learn about novel objects (e.g., Shneidman & Woodward, 2016). The growing interest in the third-party perspective has accompanied a generally increasing awareness that the extent to which children encounter direct teaching practices and, thus, rely on observational learning, varies significantly across cultural contexts (Gaskins & Paradise, 2010). Along with this finding, intentional observation has been increasingly recognized as a universal strategy enabling early cultural learning in different environmental contexts (Rogoff et al., 2007; Rogoff, Paradise, Arauz, Correa-Chávez, & Angelillo, 2003). To better account for the multiple facets of early social learning strategies, researchers have emphasized the importance of studying observational forms of learning in addition to participatory forms (Paradise & Rogoff, 2009).

In contributing to this research, the overarching aim of this thesis is to deepen our understanding of the foundations of infants' observational learning from third-party social interactions. This will be achieved by taking a broader perspective, which not only focuses on infants' attention and memory during the actual learning situation but also considers their active role in gaining access to potential learning opportunities: First, I will investigate social attentional developments and underlying motivational drivers enabling infants to detect and approach third-party interactions (Study 1 & 2). Second, I will examine infants' object encoding in the context of third-party interactions (Study 3).

The structure of this thesis is as follows: In the remainder of chapter 1, I will give a general introduction to the topic, provide an overview about the current state of research, and point out three gaps that are addressed in the empirical part of this thesis. The empirical part comprises three experimental studies that are outlined in chapters 2 to 4. The findings from all three studies will be integrated in the final chapter, the general discussion (chapter 5). Here, I will discuss the implications and limitations of the empirical work in this thesis and provide an outlook on possible future directions. Since all studies in this thesis have been inspired by phenomena in the active interaction domain, I will compare infants' observational perspective on third-party interactions with their first-party perspective in direct interactions throughout this thesis. Unless otherwise specified, I will refer to typically developing children. The primary period of interest is the infancy period (i.e., the first postnatal year), but I will also refer to developments in early toddlerhood since most previous studies on observational learning have been conducted in the second year of life.

1.1   The Infant As an Active Learner

Until around 50 years ago, preverbal infants were seen as predominantly passive receivers, exposed to the mercy of their environment to receive stimulation until possessing the motor and language abilities to produce overt interaction behaviors (for a review, see Raz & Saxe, 2020). Since then, the development of implicit technologies such as eye tracking, Electroencephalography (EEG), and functional near-infrared spectroscopy (fNIRS) has provided researchers with continuously improving tools to study the processes and correlates of early learning already shortly after birth with increasing precision. Based on the current state of this research, the picture of the infant as a passive and primary incidental learner has become obsolete. Instead, the human infant turned out to be a capable and intrinsically motivated learner who takes an active role in scaffolding their learning experience from early in ontogeny (Begus & Southgate, 2018).

Even before possessing the behavioral repertoire to interact with their environment via communicative gestures, reaching behaviors, or speech production, infants can contribute to their learning experience by controlling their visual attention. Throughout the first postnatal year, infants' initially predominantly reflexive visual attention becomes increasingly influenced by

endogenous factors (see section 1.2.1), making observational learning an increasingly active and voluntary process. In line with the focus in this thesis, infants can regulate their visual learning input in at least two ways: First, they can select what and from whom they learn during an ongoing observation. Second, they can modulate the availability of learning opportunities by selectively shifting their attention toward it. Given the importance of social interactions for early cultural learning, selective attention to *social* information (i.e., social attention) is particularly beneficial for young learners since it navigates them to situations where they can potentially engage in or observe others' social interactions.

1.2 Social Attentional Requirements of Infants' Learning From Social Interactions

What developments in social attention are required for infants to learn from social interactions? Before addressing this question from a first- and third-party perspective, I will first clarify the meaning of the terms "social attention" and "social interaction" for the context of this thesis.

1.2.1 What Defines Social Attention?

As summarized in an article by Salley and Colombo (2016), social attention has been conceptualized in various ways in the previous literature: as basic visual attention (e.g., Simion, Di Giorgio, Leo, & Bardi, 2011), as social motivation (e.g., Chevallier et al., 2012), as social behavior (e.g., Mundy & Newell, 2007), or as social cognition (e.g., Carpenter & Tomasello, 1995). In this thesis, I refer to an inclusive definition by Bertenthal and Boyer (2015), which accounts for the multiple facets of social attention by conceptualizing the construct as a dynamic process rather than a developmental outcome. More specifically, the authors define social attention as the attentional response of an observer to the presence of other people and to other peoples' actions. The developmental process begins immediately after birth (if not prenatally, Castiello et al., 2010; Hepper, Scott, & Shahidullah, 1993; Reid et al., 2017). Throughout the infancy period, it continues to gradually develop in association with developments in many other domains, including social cognition, social motivation, and perceptual skills. While infants rely on multiple perceptual channels to detect and respond to social information, the primary focus in this thesis is on *visual* social attention.

Selective and sustained attention to social information

In their daily learning environment, infants are confronted with plenty of information at once. To filter socially relevant stimuli from other competing information, they must pay selective attention to social information. As reviewed in an article by Corbetta and Shulman (2002), selective attention involves two interrelated processes: exogenous orienting (stimulus-driven attention) and

endogenous orienting (goal-directed attention). While exogenous attention is considered to be reflexive and automatic, endogenous attention depends on higher-level processes. Just as any other visual attention, infants' selective attention to social information is initially predominantly driven by exogenous processes. Endogenous social attention develops with age and experience and is significantly influenced by the infant's intrinsic motivations and emerging cognitive abilities. Another form of attention required for information processing is sustained attention. After selectively orienting their visual attention to a target, infants must maintain their attention to the stimulus to encode, store, and later recognize it (Reynolds, 2015). The duration of sustained attention periods increases significantly throughout the infancy period (Ruff & Capozzoli, 2003).

Covert and overt shifts of spatial attention

A further distinction in the attention literature and, thus, the social attention literature, is the distinction between overt and covert attention (Johnson, 1994). While overt shifts of visual attention include eye or head movements toward a target stimulus, covert shifts of attention represent attentional displacements toward targets in the visual periphery, occurring independently of what the infant looks at directly. Both forms of spatial attention can afford selective information processing (Carrasco & McElree, 2001) and can occur reflexively and intentionally (Hunt & Kingstone, 2003). The relevance of overt and covert attention for infants' referential learning will be discussed in more detail in Study 3 (chapter 4, section 4.4) and in the general discussion (chapter 5, section 5.2.2).

Focusing on behavioral measures of visual attention

In previous research on social attention, infants' response to social stimuli has been studied at the behavioral level and at the neural level (for an overview, see Puce, Latinus, & Rossi, 2015). To study the neural correlates and cortical systems involved in social information processing, previous studies have used electrophysiological and neuroimaging techniques such as EEG, fNIRS, functional magnetic resonance imaging (fMRI), or magnetoencephalography (MEG). Since the empirical work of this thesis relies on behavioral measures of visual attention, I will mainly focus on developments in infants' overt social attention behavior and less on the neural processes and brain areas involved.

1.2.2 What Defines a Social Interaction?

A social interaction can be generally defined as the contingent exchange of attention and actions between two or more autonomous agents (De Jaegher et al., 2010; Isik et al., 2020). The range of scenarios labeled as "social interaction" in previous infant studies is enormous and strongly depends on the research question being asked. For example, previous social interaction scenarios

have varied regarding: the naturalness of the interaction (e.g., unscripted and naturalistic or highly controlled), the degree of virtuality (e.g., on screen or live), the kinds of interaction partners (e.g., humans, puppets, or animated shape agents), the number of social partners (e.g., two partners in a dyadic interaction or multiple partners in a group setting), the symmetry between partners (e.g., adult-infant or peer-to-peer), the familiarity of partners (e.g., relatives or strangers), the content of the interaction (e.g., verbal or non-verbal, including touch or not, including a mutual action or not), the amount of motion involved (e.g., still images or dynamic motion), the synchronicity of the agents' movements (e.g., synchronous or asynchronous movements), the moral valence (e.g., prosocial, antisocial, or neutral), the emotional value (e.g., emotions or no emotions expressed), or whether external content was integrated within the interpersonal interaction (e.g., including an object or not).

The focus in this thesis will be on neutral social interactions between two human agents. More specifically, I will *not* focus on interactions with moral or emotional valence, group settings, or geometrical shape agents. Within the domain of interpersonal social interactions, I will differentiate between dyadic and triadic interactions. While dyadic interactions represent a merely interpersonal interaction between two people, triadic interactions include external content such as objects or events in addition to the interpersonal dyad (Siposova & Carpenter, 2019). The degree to which an object or event is integrated within an interpersonal interaction varies between different kinds of triadic interactions, such as joint attention or parallel attention interactions. I will come back to this in a later section in this chapter, when I describe the developmental progression from dyadic to triadic interactions (section 1.2.3).

First-party perspective versus third-party perspective on social interactions

Throughout this thesis, the term "first-party interaction" will refer to situations in which the infant is directly involved as an active interaction partner, and the term "third-party interaction" to interactions that the infant encounters from an observer perspective, without own participation. In the empirical studies (chapters 2–4), the interactions in the third-party conditions consist of pre-recorded dynamic interactions between two adult women in a highly controlled setting, including reduced affective stimulation, an exact timing of actions, and no perceptual distraction. In the first-party settings, the interaction either consists of a naturalistic free play situation between infant and caregiver (Study 1) or the interaction partner is portrayed by an adult on screen who acts non-contingently to the infants' response according to a pre-defined script (Study 3). In the general discussion (chapter 5), I will put the study findings in perspective with interpretive restrictions emerging from this operationalization.

The following two sections address developments in social attention that are foundational for infants to learn about their environment through first- and third-party social interactions. Figure 1.1 in the end of this section summarizes the key developments in both perspectives.

1.2.3 Infants' Social Attention in First-Party Social Interactions

According to the *Directed Attention Model of Infant Social Cognition* (Reid & Striano, 2007), five processing stages are required for the infant to learn from a direct social partner about novel objects or other external content in their environment: The infant must detect a socially relevant organism, identify this organism as a social partner, detect the partner's focus of attention, detect the partner's object-directed attention, and then infer the partner's goals and/or prepare a response. In the following, I will outline three domains of social attention that are foundational for infants to learn through direct interaction with a social partner. The domains cover the perceptual processing stages suggested by Reid and Striano (2007) and extend them with abilities required in more dynamic and coordinated interactive settings.

Detecting and identifying another individual as a social interaction partner

Human infants enter the world equipped with attentional biases enabling them to detect other humans (for a review, see Simion et al., 2011). Immediately after birth, newborns show an attentional preference for biological motion over random motion patterns (Simion, Regolin, & Bulf, 2008), face-like over non-face patterns (Goren, Sarty, & Wu, 1975), and human over non-human primate faces (Heron-Delaney, Wirth, & Pascalis, 2011). In addition, newborns show a face-specific inversion effect, which is characterized by longer reaction times and decreased precision toward upside-down compared to upright faces but not objects (Leo & Simion, 2009). This suggests that faces, different from objects, are processed holistically rather than in its separate components. The face-specific inversion effect indicates that human infants are born with a perceptual template for faces, enabling them to quickly recognize socially relevant organisms in their environment (for reviews, see Hoehl & Peykarjou, 2012; Pascalis et al., 2011). By 3 months of age, infants show an inversion effect for human bodies as well, indicating that they use body information in addition to faces to detect other individuals (Gliga & Dehaene-Lambertz, 2005; Reed, Stone, Bozova, & Tanaka, 2003). This is further supported by the finding that 3-month-old infants preferentially attend to human over non-human primate individuals when presented with face or body information in isolation, as well as when both information are presented simultaneously (Heron-Delaney et al., 2011). Together, these findings show that infants are, from early on, prepared to selectively and quickly recognize other humans.

Throughout the first year of life, infants' social perception develops further as their practical experience, their visual system, their social brain functions, and their understanding of others develop (Colombo, 2001; Grossmann & Johnson, 2007; Hunnius & Geuze, 2004). For example, the face recognition system increasingly narrows to human-specific features from 6 to 9 months of age—a phenomenon described as "perceptual narrowing" in the social attention literature (for reviews, see Anzures et al., 2013; Lee, Quinn, & Pascalis, 2017). While 6-month-old infants can

distinguish individuals from their own species (i.e., human faces) as well as from other species (e.g., monkey faces), 9-month-olds' discrimination is limited to individuals of their own species (Heron-Delaney et al., 2011; Pascalis, Haan, & Nelson, 2002; Simpson, Varga, Frick, & Fragaszy, 2011). A similar narrowing effect occurs for human faces from different races (Kelly et al., 2007, 2005), indicating that infants' perceptual specialization depends on their selective experience with faces they encounter most frequently in their social environment.

As also proposed in the model by Reid and Striano (2007), the detection and recognition of another individual is foundational for social interaction. However, not every identified person necessarily becomes a social interaction partner. Another precondition for successful interaction is that both partners are willing to establish an interactive relationship with one another. From the respondent's perspective, this requires a sensitivity to whether another person's attention is directed to the self ("self-relevance", Grossmann, 2015; Reddy, 2003). According to the *Natural Pedagogy* account, infants are sensitive to others' attention from early on (Csibra & Gergely, 2009, 2011). More than that, the account assumes that human infants have evolved a species-unique sensitivity to detect and use ostensive signals. Communicative intent can be signaled through different perceptual channels, for example, via visual cues like direct gaze or auditory cues such as infant-directed speech. Infants are sensitive to ostensive signals in multiple modalities from birth. In the visual domain, newborns show an attentional preference for faces with opened over closed eyes (Batki, Baron-Cohen, Wheelwright, Connellan, & Ahluwalia, 2000) and direct over averted gaze (Farroni, Csibra, Simion, & Johnson, 2002). In the auditory domain, newborns prefer infant-directed versus adult-directed speech (Bergelson et al., 2017; Cooper & Aslin, 1990) and the voice of their own mother over a female stranger's voice (DeCasper & Fifer, 1980; Ockleford, Vince, Layton, & Reader, 1988). Within the broader framework of this thesis, infants' sensitivity to communicative signals drives their attention toward situations in which they may have the opportunity to engage in a social interaction.

Coordinating attention with a social partner in turn-taking dyadic interactions

Around 2 months of age, infants begin to coordinate their attention with others in dyadic face-to-face interactions (Aureli, Presaghi, & Garito, 2017; Cohn & Tronick, 1987; Rochat & Striano, 1999; Tomasello, Carpenter, Call, Behne, & Moll, 2005). A defining feature of early dyadic interactions is contingency between infants' own and others' actions. Contingent responsiveness of a social partner scaffolds early learning, as it helps infants recognize others' communicative intent (Csibra, 2010). Infants' sensitivity to the turn-taking dynamic of face-to-face interactions develops around one month after birth. If an interaction partner acts in a passive way, infants begin to behave more actively during an interaction and vice versa (Tomasello et al., 2005; Trevarthen & Aitken, 2001). Around 2 months of age, infants expect their interaction partner to produce socially contingent responses. During interactions with their mother, for example, 2-month-old infants show

decreased positive responses (e.g., smiling and gazing at their mother) and increased negative responses (e.g., frowning or averted gaze), when their mother engages in behaviors disrupting the contingency of the interaction (Legerstee & Varghese, 2001; Nadel, Carchon, Kervella, Marcelli, & Reserbat-Plantey, 1999; Rochat, Striano, & Blatt, 2002; Tronick, Als, Adamson, Wise, & Brazelton, 1978).

Between 2 and 6 months of age, infants' early sensitivity to the reciprocal dynamic and timing of face-to-face interactions develops further as they begin to exchange emotions with social partners. In so called proto-conversations, infant and adult focus their attention on one another while taking turns in facial, vocal, or postural expressions of basic emotions (Aureli et al., 2017; Bråten, 1988; Trevarthen, 2011). Even though the amount to which infants engage in face-to-face visual engagement shows considerable variation across cultural contexts (Mesman et al., 2018), turn-taking proto-conversations are thought to represent a universal feature of adult-infant interactions (Keller, Schölmerich, & Eibl-Eibesfeldt, 1988; Tomasello, 2009; Trevarthen, 1993).

Following others' attention and coordinating attention in triadic interactions

During the second half of the first year of life, infants begin to engage in triadic social interactions. In contrast to merely interpersonal interactions, triadic interactions incorporate external content such as objects within the interactive situation. While 3-month-old infants show early covert attentional cueing effects (Hood, Willen, & Driver, 1998), infants between 4 and 6 months begin to follow the attentional focus of others *overtly* toward visible objects (Astor & Gredebäck, 2019). To infer another person's visual focus of attention, infants make use of referential cues such as gaze and head movements (D`Entremont, Hains, & Muir, 1997; Gredebäck, Theuring, Hauf, & Kenward, 2008) or pointing gestures (Bertenthal, Boyer, & Harding, 2014; Schmitow, Kochukhova, & Nyström, 2016).

Especially in young infants, gaze following behavior is strongly influenced by communicative cues. Senju and Csibra (2008), for example, found that 6-month-old infants only followed another person's gaze toward a visible object, if this person had addressed the infant via direct gaze or infant-directed speech prior to shifting their gaze. The authors interpreted this finding as evidence showing that communicative cues are crucial for young infants to interpret an upcoming gaze shift as meaningful. Findings from other studies have suggested that the effect of ostensive cues on young infants' gaze following depends on the highly attention-grabbing nature of ostensive cues rather than the underlying communicative intent. For example, Szufnarowska, Rohlfing, Fawcett, and Gredebäck (2015) have shown that 6-month-old infants follow another person's gaze equally well after seeing this person perform a non-communicative attention-grabbing action, such as shivering their head while looking down. According to both perspectives, the ability to follow others' gaze represents a milestone in social development as it enables infants to use other people as social cues helping them to identify content worth learning in their

environment (for a review about the underlying mechanisms of gaze following, see Del Bianco, Falck-Ytter, Thorup, & Gredebäck, 2019).

Around 9 months of age, infants begin to engage in joint attention episodes during which they coordinate their attention with the attentional perspective of another person toward content outside the interpersonal interaction (for a review about the social attentional requirements for joint attention, see Hoehl & Striano, 2015). The emergence of joint attention skills is considered an important milestone for early cultural learning, because it makes communication about cultural artifacts possible (Tomasello, 2001). At the most basic level, coordinated joint attention includes the alternation of eye gaze between two interaction partners and an object of mutual interest (Bakeman & Adamson, 1984). In contrast to parallel attention to an object, truly joint attention includes the mutual awareness of both partners about the shared experience (Carpenter & Liebal, 2011; Siposova & Carpenter, 2019).

Joint attention behaviors are typically divided into two categories: responding to joint attention and initiating joint attention (e.g., Mundy & Newell, 2007). Responding to joint attention behaviors comprise attention following behaviors produced in response to another person's invitation to share attention. Initiating joint attention behaviors, in contrast, aim to direct a partner's attention toward a common target and comprise, for example, ostensive eye contact, vocalizations, or pointing gestures. Between 9 and 18 months of age, infants increasingly progress from responding to others' joint attention bids to initiating joint attention episodes themselves (Carpenter, Nagell, et al., 1998; Mundy et al., 2007).

1.2.4   Infants' Social Attention in Third-Party Interactions

Many of the above-outlined developments are also foundational for infants' referential learning from third-party interactions. Like in a first-party setting, infants need to detect social organisms, identify them as relevant partners, detect their focus of attention, and follow their attention to content in the environment. The essential difference to the first-party perspective is that, in a third-party setting, the infant is neither involved in any social relationship nor are they directly addressed at any point. Thus, to successfully learn through third-party observation, infants require additional abilities specific to observational settings. In the following subsections, I will focus on three domains: the ability to detect a mutual relationship between two human agents, the ability to understand the turn-taking dynamic within observed dyadic interactions, and the ability to infer others' communicative intent and collaborative goals within observed triadic interactions.

Detecting a social relationship between two human agents

A salient visual cue to infer a social relationship between two agents is their body orientation toward one another. Infants begin to differentiate between face-to-face and back-to-back settings

at around 6 months of age: When observing a conversation between two adults, 6- but not 4-month-old infants perform more gaze shifts between the interlocutors when seeing them facing each other, as when they see them standing back-to-back while talking (Augusti, Melinder, & Gredebäck, 2010). From 6 to 11 months of age, the enhanced gaze shift pattern between facing dyads increases further, suggesting a developmental progression throughout the second half of the first year of life (Augusti et al., 2010). At 9 months of age, infants show the enhanced gaze shift pattern between facing dyads even for silent and static images of isolated heads and regardless of whether the eyes of the agents are opened or closed, indicating that body orientation alone provides a strong cue to inform about a social relationship (Handl, Mahlberg, Norling, & Gredebäck, 2013). Infants' ability to discriminate between third-party mutual gaze and averted gaze has been demonstrated in other paradigms as well. In violation-of-expectation paradigms, for example, 10-month-old infants look longer at a screen when seeing a man and a woman turning their heads and gaze away from one another, after being habituated to a scene showing the same people facing each other and vice versa (Beier & Spelke, 2012).

Another branch of research has investigated the underlying perceptual mechanisms of infants' attentiveness to face-to-face interactions. This research has revealed that face-to-face relationships are extracted rapidly and automatically in the visual system. For example, a study by Papeo, Stein, and Soto-Faraco (2017) has provided evidence for a "two-body inversion effect". In a visual search task, adult participants recognized face-to-face dyads faster than back-to-back dyads, but only when seeing the stimuli in an upright orientation, not when seeing them upside down. Papeo and colleagues interpreted this finding as evidence supporting the assumption that face-to-face interactions are grouped in one attentional unit in the human visual system and are stored as a chunk in the working memory (Papeo, 2020; see also Ding, Gao, & Shen, 2017). A similar effect has been found by Vestner and colleagues (Vestner, Gray, & Cook, 2020, 2021). However, in contrast to Papeo and colleagues, the researchers claimed that the search advantage for face-to-face dyads does not reflect a search advantage for social interactions per se, but that it rather derives from the lower-level saliency of a "hotspot" region present between facing dyads, but not between back-to-back dyads. While future studies are needed to disentangle the specific underlying mechanisms, both perspectives share the common assumption of a processing advantage for face-to-face over back-to-back dyads in the adult visual system (see also Hafri & Firestone, 2021).

One previous study has tested the ontogenetic origins of this processing advantage in human infants. In a series of preferential-looking experiments, Papeo, Goupil, and Hochmann (2020) presented 6-month-old infants with two images showing different arrangements and numbers of human-like bodies. The authors used a data-driven approach to identify the time window during which infants were most attentive. Longer looking times at an image were interpreted as reversely indicating that the other, competing image had been processed faster. Following this assumption, Papeo and colleagues interpreted their findings as evidence showing

that 6-month-old infants process facing dyads more efficiently than back-to-back dyads, but only when seeing the stimuli in an upright orientation. In another experiment, the authors found that infants' discrimination based on the relative positioning of entities is selective to body dyads, and does not generalize to body-object pairs, as indicated by the absence of looking time differences between an image showing a person facing an object or turning their back toward an object. Moreover, face-to-face dyads but not back-to-back dyads seem to be treated with the same efficiency as single bodies, as indicated by the absence of a looking time difference between facing dyads and single bodies, but longer looking times at back-to-back dyads compared to single bodies. In sum, the findings by Papeo and colleagues suggest that already 6-month-old infants have a perceptual template for face-to-face dyads, helping them to detect and process social interactions with high priority and efficiency.

In focusing on the neural underpinnings of third-party social interaction processing, previous studies using MEG have revealed that the adult brain spontaneously represents the presence and type of a social interaction (mutual gaze vs. joint attention, Isik et al., 2020). Moreover, studies investigating *where* in the brain social interactions are represented, have suggested a brain network selective to social interaction. By using fMRI, the dorsomedial prefrontal cortex (dmPFC) and the posterior superior temporal sulcus (pSTS) have been identified as the neural substrate for representing and interpreting social interactions in the adult brain (Isik, Koldewyn, Beeler, & Kanwisher, 2017; Wagner, Kelley, Haxby, & Heatherton, 2016). In taking a developmental perspective, a recent study supports the hypothesis that the brain basis for processing social interactions begins to specialize early in human ontogeny (Farris, Kelsey, Krol, Thiele, Hepach, Haun, & Großmann, under review). By using fNIRS, 6- to 13-month-olds' brain responses were measured while infants saw videos showing third-party social interaction scenes, non-interactive back-to-back scenes, or upside-down versions of the social interaction scenes. The videos were based on the stimuli created for Study 1 and Study 2 in this thesis. Similar to previous findings from adult studies (Isik et al., 2017; Wagner et al., 2016), infants at all ages preferentially engaged brain regions localized within the dmPFC and pSTC when viewing social interactions, compared with both non-interactive and inverted control scenes.

In conclusion, the studies reviewed thus far suggest that the infant visual system is not only attuned to isolated social entities but also to face-to-face relationships between others. Within the broader framework of this thesis this suggests that, in the second half of the first year of life, infants can recognize social relationships from an observer perspective.

Understanding the turn-taking dynamic of observed dyadic interactions

In addition to recognizing social relationships between others, infants develop an increasing understanding of the turn-taking dynamic of third-party interactions. When observing a conversation between two people, infants do not only *follow* the turn-taking flow, but they also

predict the response of an addressee: 6- and 11-month-old perform predictive gaze shifts toward an addressed person before this person has started responding, and they do so more frequently in face-to-face compared to back-to-back settings (Bakker, Kochukhova, & von Hofsten, 2011; Augusti et al., 2010). This ability increases from 6 over 12 to 26 months of age, indicating that the reciprocal understanding of dyadic conversations increases throughout the first two years (Bakker et al., 2011; von Hofsten, Uhlig, Adell, & Kochukhova, 2009). In addition, infants' ability to predict turn-taking events in observed conversations becomes increasingly selective to social partners. Twelve-, but not 6-month-old infants make more predictive gaze shifts between talking humans than between toy trucks producing turn-taking motor sounds (Bakker et al., 2011). Moreover, when seeing an adult uttering a sentence or producing non-speech sounds toward a another adult in a face-to-face setting, 12- and 24-month-old infants look faster and longer at the recipient following speech than non-speech sounds, suggesting that they expect the recipient to selectively respond to speech (Thorgrimsson, Fawcett, & Liszkowski, 2015). This effect disappears when seeing the same adults turning their backs toward one another, suggesting that it is specific to face-to-face settings.

The assumption that infants can differentiate between animate and inanimate interaction partners in observed interaction contexts receives further support from previous studies using violation-of-expectation paradigms. In most of these studies, infants were first habituated to an event during which they saw an actor talking or gesturing while looking at an occluded area. Then, the occluded area was uncovered so that infants could see what was previously hidden. In most studies that was either an inanimate object or a person. Longer looking times to the uncovered target were interpreted as a surprise response, reversely indicating that infants had developed a different expectation based on their observations during habituation. Corresponding studies revealed that 6-month-old infants expect people to behave differently toward persons and inanimate objects, in that they expect a person to talk to another person but not to an object, and to reach for an object but not for another person (Molina, Van de Walle, Condry, & Spelke, 2004; Legerstee, Barna, & DiAdamo, 2000). Relatedly, 10- but not 9-month-old infants expect a talking person to look in the direction of a human interaction partner rather than an animated toy truck (Beier & Spelke, 2012, Experiment 3). Taken together, these findings demonstrate that, during the second half of the first year of life, infants develop a sensitivity to the turn-taking dynamic of third-party interactions, with an increasing specialization for social partners.

Inferring others' communicative intent and collaborative goals in observed triadic interactions

Around their first birthday, infants begin to develop an understanding of communicative signals and intentions within observed triadic interactions. Twelve-month-old infants, for example, anticipate a person to respond to another person's action only if this action serves a communicative goal. After watching a communicator grasping one of two target objects, infants

expect the responder to reach for the target object, but only if the communicator had produced a speech vocalization in their direction, not if she had produced a cough vocalization (Yamashiro & Vouloumanos, 2018). By 14 months of age, infants can furthermore recognize observed non-verbal communicative acts. When seeing a gesturer pointing to an object or using a palm-up request, 14-month-olds anticipate an addressee to give the object to the gesturer, suggesting that they ascribed a communicative motive to the gesture. In contrast, when the gesturer reaches for the object or no action takes place, infants do not anticipate the addressee to respond (Thorgrimsson, Fawcett, & Liszkowski, 2014; Elsner, Bakker, Rohlfing, & Gredebäck, 2014).

At least by 18 months of age, infants take ostensive context between two people into account when interpreting the informative value of a pointing gesture in a third-party context. After observing an adult gazing ostensively in the direction of another adult while pointing toward one of two hiding locations, 18-month-olds can successfully predict the location of a hidden toy when searching for the toy themselves (Gräfenhain, Behne, Carpenter, & Tomasello, 2009). In contrast, infants perform at chance level when observing the adult using non-ostensive similar behaviors. Relatedly, 18-month-old infants consider third-party ostensive context when assessing the relevance of others' gaze-object relations. When seeing a woman gazing at one of two toys after either facing a social partner (face-to-face condition) or looking away from them (back-to-back condition), 9- and 12-month-old infants follow the woman's gaze in both conditions, while 18-month-olds follow her gaze only if both partners have engaged in mutual eye contact beforehand (Meng, Uto, & Hashiya, 2017). Together, these findings show that infants are sensitive to third-party communicative signals in the first half of the second year of life.
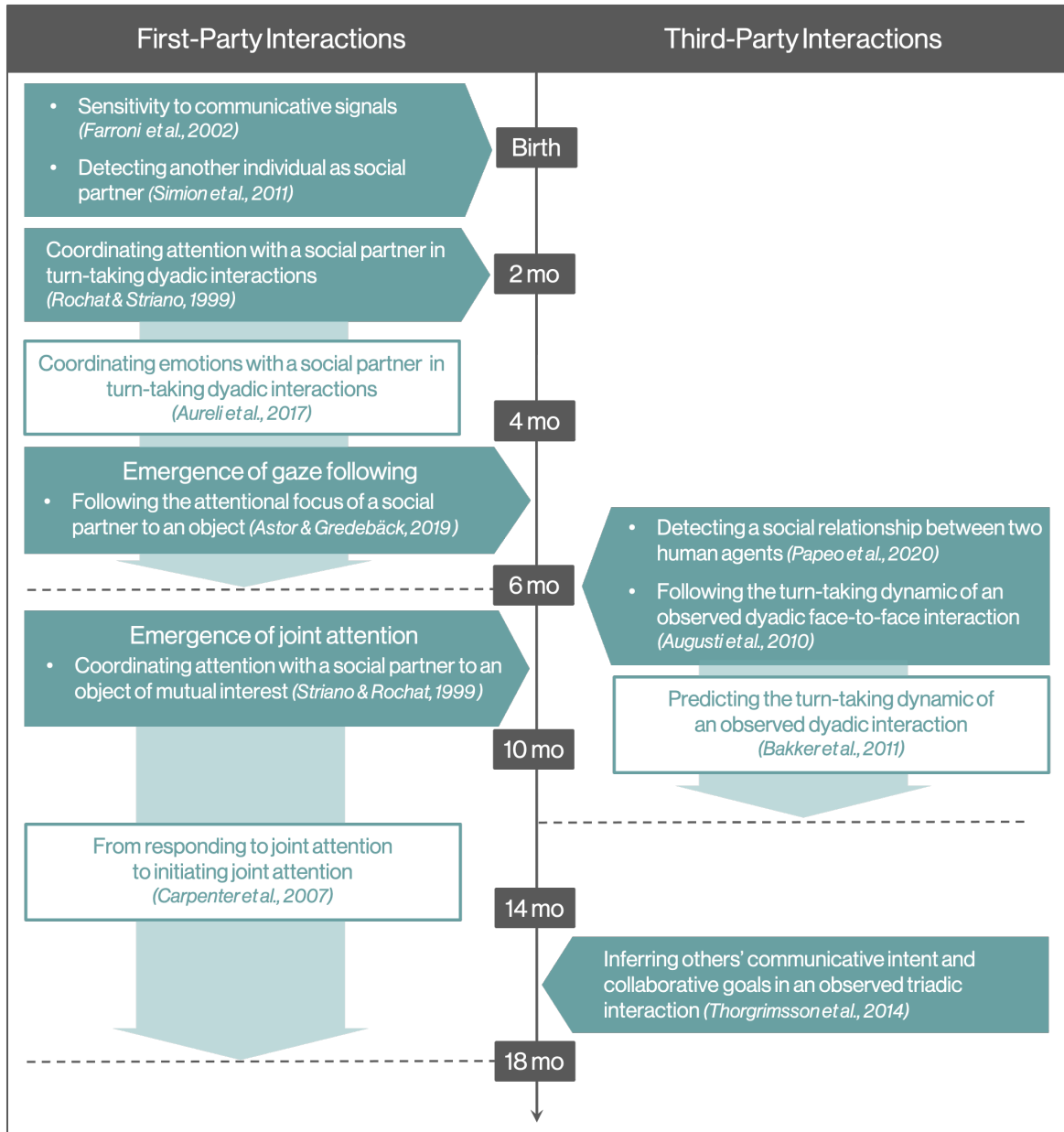
During the same period, infants develop the ability to infer others' collaborative action goals in an observed interaction. When seeing a collaborative setting during which two actors smile at each other, engage in joint attention, and pursue a common collaborative action goal, 14-month-old infants understand that the hierarchically structured actions of the two collaborative partners complement each other and that both actions are critical to accomplish the mutual goal. However, when the actions of the actors are not causally related, 14-month-olds do not show signs of interpreting the actions of the two individuals in terms of a collaborative goal (Henderson & Woodward, 2011). Furthermore, 14-month-old infants can use these inferences to make predictions regarding the future behavior of the single interaction partners: When seeing one of two actors again after observing them engaging in a collaborative interaction, infants expect this individual to reach toward an object that has been the shared goal of the previously observed collaborative interaction (Krogh-Jespersen, 2020). In contrast, infants do not systematically predict the future behavior of an actor, if this person was engaged in a non-collaborative interaction in which two actors looked at each other but pursued different action goals. Further indication for a developing understanding of collaborative action goals comes from a study showing that 18-month-olds interpret a subsequently observed action sequence as parts of a joint goal only after observing two

actors engaging in a face-to-face conversation (Fawcett & Gredebäck, 2013). When the actors previously engaged in a non-interactive activity during which they turned away from one another, infants interpreted their actions as individual activities. This supports the idea that others' mutual social engagement in a face-to-face conversation leads 18-month-old infants to bind the individual actions of two people together as part of a larger joint goal.

In conclusion, the current state of research indicates that infants develop foundational skills for observational learning from third-party interactions during the first 1.5 years after birth: From 6 months onwards, they conceive others as social partners, recognize the presence of a relationship between two human agents, pay selective attention to third-party social interactions, and begin to anticipate the turn-taking sequences of observed reciprocal exchanges. Throughout the second half of the first year of life, these abilities develop further. By their first birthday, infants can flexibly represent different kinds of social relations between agents in different action domains and they predict others' action goals. Slightly later, between 14 and 18 months of age, they begin to infer others' communicative intent and collaborative action goals within observed interactions. Figure 1.1 provides an overall summary of the above-outlined findings in comparison between first- and third-party perspective.

Figure 1.1

Overview of developments in social attention required for infants to learn from first- and third-party social interactions

First-Party Interactions		Third-Party Interactions
• Sensitivity to communicative signals *(Farroni et al., 2002)* • Detecting another individual as social partner *(Simion et al., 2011)*	**Birth**	
Coordinating attention with a social partner in turn-taking dyadic interactions *(Rochat & Striano, 1999)*	**2 mo**	
Coordinating emotions with a social partner in turn-taking dyadic interactions *(Aureli et al., 2017)*	**4 mo**	
Emergence of gaze following • Following the attentional focus of a social partner to an object *(Astor & Gredebäck, 2019)*	**6 mo**	• Detecting a social relationship between two human agents *(Papeo et al., 2020)* • Following the turn-taking dynamic of an observed dyadic face-to-face interaction *(Augusti et al., 2010)*
Emergence of joint attention • Coordinating attention with a social partner to an object of mutual interest *(Striano & Rochat, 1999)*	**10 mo**	Predicting the turn-taking dynamic of an observed dyadic interaction *(Bakker et al., 2011)*
From responding to joint attention to initiating joint attention *(Carpenter et al., 2007)*	**14 mo**	
	18 mo	Inferring others' communicative intent and collaborative goals in an observed triadic interaction *(Thorgrimsson et al., 2014)*

Notes. All cited references represent study examples outlined in more detail together with further evidence in section 1.2 in this chapter.

1.3 Motivational Mechanisms Affording Opportunities to Learn From Social Interactions

So far, I have taken a rather cognitive view on the social attentional foundations of infants' learning from social interactions. However, to gain a comprehensive understanding of the processes involved, we must also consider the influence of motivational factors—especially given the increasing influence of endogenous motivations on infants' visual attention. Within the framework of this thesis, intrinsic motivations like infants' social motivation or their motivation to acquire knowledge are considered "door openers" to potential learning opportunities as they drive infants' attention toward social interactions (e.g., Su, Rogers, Estes, & Yoder, 2021).

1.3.1 Infants' Intrinsic Social Motivation

One proximal driver toward social interactions is infants' intrinsic social motivation (Tamir & Hughes, 2018). According to a definition by Chevallier and colleagues (2012), social motivation represents an evolutionary adaptation that has evolved in humans to enhance the individual's fitness in collaborative environments (see also Tomasello & Vaish, 2013). At the proximal level, social motivation is described as "[…] a set of psychological dispositions and biological mechanisms biasing the individual to preferentially orient to the social world (social orienting), to seek and take pleasure in social interactions (social reward), and to work to foster and maintain social bonds (social maintaining)" (p. 231, Chevallier et al., 2012). In the following subsections, I will give a brief overview of the current state of research on each of these three components. For a summary, see Figure 1.2 in the end of this section.

Attentional preference for social interactions ("social orienting")

Chevallier and colleagues (2012) define social orienting as prioritized attention to social signals, with prioritized attention meaning rapid attention capture and selective attention over non-social stimuli. Infants show early social orienting behaviors immediately after birth. The newborn attentiveness to biological motion, human faces, and direct gaze bootstraps infants' orienting toward potential interaction partners from early in ontogeny (Simion et al., 2011). Throughout the first year of life, infants' social orienting increases and becomes more fine-tuned. For example, infants show an increasing visual preference for socially meaningful faces from 6 to 9 months of age (Lee et al., 2017). By 14 months, infants also show prioritized attention to third-party interactions. When prompted to choose whether to attend to a video showing face-to-face interaction scene or a simultaneously displayed video showing a non-interactive back-to-back scene, 14-month-old infants look longer at the face-to-face stimulus (Galazka et al., 2014).

Intrinsic reward value of social interactions ("social reward")

According to the framework by Chevallier and colleagues (2012), the social reward component of social motivation refers to the incentive salience and hedonic value of social stimuli. At the behavioral level, social reward value is manifested in approach behaviors and expressions of positive affect. During the second half of the first year of life, infants demonstrate a series of behaviors indicating that they increasingly seek social engagement with others: For example, 7- and 10-month-old infants try to re-engage an experimenter when this person suddenly stops reacting to the child (Striano & Rochat, 1999). Moreover, between 9 and 14 months of age, infants increasingly perform bids for social interaction and coordinated attention toward social partners, including gestures, vocalizations, or facial cues (Clearfield et al., 2008). During the same period as infants begin to seek social interactions, they show signs of hedonic pleasure during social engagement: Around 6 months of age, infants smile more frequently (Hains & Muir, 1996) and engage with others over longer periods of time (Symons, Hains, & Muir, 1998) when their interaction partner is looking at them compared to looking away from them. Around the same age, infants engage in protoconversations during which they take turns in exchanging positive emotions—both with their caregiver, as well as with strangers (Rochat, Querido, & Striano, 1999). This suggests that infants' positive affect during protoconversations goes beyond infant-caregiver bonding, signaling a more general hedonic pleasure during turn-taking social exchange. From 5 to 9 months of age, infants show increasing positive affect during coordinated attention episodes (Striano & Bertin, 2005) and from 8 to 12 months, they increasingly smile in anticipation of successfully initiated joint attention (Venezia et al., 2004; Venezia Parlade et al., 2009).

In addition to seeking and liking social exchange, Berridge and Robinson (2003) propose reinforcement learning as a third component of social reward. Following reward-learning principles, stimuli with intrinsic reward value reinforce actions and behaviors that lead to the rewarding outcome (Skinner, 1938). Focusing on *social* rewards more specifically, stimuli with high social-emotional value such as the face of the own mother increase 8-month-olds' pupillary response and promote their learning of spaciotemporal patterns (Tummeltshammer et al., 2019).

It remains less clear whether infants' orienting toward third-party social interactions is driven by similar reward mechanisms. Indication comes from a study revealing that 12-month-old infants show enlarged pupil sizes while observing face-to-face conversations compared to mirrored back-to-back scenes (Gustafsson et al., 2016; see also Cheng, Liu, Yuan, & Jiang, 2021). It has been suggested that pupil size is modulated by reward-related processing, among many other influential factors such as physiological arousal, focused attention, and cognitive processing (Bijleveld, Custers, & Aarts, 2009; Tummeltshammer et al., 2019). Following this assumption, the increased pupillary response toward face-to-face interactions may indicate an underlying influence of social reward mechanisms. This possibility is further supported by a previous study

indicating the influence of affect during infants' observation of third-party interactions. At 7 months, infants show more positive facial expressions when observing a prosocial interaction between two puppets compared to observing an antisocial interaction (Steckler et al., 2018). However, this finding does not reveal whether infants find social interaction in itself, that is, without explicit positive or negative moral value, inherently rewarding.

Fostering and maintaining social bonds ("social maintaining")

Another key manifestation of social motivation is social maintaining, which Chevallier and colleagues (2012) define as "an individuals' desire to engage with others over sustained periods of time" (p. 232). According to the authors, strategies for social maintaining include behaviors enabling an individual to establish, maintain, and foster social bonds and relationships with others. Behavioral manifestations of social maintaining typically emerge around preschool age and include, for example, ingratiating behaviors reflecting the concern for others' acceptance (Fu & Lee, 2007), long-term commitment to social relationships (Misch, Over, & Carpenter, 2016), or the emergence of stable friendship patterns (Gifford-Smith & Brownell, 2003). In the remainder of this thesis, I will mainly refer to social orienting and social reward as more age-appropriate manifestations of social motivation in the first year of life. Nevertheless, it should be noted that there is some indication that precursor abilities of social maintaining emerge already during the first 1.5 years of life: In the second half of the first year, infants can distinguish and recognize individual faces (Pascalis et al., 2011) and make attempts to maintain reciprocal social exchange once established (Striano & Rochat, 1999). Moreover, according to Over (2016), 9-month-olds' joint attention (Carpenter, 2010), 14-month-olds' helping (Warneken & Tomasello, 2007), and 18-month-olds' action imitation (Nielsen, 2006) are at least partly influenced by an early motivation to affiliate with others. In addition, infants show precursor preferences required to form long-lasting bonds with social group members. For example, they show a preference for particular peers and spend more time with certain individuals over others (Hay, Payne, & Chadwick, 2004).

It remains unclear from the previous literature as to which the desire to form long-lasting social bonds extends to third-party interactions. Following the assumption that the motivation to foster and maintain social relationships has evolved from the human reliance on social group members (Over, 2016), it would be adaptive to develop long-lasting preferences for interactions between particular partners as well, for example, members of the own social group. In line with this idea, it has been theorized that humans do not only seek isolated bonds with single individuals, but also with particular social groups (Swann, Gómez, Seyle, Morales, & Huici, 2009; as cited in Over, 2016). Moreover, a recent study by Liberman, Kinzler, & Woodward (2021) has shown that infants can make inferences about third-party affiliative relationships based on observed homophily. In the study, 14-month-olds expected two people with a shared food preference to be more likely to affiliate (i.e., to face each other instead of turning away from one another) than two

people with opposite food preferences, even without any signals of a friendship between the agents. Taken together, previous findings from the first-party interaction literature suggest that social motivational mechanisms bias infants' attention and behavior toward social interactions. As illustrated in the figure below, comparatively little is known about the influence of social motivational mechanisms on infants' attention to third-party interactions.

Figure 1.2

Overview of developments in social motivational mechanisms driving infants toward first- and third-party social interactions
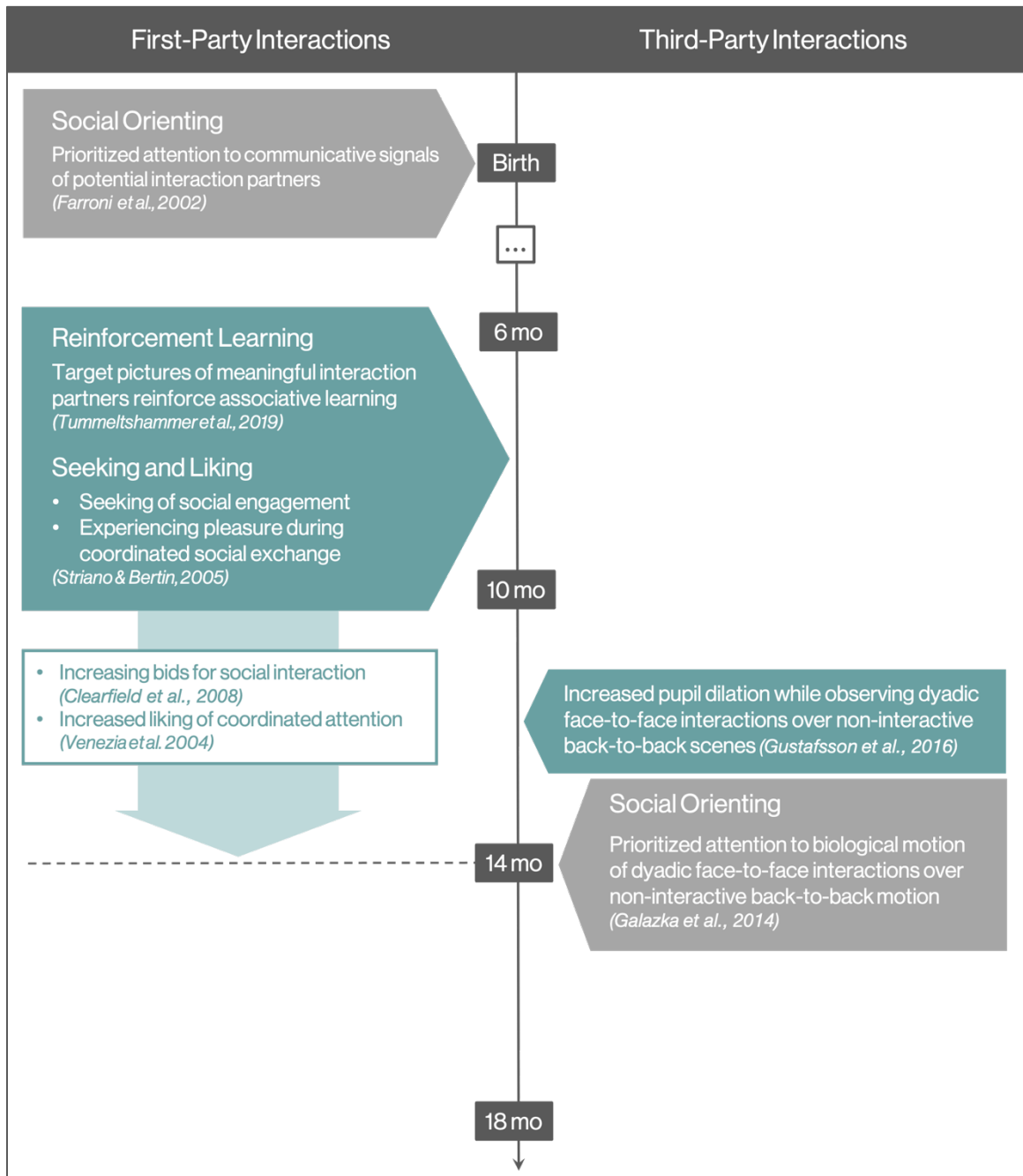
Notes. The turquoise-colored text boxes represent behavioral manifestations of social reward (Berridge et al., 2009; Chevallier et al., 2012). All cited references represent study examples outlined in more detail together with further evidence in section 1.3.1 in this chapter.

1.3.2   Infants' Intrinsic Motivation to Learn From Others

According to curiosity-driven theories, another influential motivation is infants' intrinsic motivation to learn from others. Litman (2005) defines curiosity as the "desire to know, to see, or to experience; that motivates exploratory behavior directed towards the acquisition of new information" (p. 79, Litman, 2005; see also Jirout & Klahr, 2012 for an overview of different definitions). As summarized in Litman's article about the "pleasures of learning", curiosity is thought to elicit approach behavior and experiences of intrinsic reward as well (Litman, 2005; Berlyne, 1966; Loewenstein, 1994). Similar to the terminologies used in the social motivation literature, the author refers to the "wanting and liking" of novel information.

A common underlying assumption of curiosity-driven learning accounts is that knowledge-seeking motives organize infants' behavior and learning towards acquiring novel information. Supporting evidence comes from first-party studies showing that infants actively modulate their social behavior in favor of acquiring novel information. For example, 12-month-old infants perform more pointing gestures when their behavior elicits an "informing" response in their interaction partner (i.e., showing an emotional attitude toward the object) compared to when it elicits a "sharing" response (i.e., gaze alternation between the object and the infant, Kovács, Tauzin, Téglás, Gergely, & Csibra, 2014). Moreover, 16-month-olds perform more pointing behaviors in the direction of a novel object when they interact with a knowledgeable informant compared to an incompetent informant, indicating a preference to gather information from knowledgeable partners (Begus & Southgate, 2012; see also Crivello, Phillips, & Poulin-Dubois, 2018). Around the same age, toddlers use their previous learning experience to adjust their behavior in future learning situations. For example, 18-month-olds are more likely to imitate an action if the demonstrator of this action has been a reliable compared to an unreliable informant in a previous word learning task (Brooker & Poulin-Dubois, 2013). Taken together, these findings indicate that infants and toddlers around 12 months and older can evaluate whether an ongoing first-party interaction carries the potential to acquire knowledge. Moreover, they can use this information to invest their resources selectively in situations where they expect to maximize their learning benefit.

It should be noted that the studies in the empirical part of this thesis do not directly examine the influence of information-seeking motives on infants' attention and learning. Nevertheless, it needs to be acknowledged that infants' desire to learn from others represents another candidate mechanism that may steer infants' visual attention toward third-party interactions. I will elaborate more on this possibility in the general discussion, where I interpret the findings from this thesis in the light of infants' motivation to acquire knowledge (chapter 5, section 5.2.1).

1.4 Infants' Learning From Social Interactions

At the beginning of this chapter, I gave an overview about developments in social attention that are essential for infants to learn from own and others' social interactions. Then, I highlighted the importance of motivational mechanisms as intrinsic drivers toward social interaction opportunities. What remains open is the question of what we know about infants' actual *learning* from social interactions. Can infants in the first year of life encode information provided within an interactive situation, store it over time, and retrieve it outside the interactive situation (memory definition by McDermott & Roediger, 2018)? If so, what factors facilitate this learning?

Generally speaking, infants can acquire a range of socially and culturally relevant information from social interactions, including knowledge embodied in merely interpersonal interaction such as coordinated action rituals (Legare & Nielsen, 2020) or knowledge about external content such as tools or novel objects (Cleveland, Schug, & Striano, 2007). In the following sections, I will primarily focus on the latter case, given that the empirical part of this thesis relies on object encoding as a measure of early learning (Study 3).

1.4.1 Infants' Learning About Objects Through Own Participation in First-Party Social Interactions

To investigate whether and how young children use social interactions to gather knowledge about objects in their environment, previous studies have examined different domains of knowledge, including object-directed actions (Carpenter, Akhtar, & Tomasello, 1998; Killen & Uzgiris, 1981; Nielsen, 2006) and object-labels (Hirotani, Stets, Striano, & Friederici, 2009; Werker, Lloyd, Cohen, Casasola, & Stager, 1998; Woodward & Hoyne, 1999). Both action imitation and word learning require comparatively complex skills such as motor or language skills and have been—in the context of social interactions—mainly studied in the second year of life. A frequently used measure of object-related memory process within the first year of life is object encoding (McDermott & Roediger, 2018).

Object encoding as a measure of early learning about objects

The encoding of visual object features represents a fundamental requirement for any more complex form of learning about objects, such as the learning of object-related functions, emotions, actions, or words. To learn an object-specific label, for example, infants must encode the word label, the visual features of an object, and the association between auditory information and referenced visual information ("label-object mapping", Spiegel & Halberda, 2011). As an indicator of object encoding, previous studies have used infants' novelty preference, a defining feature of the infant visual recognition memory (Fantz, 1964; Rose et al., 1982). Infants' novelty preference can be measured with behavioral measures and measures of neural activity. The behavioral version

of the object encoding paradigm typically includes an object encoding phase during which the infant is familiarized with a target object, and a subsequent paired preference phase during which the previously familiarized object reappears next to a novel object. Depending on the study design, the "novel" object is either an entirely new object that the infant sees for the first time when it appears in the paired preference phase (e.g., Cleveland et al., 2007) or a non-cued object that has been visible during the previous encoding phase without being highlighted by any cue (e.g., Theuring, Gredebäck, & Hauf, 2007). As an indirect measure of previous encoding success, the paradigm relies on infants' attention response in the paired preference phase. An increased response to the novel, yet unprocessed object is assumed to reversely indicate recognition of the fully encoded familiarized object. As behavioral indicators for an increased novelty response, looking times or touching preferences have been used in previous studies (for reviews on infant recognition memory, see Pascalis & de Haan, 2003; Reynolds, 2015).

Factors influencing object encoding in first-party social interactions

In direct interactions with a social partner, a crucial determinant of infants' object encoding is the partner's object-directed gaze. When seeing an adult moving their gaze toward an object after gazing in the direction of the infant, 4-month-olds spend less time looking at this object when it subsequently appears along with a novel object, reversely indicating that the gaze-cued object had been encoded during the previous encoding phase (Reid & Striano, 2005; Wahl, Michel, Pauen, & Hoehl, 2013). Previous EEG studies using event-related potentials (ERPs) as a measure of attention and memory encoding, have revealed a corresponding result at the neural level: 4-month-olds show an enhanced brain activity in response to non-cued objects compared to previously gaze-cued objects (enhanced positive slow wave activity: Reid, Striano, Kaufman, & Johnson, 2004; Hoehl, Wahl, Michel, & Striano, 2012; enhanced negative central component: Hoehl, Reid, Mooney, & Striano, 2008; Wahl et al., 2013). The promoting effect of object-directed gaze cues on young infants' object encoding remains stable for movements of isolated schematic eyes, suggesting that facial context is not required to elicit the effect (Wahl, Marinović, & Träuble, 2019). As this only applies to schematic eyes with intact contrast polarity (i.e., black circles moving on white oval shapes), not to cues with reversed polarity (i.e., white circles moving on a black background), it is assumed that infants' sensitivity to referential gaze cues relies on the typical perceptual pattern of the eyes (Michel, Pauen, & Hoehl, 2017; Wahl et al., 2019).

According to the *Natural Pedagogy* account (Csibra & Gergely, 2006, 2009), observed direct gaze to an object is not sufficient for infants to learn about objects in direct interactions. Instead, it is assumed that infants rely on communicative signals to evaluate the relevance of an object-directed gaze shift and show increased responsiveness to the cued content. Okumura and colleagues (2020) have systematically tested this assumption in a series of experiments. In supporting the assumption of the *Natural Pedagogy* account, the researchers found that 9-month-

old infants' encoding of an object was only enhanced, if an adult had addressed the infant through direct gaze or infant directed speech before shifting their gaze toward the target object (see chapter 4 for more details). Moreover, studies comparing human interaction partners with robot partners indicate that infants rely on human sources of information when learning about objects. When seeing a robot agent looking in the direction of the child before shifting their head and gaze toward an object, 12-month-olds infants can follow the robot's gaze (Meltzoff, Brooks, Shon, & Rao, 2010), but they do not encode the gazed-at object unless the gaze shift is preceded by infant directed speech (Okumura, Kanakogi, Kanda, Ishiguro, & Itakura, 2013). The influence of ostensive signals on infants' object encoding has been demonstrated in screen-based setups in which the interaction partner was portrayed by an adult on screen (Okumura et al., 2020), as well as in real-interactive setups during which the infant was actively engaged with a real adult (Cleveland et al., 2007; Cleveland & Striano, 2007).

As described in more detail in Study 3 (chapter 4, section 4.1), the availability of ostensive signals alone, like isolated object-directed gaze alone, is insufficient for infants' object encoding. This leads to the conclusion that the combination of ostensive signals and object-directed gaze is important for early object encoding in first-party interactions. This assumption receives further support from studies showing that joint attention, by definition including the interplay between ostensive eye contact and mutual gazing to an external content, represents an important role for the successful learning of word-object relations in toddlerhood (Baldwin, 1995; Hirotani et al., 2009; Tomasello & Farrar, 1986).

1.4.2 Infants' Learning About Objects Through Observing Interactions Between Third Parties

It remains unknown whether infants in the first year of life can learn about novel objects by observing social interactions between third parties. The earliest age at which learning from third-party interactions has been demonstrated is 18 months. Around this age, toddlers begin to imitate object-related actions they have observed in a demonstration directed toward another person and they do so more frequently compared to a solitary observational situation in which they have seen the demonstrator acting on their own, without addressing anyone (Matheson et al., 2013). In addition, 18-month-old toddlers can learn novel object labels by merely overhearing conversations between others (Floor & Akhtar, 2006; Gampe, Liebal, & Tomasello, 2012). In corresponding studies, toddlers have been typically presented with a "finding game" scenario, during which they observed an adult (E1) introducing another adult (E2) to a range of novel objects hidden in separate opaque boxes: one target object and 2-3 distractor objects. According to the procedure of the game, E1 opened one box after the other, removed the object from the inside, held it up, and handed it to E2. Before opening each box, E1 introduced the hidden object. The target object was introduced with a labeling statement (e.g., "I'm going to show you the toma. Let's see the toma. I'm going to

find the toma."). The distractor objects were introduced with a non-labeling statement (e.g., "I'm going to show you this one. Let's see this one. I'm going to find this one."). As a measure of word learning success, the participants were subsequently asked to identify the target object when seeing all previously introduced objects at the same time. For example, E1 said: "Can you show me the toma? Which one is the toma?".

Studies using this task have shown that, toward the end of the second year of life, toddlers' word learning through overhearing is equally good as their learning in direct pedagogical settings at this age (Akhtar, Jipson, & Callanan, 2001; Gampe et al., 2012). Moreover, the ability to learn novel object labels via overhearing has been demonstrated in diverse settings, including live interactive settings and screen-based settings during which pre-recorded conversations have been presented on a screen (O'Doherty et al., 2011). Shneidman and Woodward (2016) provide a systematic overview of previous studies comparing word learning and action imitation in child-directed and third-party observational contexts.

*Factors influencing the learning about objects in third-party interactions*

What factors make an observed interaction a good learning opportunity? In addressing this question, a study by Fitch, Lieberman, Luyster, & Arunachalam (2020) has revealed that, similar to first-party settings, one influential factor seems to be observed joint attention: 20- to 30-month-old toddlers learned a novel object label only when seeing speaker and listener attending to the object together while the speaker labeled the object. Toddlers did not learn the object label when the listener in the observed scenario was engaged with another individual activity without looking at the object or the other person.

Another influential factor seems to be observed reciprocal action. In a study by O'Doherty and colleagues (2011), 30-month-old toddlers learned a novel object label better if the observed interaction included a reciprocal interaction between speaker and listener than an observed one-sided demonstration during which the listener merely observed the speaker. During the reciprocal interaction, the speaker labeled the object, performed an action on it, and then handed it to the listener, who imitated the observed action on the object. In contrast to the one-sided demonstration, the listener did not only share attention with the speaker but furthermore made an active contribution to the interaction by accepting the object and imitating the observed action.

Taken together, the previously outlined findings indicate that, at least by the end of the second year of life, toddlers can learn about objects through merely observing social interactions between others. Figure 1.3 summarizes the main findings from the above-outlined studies on infants' learning about objects in first- and third-party interaction contexts.

Figure 1.3

Overview of developments related to infants' learning about objects from first- and third-party social interactions
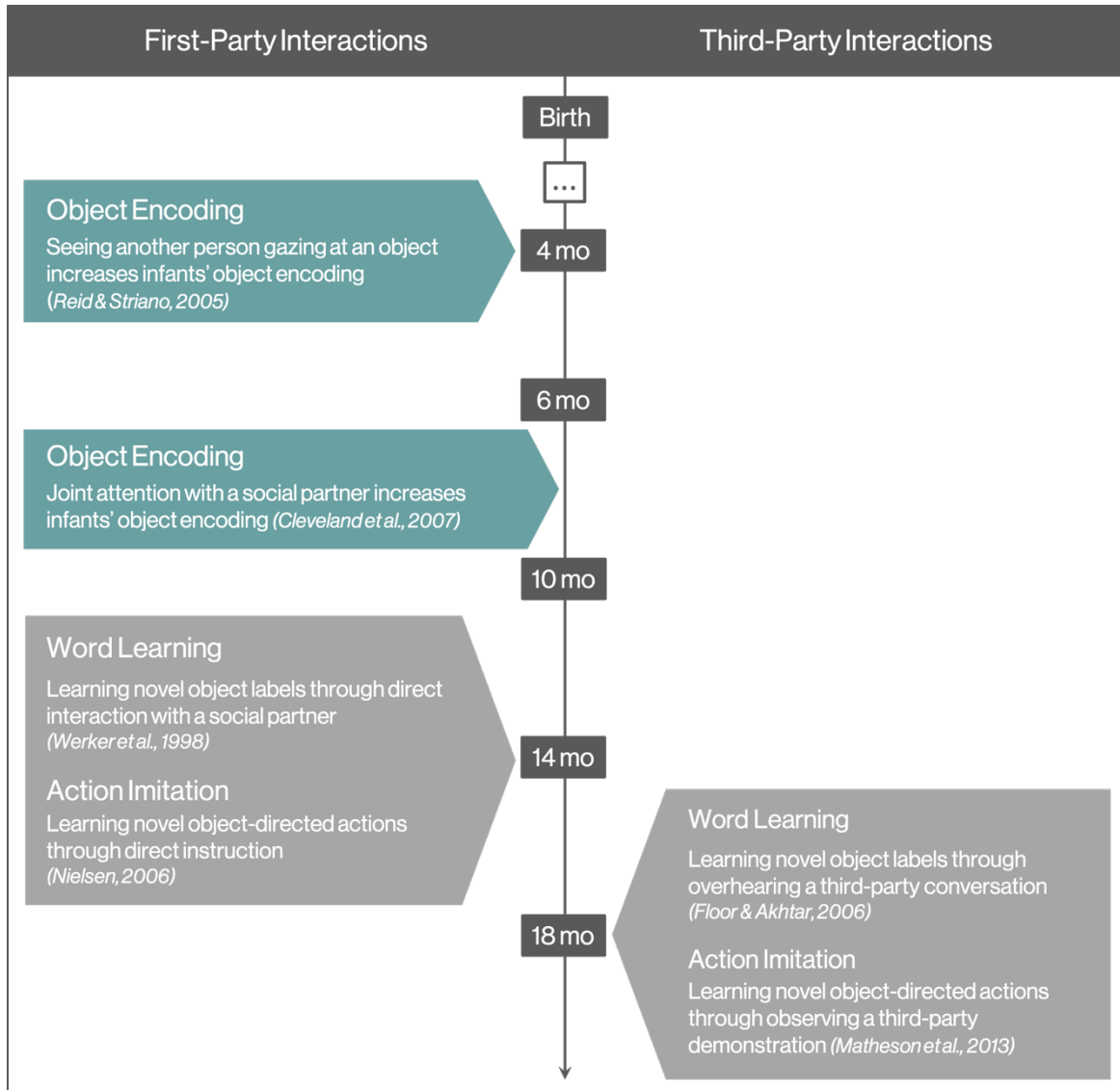
Notes. All cited references represent study examples outlined in more detail together with further evidence in section 1.4 in this chapter.

1.5 Research Gaps

As the literature review in this chapter shows, the infancy research in the past ten years has substantially enhanced our understanding of how infants represent social relationships between third parties and how they develop an understanding of third-party interactions throughout the first year of life. Nevertheless, many questions remain to be studied to gain a comprehensive understanding of how infants come to learn from third-party interactions. In this section, I will outline three shortcomings in the previous literature, specifically focusing on the ones that will be addressed in the empirical part of this thesis. In the general discussion (chapter 5), I will discuss other research gaps that should be addressed in future studies.

As illustrated in Figure 1.2, one gap in the third-party interaction literature is the lack of studies investigating the influence of motivational factors on infants' visual attention to third-party interactions. Previous first-party studies have suggested a significant increase in infants' orienting to and seeking of social interactions during the second half of the first year of life. Moreover, toward the end of the first year, infants engage in behaviors indicating that they find it intrinsically rewarding to engage in social interactions with others. It remains unclear whether infants' attention to third-party interactions is influenced by similar motivational mechanisms. It would be possible that infants' intrinsic motivations steer their attention toward social interactions more generally, independent of whether the social interaction is experienced through direct participation or third-party observation. Investigating this possibility would increase our understanding of the processes and mechanisms involved in infants' orienting toward third-party interactions. Moreover, it would shed light on whether infants take an active role in shaping the availability of observational learning opportunities.

Another significant gap is that the ontogenetic origins of infants' memory and learning in observational contexts remain understudied. As Figure 1.3 illustrates, 18 months is the earliest age at which object learning has been previously demonstrated in third-party interaction contexts. It remains unclear when this ability emerges and whether already infants in the first year of life can learn about their environment through observing third-party interactions. Indication comes from the social attention literature suggesting that already in the second half of the first year of life, infants possess foundational social attention abilities and mechanisms required to learn from third-party interactions (see Figure 1.1). This raises the possibility that the ability to learn through observing others' interactions may evolve at a younger age than previously shown.

Furthermore, following from the lack of evidence about whether infants in the first year of life can learn from third-party interactions, it remains unknown which factors within an observed interaction promote infants' learning about objects. Previous first-part studies have revealed that around 9 months of age, joint attention represents an essential context for object-related learning in direct interactions. Studies with 18-month-old toddlers have suggested that similar factors may

promote object learning in third-party interaction contexts. Studying this possibility in 9-month-old infants would deepen our understanding of the factors involved in early observational learning. Moreover, it would allow to compare object-related learning in observational contexts with object-related learning in direct interactive settings around the same age.

1.6 Focus of This Dissertation

The overarching aim of this dissertation is to deepen our understanding of the underlying processes and mechanisms involved in infants' observational learning from third-party social interactions. As introduced in this chapter, this will be achieved by taking a broader perspective, which not only focuses on infants' memory and learning during the actual observation, but also considers social attentional developments and underlying motivational influences enabling infants to detect and approach situations in which they can observe social interactions. The empirical part of this thesis (chapters 2–4) comprises three pre-registered experimental studies that address open gaps in the previous literature on infants' attention to and learning from third-party social interactions. In the following subsections, I will give an overview of each of these studies: First, I will introduce the research questions of the studies and give a brief description of the corresponding study designs and dependent measures. Then, I will give an overview of the overall study population, introduce the methodological focus on eye tracking related measures, and describe the stimuli used across studies. Table 1.1 in the end of this section gives an overview about the three studies at a glance.

1.6.1 Research Questions

Study 1: How does infants' orienting to third-party interactions develop during the second half of the first year of life, and how does this development coincide with changes in infants' active social attention behavior?

Previous research has revealed that 14-month-old infants preferentially orient their attention to situations where they can observe face-to-face interactions between two people (versus non-interactive back-to-back scenes, Galazka et al., 2014). However, the ontogenetic origins of this preference remained unclear. Considering the possibility that more general motivational mechanisms may navigate infants' social attention in first- and third-party interactive contexts, the objective of Study 1 was to investigate how infants' orienting to third-party interactions develops between 7 and 13 months of age—a period during which infants' social orienting in direct interactions undergoes considerable developments. Moreover, by systematically examining infants' social attention behavior in direct interactions in addition to their orienting toward others' interactions, this study aimed to compare the developmental trajectories in both modalities at the group level and at the individual level. To measure infants' orienting to third-party interactions,

we measured their looking times while they were simultaneously presented with two types of videos in a forced-choice preferential-looking task: one video showing a dyadic face-to-face interaction between two adults, and the second video showing a non-interactive scene during which the same actors performed the identical motions standing back-to-back. To measure infants' social attention behavior in direct interactions, we recorded their natural interaction behavior during a free play session with their caregiver and coded four infant looking behaviors based on these recordings: a general attentional interest in their caregiver, looks at the caregiver's face, eye contact, and joint attention looks.

Study 2: Is it intrinsically rewarding for infants to observe third-party interactions?

Another question that remains open from previous research is what proximal mechanisms drive infants toward third-party social interactions. Previous first-party studies have revealed that infants find it intrinsically rewarding to engage in direct social interactions with others. Building up on this finding, the aim of Study 2 was to investigate whether similar reward mechanisms underlie infants' attentiveness to third-party interactions. To achieve this aim, we examined if third-party interaction videos can serve as a primary reinforcer in an associative learning task. Thirteen-month-old infants participated in a gaze-contingent associative visual learning task during which they saw two non-social cues (a circle or a triangle) repeatedly paired with two kinds of target videos appearing left or right from the cue: a dyadic face-to-face interaction or a non-interactive back-to-back scene. To compare infants' learning of the interaction-predictive association with their learning of the non-interaction predictive association, we measured the change in infants' saccadic latencies from the central cue to the correct target region over trials.

Study 3: Can infants in the first year of life use third-party interactions to learn about objects? If so, do similar processes contribute to infants' observational learning from third-party interactions as to their referential learning in direct social interactions?

Previous studies have revealed that 18-month-old toddlers can learn novel object labels and novel object-related actions by merely observing others' interactions. However, it remains unclear whether preverbal infants can learn about objects by observing others' interactions and, if so, whether similar factors contribute to this observational learning as to their referential learning through direct social engagement, that is, an interplay between eye contact and direct gazing to an object. To address these questions, the objective of Study 3 was to investigate 9-month-old infants' object encoding in the context of third-party interactions. In a screen-based object encoding task, infants were presented with four kinds of videos in which they saw two adults together with one object. Based on a 2×2 design, the videos were manipulated regarding the eye contact between the two agents (eye contact or no eye contact) and the agents' object-related gaze (looking toward the object, looking away from the object). As a measure of object encoding, the

study relied on infants' looking time response when the familiarized object reappeared next to a novel object in a subsequent preferential-looking phase, assuming that longer looking times to the novel object would reversely indicate an increased previous encoding of the familiarized object. We compared infants' object encoding in the third-party observational setting with a matched first-party setting in which another sample of 9-month-old infants was directly addressed by an adult on screen.

1.6.2 Study Population

The overall study population consisted of 7- to 14.5-month-old typically developing infants growing up in a Western, urban, industrialized environment. Study 1 and Study 3 were tested in Leipzig, a mid-sized city in Germany with approximately 605,000 inhabitants. Study 2 was tested in Uppsala, a mid-sized city in Sweden with approximately 177,000 inhabitants. In both contexts, infants' social experiences occur primarily within the nuclear family setting, typically consisting of three to five household members. Infants in the tested age range typically spend their days in the presence of their primary caregiver or, around one year and older, partly in daycare facilities. The prototypical socialization context in Western middle-class societies, like Germany and Sweden, is a child-centered setting during which infants experience high levels of direct pedagogy (Keller, 2007; Tamis-Lemonda & Song, 2012). Adults use infant-directed speech and communicative signals to engage the infant in face-to-face interactions and object play (Keller, Hentschel, et al., 2004). In the general discussion, I will discuss the implications of the findings from this thesis from a cross-cultural perspective (Gaskins & Paradise, 2010; Rogoff et al., 2003).

1.6.3 General Measures

All three studies in this thesis relied on measures of overt visual attention, recorded via eye tracking (Study 1–3) and manual coding of infants' looking behavior during a naturalistic interaction (Study 1). Since eye tracking methodology was the primary tool across studies, I will give a brief overview about the general eye tracking setup and the eye tracking related measures.

Even though the three studies were tested in different laboratories, all participants experienced the same eye tracking setup. They sat on their caregiver's lap in front of a screen with an attached remote eye tracker. The eye tracking area was separated by cloths or partition walls in all testing rooms to shield the participant from any visual distraction. All eye trackers used had a minimum sampling rate of 120 Hz and a similar gaze tracking accuracy around 0.4°. Overall, three different eye tracking paradigms were applied in the included studies: a preferential-looking paradigm (Study 1), an associative visual learning paradigm (Study 2), and an object encoding paradigm (Study 3). All paradigms relied on fixation-based measures, including duration of fixation as a measure of looking time (Study 1 & 3) and duration until first fixation as a measure

of saccadic latency (Study 2). In the preferential-looking paradigm, longer looking times at a stimulus were used as an indicator of a visual preference for this stimulus over a simultaneously presented stimulus (see also Spelke, 1985; Teller, 1979). In the object encoding paradigm, shorter looking times at a previously familiarized object compared to a novel object were used as an indicator of infants' previous encoding of this object (see also Cleveland et al., 2007). Finally, in the associative visual learning task, saccadic latencies were used as a measure of visual reaction time from a central cue to a subsequently displayed target stimulus (see also Wang et al., 2012). Taken together, infants' fixation behaviors were used as measures of their visual preferences, reaction times, and memory processes.

1.6.4 Stimuli

Third-party interaction stimuli

Across studies, the third-party interaction stimuli consisted of highly controlled silent videos showing dynamic interactions between two adult women. In total, seven women acted in the videos. All women came from Leipzig (Germany) and were between 27 and 34 years of age at the timepoint of filming. The external characteristics of the actors were kept as neutral as possible, meaning that they did not express any emotions, wore the same white t-shirts, had their hair tied back, and did not wear any jewelry or glasses.

Overall, two categories of social interaction videos were created for the studies included in this thesis: merely interpersonal dyadic interactions during which the actors engaged in a clapping game, touched their hands, or leaned their heads toward one another (Study 1 & 2), and triadic interactions including an object in addition to the social partners (Study 3). In Study 1 and 2, the same videos were used with different durations. To manipulate the relationship between the women as interacting or non-interacting, we used the movements of their bodies (turning toward or away from one another), the relative positioning of their bodies (face-to-face or back-to-back), and the actors' gaze direction (eye contact or looking in opposite directions) across videos.

In Study 1 & 2, we additionally included the execution of an action (co-regulated versus individually) and the amount of touch (mutual touch versus no touch) within the interactions. In Study 3, in contrast, no touch or manual action was included within the interactions. Instead, the interactions contained mutual looking toward a visible object (yes or no) to manipulate the "triadicness" of the interaction. The control conditions resulting from the individual study designs are described in the individual study chapters.

First-party interaction stimuli

While Study 2 focused only on the third-party perspective, Study 1 and Study 3 included a comparison with infants' first-party perspective in social interactions. In Study 1, the focus was

on infants' naturalistic social attention behavior. Thus, the social "stimulus" was the caregiver of the child during a free play interaction. In Study 3, in contrast, a screen-based setup was used during which infants were addressed by an adult on screen who acted according to a standardized script. The content of the videos was closely matched to the third-party interaction stimuli created for this study.

Video creation

The general steps of video creation were the same for all stimuli. All actors were filmed individually in front of a green screen to ensure the flexible and accurate positioning of dyad partners and control for color and luminance differences between and within videos. The post-production of the footage consisted of multiple steps, including a frame-by-frame adjustment of the action timing for each actor and action and the replacement of the green background with a grey background layer which was identical across videos and studies. The edited individuals were paired in dyads for the third-party interaction stimuli—depending on the condition in a face-to-face or back-to-back arrangement. Technical details about filming and editing are provided in the supplemental materials of each study and can be found in the Appendix.

Table 1.1

Overview of the studies included in the empirical part of this thesis

	Study 1	Study 2	Study 3
Research Question	*How does infants' attentional orienting to third-party interactions develop in relation to changes in infants' active social attention behavior?*	*Is it intrinsically rewarding for infants to observe third-party interactions?*	*Can infants learn from third-party interactions and, if so, do similar factors contribute to this learning as to their learning from first-party interactions?*
Population	• Leipzig (Germany) • 7.0 – 13.5 months of age	• Uppsala (Sweden) • 13.0 – 14.5 months of age	• Leipzig (Germany) • 9.0 – 10.0 months of age
Paradigm/ Task	I. Forced-choice preferential-looking task (eye tracking) II. Parent-infant free play	• Associative visual learning task (eye tracking)	• Object Encoding task (eye tracking)
Dependent Measures	I. Looking time (duration of fixations) II. Frequency of infant social looking behaviors (coded from free play recordings)	• Saccadic latency (duration until first fixation)	• Looking time (duration of fixations)
Third-Party Stimuli	Dyadic interactions including coordinated actions between two adults (same stimuli in Study 1 and 2).		Triadic interactions between two adults and an object.
Comparison First-Party?	Yes	No	Yes

2

Study I

Infants' preference for social interactions increases from 7 to 13 months of age

Abstract

This study examined 7- to 13.5-month-old middle-class Western infants' visual orienting to third-party interactions in parallel with their social attention behavior during own social interactions (Leipzig, Germany). In Experiment 1, 9.5-to-11-month-olds ($N = 20$) looked longer than 7- to 8.5-month-olds ($N = 20$) at videos showing two adults interacting with one another when simultaneously presented with a scene showing two adults acting individually. Moreover, older infants showed higher social engagement (including joint attention) during parent-infant free play. Experiment 2 replicated this age-related increase in both measures and showed that it follows continuous trajectories from 7 to 13.5 months ($N = 50$). This suggests that infants' attentional orienting to others' interactions coincides with parallel developments in their social attention behavior during own social interactions.

Keywords: social orienting, joint attention, 9-month-shift, social interaction, eye tracking

2.1 Introduction

Human infants strongly rely on social interactions to acquire culturally relevant knowledge about their environment. Not only active social engagement but also the observation of others' social interactions represents an essential source of social learning opportunities (Paradise & Rogoff, 2009; Tomasello, 2016). Already 9-month-old infants can encode novel objects by merely observing triadic joint attention interactions between two adults (Thiele, Hepach, Michel, & Haun, in press). Moreover, 18-month-olds can learn novel words through overhearing a third-party conversation between two people (Floor & Akhtar, 2006) and imitate actions they have observed in a demonstration directed toward another person (Herold & Akhtar, 2008; Matheson, Moore, & Akhtar, 2013). To learn from one's own or others' social interactions, infants first need to gain access to a potential learning opportunity. Theories highlighting the infant's active role in this process suggest that infants develop capacities and motivations guiding them toward social interactions.

Typically developing infants orient to social information from early on. This preference is crucial for infants to detect potential interaction partners and to structure and filter the large amount of information they are confronted with (Reid & Striano, 2007). Newborns preferentially orient toward face-like over non-face patterns (Goren et al., 1975), show enhanced neural processing of direct over averted gaze (Farroni, Csibra, Simion, & Johnson, 2002), spend more time looking at faces with opened than closed eyes (Batki, Baron-Cohen, Wheelwright, Connellan, & Ahluwalia, 2000), and prefer looking at biological motion over random motion patterns (Simion et al., 2008). During the first year of life, infants' social perception matures as their visual system, their practical experiences, and their understanding of others develop (Bertenthal & Boyer, 2015). For example, infants' preference for faces becomes increasingly robust from 3 to 6 months (Di Giorgio, Turati, Altoè, & Simion, 2012) and the face recognition system becomes attuned to human-specific features from 6 to 9 months of age (Pascalis et al., 2002).

In addition to identifying potential partners for direct interaction, infants can detect social relations between other people. From 6 months onwards, they perform more gaze shifts between two people facing each other during a turn-taking conversation compared to two people standing back-to-back while talking, and their gaze shifts become increasingly predictive toward the end of the first year (Augusti et al., 2010; Bakker et al., 2011; for equivalent findings with silent and still image stimuli, see Handl et al., 2013). Other findings suggest that the sensitivity to face-to-face arrangements emerges slightly later: 10- but not 9-month-old infants show increased looking times when seeing two people facing each other during a conversation, after a habituation phase showing the same individuals standing back-to-back while talking (and vice versa, Beier & Spelke, 2012).

The second half of the first year of life is marked by significant changes in infants' active interaction behavior (e.g., Callaghan et al., 2011). While infants engage in dyadic face-to-face interactions from 2 months on (Aureli et al., 2017; Striano, 2001), they begin to develop competencies for triadic social interactions in the second half of the first year (Carpenter, 2010; Carpenter, Nagell, et al., 1998; Striano & Reid, 2006). In addition to social cognitive developments (including an emerging understanding of others as intentional agents, Tomasello & Carpenter, 2007), infants' social attention is marked by significant changes in social motivation, including an increasing interest in coordinating attention with others. From 9 to 12 months of age, infants engage with an increasing frequency in joint attention and begin to initiate joint attention episodes themselves (Carpenter, Nagell, et al., 1998). Between 9 and 14 months, infants start to signal communicative intent toward an interaction partner (e.g., using ostensive gaze cues, gestures, vocalizations, Clearfield et al., 2008), and 7- to 10-month-old infants make increasing attempts to re-engage a person who stops reacting to them (Striano & Rochat, 1999). The exact onset age of joint attention has been a matter of debate. Some researchers suggest an early onset and gradual increase starting around 6 months (Bakeman & Adamson, 1984; Striano & Bertin, 2005), while others argue that truly joint attention abilities do not emerge before 9 months of age (Carpenter, Nagell, et al., 1998; Tomasello, 1995). According to both perspectives, however, the second half of the first year of life marks a critical period in infants' social development and learning, as infants' emerging capacity to coordinate attention with others provides the necessary basis for teaching and cooperation (Csibra & Gergely, 2006; Tomasello et al., 2005). Moreover, joint attention facilitates 7- and 9-month-old infants' processing of novel objects (Cleveland & Striano, 2007; Striano, Chen, Cleveland, & Bradshaw, 2006) and promotes future language learning (Morales et al., 2000). Combining these two strands of evidence, infants' increasing motivation to engage in joint attention enhances the availability of potential learning opportunities.

There is some indication from previous studies that infants' attention to third-party interactions is influenced by motivational factors as well. At least by the end of the first year of life, infants *prioritize* face-to-face interactions when choosing between attending to a face-to-face or a back-to-back scene including two human agents. Fourteen-month-olds look longer at biological motion of face-to-face interactions (point-light displays of two people engaging in a falling-catching or a pushing interaction), compared with biological motion of mirrored back-to-back scenes (two people performing the identical movements while standing back-to-back; Galazka, Roché, Nyström, & Falck-Ytter, 2014). Infants' looking preference disappears when seeing the same stimuli upside down, indicating that the longer looking times at upright face-to-face scenes do not reflect a response to the low-level perceptual features, but rather a greater interest compared to the "competing" back-to-back scene (Galazka et al., 2014). This interpretation is further supported by a previous study showing that 13-month-old infants organize their attention

and associative learning in favor of predicting and actively approaching situations in which they can observe a face-to-face interaction (Thiele, Hepach, Michel, Gredebäck, & Haun, 2021).

Together, these findings suggest that, at least by 14 months of age, infants selectively attend to situations in which they can observe a third-party interaction. What remains unclear, however, is how this attentional preference develops during the second half of the first year of life, when infants' social attention in direct interactions undergoes decisive changes. In contrast to previous work, this requires a systematic investigation of both infants' attentional preference for others' social interactions, as well as of their social attention behavior during active social engagement. Infants' emerging awareness of others as communicators of learnable content and their increasing motivation to seek social interactions may not only contribute to their social attention behavior in own interactions, but also enhance their attention to situations in which they can observe others' interactions. Support for this idea comes from active-learning accounts highlighting the influence of motivational mechanisms on infants' behavior and learning (for a review see Raz & Saxe, 2020). Theories of curiosity-driven learning, for example, claim that infants are intrinsically motivated to acquire knowledge and to learn from others (for a review see Begus & Southgate, 2018). Since both active social engagement and observations of others' interactions represent potential sources of social learning opportunities, it would be functionally adaptive if infants increasingly oriented their attention toward both situations. Moreover, social motivation theories raise the possibility that infants' increasing intrinsic social motivation may modulate their interest in social interactions beyond situations in which they are directly involved (see, e.g., Chevallier et al., 2012).

2.1.1 The Current Study

In this study we aimed to investigate developmental trajectories of infants' attentional orienting toward third-party social interactions and, moreover, examine how these changes coincide with infants' social orienting behavior during active social interaction. For this purpose, we assessed both infants' visual attention to third-party interactions and their active social attention behavior within the same testing sessions. Like most of the previous studies cited above, the current study was conducted in a Western, industrialized context where infants typically experience high levels of face-to-face interactions and direct pedagogy.

We conducted two experiments. In Experiment 1, we systematically investigated developments from before to after the previously suggested 9-month-threshold by comparing infants from two age groups (7 to 8.5 months and 9.5 to 11 months). All participants were tested in the same two experimental phases. First, we measured their looking times while they were simultaneously presented with two videos. One video showed two people turning towards one another while engaging in a social interaction, whereas the second video showed the same agents acting individually while standing back-to-back. To manipulate the relation between the persons

as interacting or non-interacting, we used the relative positioning of their bodies (face-to-face versus back-to-back), gaze direction (eye contact versus looking away), the execution of an action (co-regulated versus individually), and the amount of touch (mutual touch versus no touch). In the second phase, we observed the participant's behavior during free play with their parent and coded four kinds of looks in the direction of their parent (general looks at their parent, looks at their parent's face, eye contact, and joint attention looks).

We hypothesized that if infants from before to after 9 months of age develop an increasing interest in observing others' interactions, infants in the older (versus younger) age group should look relatively longer to the social interaction videos. Moreover, to probe infants' attentional preference for the social interaction videos, we tested infants' looking time score against chance level within the two age groups. We further hypothesized that if infants have an attentional preference for others' social interactions, they should spend more than 50% of their total looking time attending to the face-to-face interaction videos. Regarding infants' social attention during active social interaction, we hypothesized that if infants' social interest during active interaction increases with age, then infants in the older age group should perform more social looking behaviors during free play compared to the younger group. To examine the relation between infants' active social attention behavior and their attentional preference for others' interactions, we compared the developmental trajectories of both measures at the group level and explored the correlational relation at the individual level. Given the scarce literature about the immediate relation between infants' active social attention in direct interactions and their attentional orienting toward others' interactions, we did not pre-register any specific predictions concerning the degree of correlation between the two modalities but sought to explore this relation in reference to parallel findings at the group level. We hypothesized that if infants' attention to third-party interactions relates to their social orienting behavior in direct interactions at the individual level, the two measures should be correlated with one another.

In Experiment 2, we aimed to replicate our findings from Experiment 1 and build on them by testing infants at a broader and continuous age range between 7 and 13.5 months of age. This way, we aimed to gain a more comprehensive insight into the developmental trajectories in both modalities. We made analogous predictions as in Experiment 1, except that we did not test infants' looking preference against chance level. Specifically, we hypothesized that if infants' attentional preference for social interactions increases from 7 to 13.5 months of age, then the proportional looking time to the social interaction videos should increase with age. If infants' social interest during active interaction increases during this period, then infants' social engagement score should increase with age. Our assumptions regarding the relation between the two measures were the same as in Experiment 1.

The study was approved by the ethics committee of the Medical Faculty of Leipzig University. We pre-registered the hypotheses, methods, procedures, and the data analysis plans

for both experiments at the Open Science Framework (OSF). The pre-registration forms, all data, scripts for analyses, and supplementary materials are publicly accessible (Experiment 1: https://osf.io/42nyv/; Experiment 2: https://osf.io/s4uy7/).

2.2 Experiment I

2.2.1 Methods

Participants

Forty infants from two age groups provided valid data for both eye tracking and free play measures. The younger sample consisted of 20 infants between 7 months, 2 days and 8 months, 14 days (n = 10 female; M = 240.4 days, SD = 13.24 days). The older sample consisted of 20 infants between 9 months, 15 days and 10 months, 25 days (n = 10 female; M = 313.6 days; SD = 11.65 days). Data from 13 additional infants were excluded due to technical error (n = 1), failure of calibration (n = 8), preterm birth (n = 2), or because they were older than the inclusion criterion (n = 2). The aimed sample size was based on the upper range of the sample sizes in previous similar studies (e.g., Augusti et al., 2010; Handl et al., 2013). For the separate analyses of the free play data, we included all participants who provided valid data in at least the free play phase, resulting in a larger sample of 27 participants in the younger age group (n = 15 female; M = 236.81 days, SD = 14.94 days). All infants were born full term (M = 40.4 weeks; SD = 1.32 weeks). The primary caregiver participated in the free play phase of the study, that is, the person spending most time of the day with their child at the time of testing. Five fathers (younger sample: n = 2; older sample: n = 3) and 42 mothers (younger sample: n = 25; older sample: n = 17) participated in the free play phase of the study. All participants came from Leipzig (Germany) or surrounding areas, an urban Western, industrialized context. They were recruited on a voluntary basis via phone from a database of Max Planck Institute for Evolutionary Anthropology in Leipzig. We did not collect individual data regarding the participants' socioeconomic or ethnic background, but families in this database come from a predominantly white population with mixed, mainly mid to high socioeconomic backgrounds. Written informed consent was obtained from one parent of each infant prior to testing.

Stimuli and design

To investigate infants' attentional preference for third-party social interactions, we measured their looking times while they were simultaneously presented with two video clips: one social interaction stimulus and one non-interactive control stimulus. Both videos were presented without sound against a black background. The social interaction stimulus showed two women initially facing forward before they turned towards one another and engaged in one of three social

interactions while facing each other: playing an interactive clapping game, leaning towards one another, or touching their hands. The control stimulus showed the same two women facing forward before they turned away from one another, performing the identical movements as in the social interaction scene while standing back-to-back. All actors were female, wore white t-shirts, and were visible from the waist up. To avoid actors between stimuli being interpreted as interacting with one another and to maximize the visual distance between the two videos, the videos were positioned diagonally on the screen.

Every trial lasted 12 seconds (see Figure 2.1). Before each trial, an attention-grabbing sequence was presented in the center of the screen until the infant looked at it. Every participant saw twelve trials in a randomized order: Each of the three interactions (and the corresponding control video) were shown in four possible diagonal arrangements on the screen. All four trials within one interaction showed a different dyad. The video stimuli were created by using Adobe Premiere Pro. Although seemingly acting in dyads, the actors were filmed individually. The control stimuli were created by horizontally mirroring the actions of the individual actors. All actors were filmed in front of a green screen to control for color and luminance differences between and within videos. Each video covered an approximate area of 13.9° width × 7.8° height (at a screen distance of 60 cm). In the supplementary materials we provide detailed information regarding stimulus development. To measure infants' looking time, an SMI eye tracker (RED250mobile, SensoMotoric Instruments, 8.2) and SMI eye tracking computer programs (Experiment Center 3.7.60 and BeGaze 3.7.42) were used. Data were recorded separately for the left and the right eyes at a sampling frequency of 250 Hz.

To investigate infants' social attention during own social interaction, we coded their looking behavior during a five-minute free play phase with their parent. We placed three toys (two rattles and a rubber duck) within reaching distance between the infant and their parent.

Procedure

The testing took place at the Leipzig Research Center for Early Child Development (Leipzig University) between July and December 2017. Each testing session was divided into two phases: eye tracking (10 min) and free play (5 min).

During the eye tracking phase, the parents sat down in front of a screen, holding their child on their lap. We used a 25″ monitor with 117.5 dpi and 1920 × 1080 screen resolution. The parents were instructed to close their eyes or lower their gaze during the experiment, hold their child as still as possible, and avoid any kinds of communication. We used SMI five-point calibration to calibrate the eye tracker to the participant's eyes. To check the quality of the calibration, a manual calibration check was performed for each participant. Based on visual inspection, the experimenter evaluated the accuracy of each infant's gaze shifts, while they saw a colorful ball in the center of

the screen and in all four stimulus regions of the preferential-looking task. A participant was only included if providing valid gaze data according to this assessment.

Figure 2.1

Exemplary sequence of one experimental trial (clapping interaction) with the social interaction stimulus in the upper right corner and the control stimulus in the lower left corner
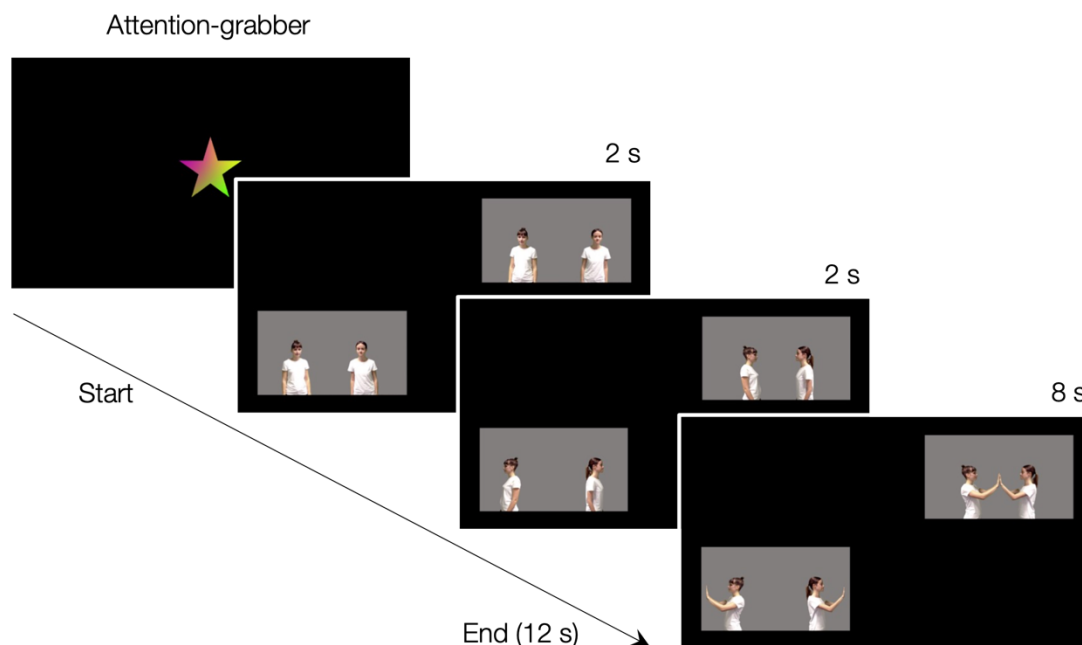
At the beginning of the free play phase, the experimenter instructed the parents to engage with their child and the toys in "normal play". The parents were further told not to touch the toys themselves during the first 90 seconds of play, to allow infants to actively initiate joint engagement (Bigelow, MacLean, & Proctor, 2004). A short notification sound indicated the end of the 90 second interval. We did not find any statistically relevant differences in infants' social engagement score from before to after the 90-second threshold (see supplementary materials for details). After instructing the parents, the experimenter left the room and came back after 5 minutes. All free play sessions were video-recorded.

Data Analyses and Coding

Attentional preference for others' social interactions. By using SMI BeGaze 3.7.42, we defined rectangular-shaped areas of interest (AOI) for the social interaction and the control stimulus. Each AOI covered an area of 15.8° width × 9.7° height (at a screen distance of 60 cm). To accommodate for inaccuracies in calibration, the AOIs were defined 1° visual angle larger than the maximal dimensions of the stimulus (Gredebäck, Johnson, & von Hofsten, 2009). In a second step, we calculated the total duration of fixations within the social and the control AOI for each individual trial. Data for both the left and the right eyes of each participant were averaged. We

included fixation data from the entire trial sequence. The results did not differ when including only the last 10 seconds of each trial (i.e., after the actors had started turning, see supplementary materials). To define the gaze events, we used the SMI BeGaze 3.7.42 high speed event detection filter. In a third step, we calculated the relative looking time at the social interaction stimulus for each individual trial:

$$\text{Proportional looking time at social stimulus} = \frac{\text{Cumulative length of fixations in social AOI}}{\text{Cumulative length of fixations in social AOI + control AOI}}$$

The score could take values between 0 and 1, with values above 0.50 indicating a relatively longer looking time at the social interaction stimulus. For statistical analyses, the proportion scores were averaged over all trials. A Shapiro-Wilk test of normality revealed that the proportion score was normally distributed (p = .72). A trial was excluded from the analysis if the participant did not look at the screen at all. To compare the averaged preference scores between the two age groups, we conducted a two-way analysis of variance (ANOVA) with age group as a between-subject factor. As some previous studies suggest gender differences in social attentional preferences in infancy (e.g., Lutchmaya & Baron-Cohen, 2002), we controlled for gender. To assess whether the proportion score significantly differed from chance level, we ran a one sample t-test (against .50) for both age groups.

Active social attention behavior. The occurrence of four infant looking behaviors was coded from video recordings of the free play sessions (see Figure 2.1). The reason for choosing these behaviors was to assess variability in different hierarchical levels of social attention (i.e., beginning with a very general social interest over face-to-face interactions up to joint attention looks). Note that the category "looking at the parent's face" was not included in the pre-registered coding-scheme. In aiming to get a more precise picture of infants' social attention behavior, we decided to differentiate general looks at the parent from looks at the parent's face after watching the recordings for the first time and prior to running any statistical analyses. In addition to the coding category "eye contact", infants' "looks at their parent's face" would consider situations in which infants made an attempt to engage in eye contact with their parent, without the parent looking back.

The coder watched every video recording in 5 second intervals (see also Hirshberg & Svejda, 1990). For each interval she decided if the infant showed one of the four looking behaviors. If none of the behaviors was shown, the infant received a "0" in the respective interval. If an infant showed one of the four behaviors at least once, they received a "1" in the respective category. Based on the hierarchical structure of the coding behaviors, each interval was coded with the highest occurring looking behavior during this interval. The primary coding was done by the first author. For inter-observer reliability, a second coder, naive to any hypothesis, coded a random 25% of the free play sessions after data collection was completed. The reliability coder was trained on a shared set of

videos prior to coding. The inter-rater agreement was good (*ICC* = .85). The coding of the free play sessions was conducted in Microsoft Excel. For statistical analyses, the following preparatory steps were taken for each individual. First, we calculated the frequency of occurrence of the four relevant behaviors over all coding intervals (i.e., the number of intervals during which a behavior was shown). Then, the total frequencies of the individual behaviors were integrated in the following proportion score:

$$\text{Social Engagement Proportion Score} = \frac{\text{Frequency of occurence of behaviors } 2+3+4}{\text{Frequency of occurence of behaviors } 1+2+3+4}$$

The score could take values between 0 and 1, whereby higher scores indicated greater levels of social interest. The specific equation for calculating the proportion score was not pre-registered prior to data collection. We based it on our observation that infants' "general looks at their parent" were mainly looks at toys in their parent's hand. As a consequence, infants' "general looks at their parent" (category 1) seemed to be strongly confounded by the parents' activity level. To include all pre-registered infant behaviors while extracting infants' "real" social looking behaviors from the overall number of coded behaviors, we relativized the sum of the higher-order social looking behaviors at the total amount of all coded behaviors for each individual infant. Our results remained stable when including the sum of frequencies of the behaviors "look at parent's face" (category 2), "eye contact" (category 3), and "joint attention" (category 4) without relativizing them at the total amount of behaviors (see supplementary materials).

A Shapiro-Wilk test of normality revealed that the social engagement score was normally distributed (*p* = .14). To compare the social engagement scores between the two age groups, we ran a two-way analysis of variance, controlling for gender. We explored the data further by running separate analyses for each of the four behaviors. Four Mann-Whitney-U-tests for independent samples were conducted to compare the mean frequency of occurrence of the behaviors between the two age groups.

Table 2.1

Infant looking behaviors coded during free play with their parent

	Infant looking behavior	Description
1.	General look at the parent	Infant looks at their parent (including looks at objects, if the parent holds it in their hands).
2.	Looking at the parent's face	Infant looks at the face of their parent but the parent does not look back.
3.	Eye contact between parent and infant	Infant and parent look at each other's eyes.
4.	Joint attention looks between parent, infant and an object	Before or after infant and parent look at each other's eyes they both look at the same object.

Relation between attentional preference for others' interactions and active social attention behavior

We correlated the proportional looking time at social interactions with the social behavior score by using Pearson's *r* correlation. In addition to the pre-registered plan, we calculated separate Pearson's *r* correlations for both age groups. All statistical tests were two-tailed with an alpha-level of .05, except the exploratory pair-wise comparisons of the four infant looking behaviors during free play (Bonferroni-corrected alpha = .0125). R software environment was used for processing and analyzing the data.

2.2.2 Results

Attentional preference for others' social interactions

We found no effects for gender, neither as main effect ($F(1,36) = 0.05$, $p = .83$, $\eta^2 = .001$) nor in interaction with age group ($F(1,36) = 2.40$, $p = .13$, $\eta^2 = .06$) and thus excluded gender from the following analyses. The mean proportion of looking time at social stimuli was significantly greater in the older compared to the younger sample ($F(1,38) = 7.50$, $p = .009$, $\eta^2 = .16$, Table 2.2a). Only infants in the older age group preferentially looked at the social interaction stimuli ($M = .54$, $SD = .07$; $t(19) = 2.38$, $p = .03$, $d = 0.53$), whereas infants in the younger age group did not show any preference ($M = .47$, $SD = .08$; $t(19) = -1.56$, $p = .13$, $d = 0.35$, Table 2.2a).

We ran the following analyses in addition to the pre-registered analysis to explore the data further. First, we repeated our main analysis after excluding trials in which infants exclusively looked at one stimulus, revealing the same pattern with even stronger effects (older sample: $M = .57$, $SD = .06$, $t(19) = 5.32$, $p < .001$, $d = 1.20$; younger sample: $M = .47$, $SD = .07$, $t(19) = -1.63$, $p = .12$, $d = 0.36$; difference between age groups: $F(1,38) = 21.11$, $p < .001$, $\eta^2 = .36$). The average number of trials discarded in this way per infant was 1.40 ($SD = 1.79$, total = 27) for the younger age group and 1.60 ($SD = 1.64$, total = 32) for the older age group. Secondly, we explored possible inter-trial variability over the course of the experiment. We did not find any effect of trial on infants' preference score, neither in interaction with age group ($\chi^2(1) = 1.12$, $p = .29$, estimate = -0.03, $SE = 0.03$), nor as overall main effect ($\chi^2(1) = 2.11$, $p = .15$, estimate = 0.01, $SE = 0.01$).

Active social attention behavior

We found no effects of gender, neither as main effect ($F(1,43) = 1.71$, $p = .20$, $\eta^2 = .03$) nor in interaction with age group ($F(1,43) = 1.02$, $p = .32$, $\eta^2 = .02$) and thus removed gender from the following analyses. Social engagement scores were significantly higher in the older age group ($M = .36$, $SD = .18$) compared to the younger age group ($M = .24$, $SD = .20$; $F(1,45) = 5.06$, $p = .03$, $\eta^2 = .10$, Figure 2.2b). Exploratory pair-wise tests regarding the mean frequency of occurrence of the separate looking behaviors revealed an age group difference in only joint attention looks, with

infants in the older group performing more joint attention looks compared to the younger age group (U = 154.5, p = .01, see Table 2.2). The proportional looking time at others' social interactions did not correlate with the social engagement scores—neither in the total sample (N = 40; r(38) = .15, p = .36), nor in both age groups separately (younger sample: r(18) = .30, p = .19; older sample: r(18) = −.34, p = .15).
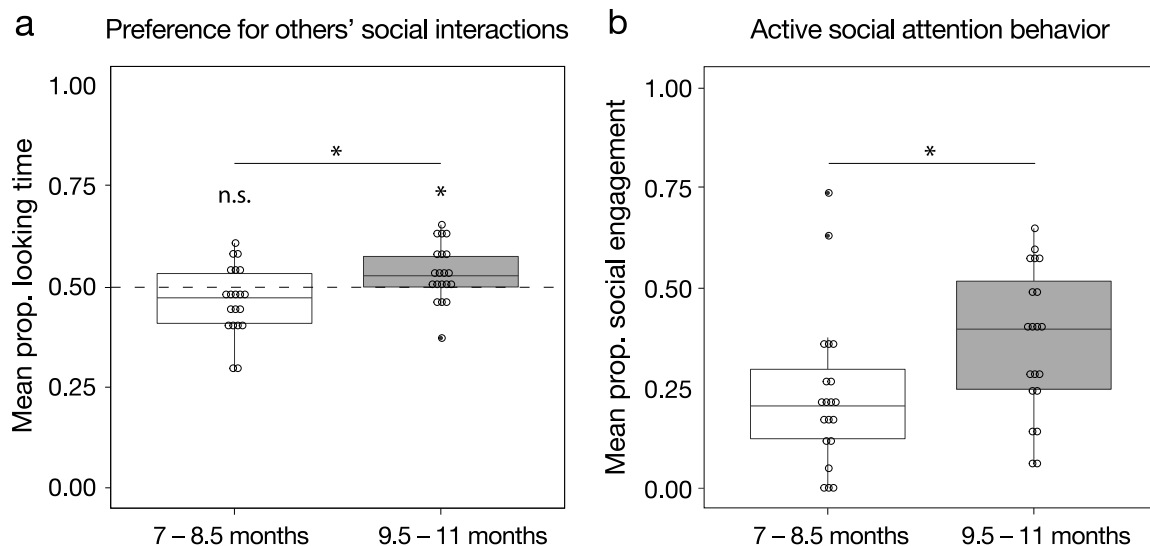
Table 2.2

Mann-Whitney-U-tests for the mean frequency of occurrence of the four coded infant behaviors in Experiment 1

Looking behavior	7 – 8.5 months Total N = 27		9.5 – 11 months Total N = 20		
	N	M *(SD)*	N	M *(SD)*	U
General look at parent	27	21.81 *(11.06)*	20	20.0 *(8.35)*	292
Look at parent's face	0	—	5	0.35 *(0.67)*	202.5
Eye contact	20	3.0 *(2.87)*	18	4.65 *(3.80)*	202
Joint attention look	19	3.41 *(4.33)*	20	6.05 *(4.30)*	154.5**

*Notes. **p = .011*

Figure 2.2

Boxplots with individual data points

a   Preference for others' social interactions

b   Active social attention behavior

Notes. (a) Mean proportional looking time to social interaction stimuli for both age groups tested against .50 and compared between age groups. The dashed line at .50 represents chance level. (b) Mean proportion score of active social engagement compared between age groups. n.s. = not significant. *p < .05

2.2.3 Discussion

We found an increase in infants' attentional preference for third-party social interactions from before to after 9 months of age. Infants at 9.5 to 11 months, but not at 7 to 8.5 months, showed a preference to watch others' social interactions over individual actions. Moreover, 9.5- to 11-month-olds showed more social looking behaviors during active social engagement compared to younger infants. At the individual level, infants' social attention behavior during own social interaction was not correlated with their attentional preference for others' interactions.

Together, these findings are in line with the idea that the age of 9 months represents a critical age in infants' social-motivational development (Tomasello, 1995). By comparing infants from two age groups close to before and after 9 months, we could demonstrate that infants do not only show an increasing interest in direct interaction partners (Carpenter, Nagell, et al., 1998) but also develop an increasing visual preference for others' social interactions. The developmental differences that we found are particularly relevant given the small difference of a minimum of 4 weeks between age groups. It remains unclear, however, *how* infants' social behavior and especially their attention toward others develop during the critical transition period—abruptly, or following a gradual and continuous increase (e.g, Striano & Bertin, 2005).

To gain a more comprehensive insight into developmental trajectories, we ran a second experiment using exactly the same tasks as in Experiment 1, but including infants at a broader and continuous range between 7 and 13.5 months of age. We aimed to test whether we could replicate our findings from Experiment 1 and extend them in three regards. First, by measuring age continuously and throughout the 9-month-period, we aimed to get an insight into the kind of transition taking place around 9 months. Moreover, by including infants up to 13.5 months of age, we aimed to broaden our understanding of developments after 11 months since infants' active social engagement has been previously found to continue increasing after 11 months (Adamson & Bakeman, 1985; Carpenter, Nagell, et al., 1998). Finally, we aimed to examine the non-significant correlation further by testing an additional and bigger sample.

2.3 Experiment II

2.3.1 Methods

The experimental design, procedure, as well as data pre-processing and coding procedures were identical to Experiment 1. The testing took place between July and October 2019 at the Max Planck Institute for Evolutionary Anthropology in Leipzig (Germany). To measure infants' looking time in the eye tracking task, we used a different SMI eye tracking hardware compared to the first experiment (RED-m, SensoMotoric Instruments, 8.2), recording data at a sampling frequency of 120 Hz. The eye tracking model did not have an effect on the results of the merged analyses (see

supplementary materials). Other deviations from Experiment 1 are described in the corresponding sections.

Participants

Fifty infants between 7 months, 0 days and 13 months, 13 days provided both eye tracking and free play data and were included in the correlation analysis (n = 21 female; M = 316.9 days, SD = 58.42 days). Another 24 infants were tested but excluded due to calibration error or technical failure during eye tracking (n = 14), technical failure during free play (n = 5), or because the infant did not remain in the camera field during free play (n = 5). For the separate analyses of the eye tracking and free play data, we included all infants who contributed valid data for either of the two measures. Accordingly, 51 infants were included in the eye tracking analyses (n = 21 female, M = 318.43 days, SD = 58.86 days), and 64 infants in the free play analyses (n = 30 female; M = 311.9 days, SD = 57.79 days). All infants were born full term (M = 40.09 weeks; SD = 1.40 weeks). The primary caregivers, that is, 5 fathers and 59 mothers participated in the free play phase of the study. Participants were partly recruited from the data base described in Experiment 1 (n = 39), and partly from the database of the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig (n = 35). The general sample characteristics and contexts are similar between the two data bases. The sampling plan was planned in the pre-registration and is available on the OSF. We ran additional analyses over a merged sample combining all participants from Experiment 1 and 2. A total of 90 infants between 7 months, 0 days and 13 months, 13 days (M = 299.17 days, SD = 54.27 days) were included in the overall correlation analysis, 91 infants in the overall eye tracking analysis (M = 300.22 days, SD = 54.89 days), and 111 infants in the overall free play analysis (M = 293.95 days, SD = 55.21 days).

Data analysis and coding

The coding of the free play sessions was identical to Experiment 1, except that a different second coder performed the inter-reliability coding. The inter-rater agreement between the first and second coder was good (ICC = .88). Shapiro-Wilk tests revealed that both dependent variables were normally distributed (p > .05). To investigate the effect of age on infants' visual preference for third-party social interactions, we ran a linear model for the mean proportional looking time to the social interaction videos in the eye tracking task, using age (in days) as continuous predictor. To investigate the effect of age on infants' social attention behavior during own social interaction, we ran a second linear model for infants' active social engagement score, including the same predictor as in the first model. To assess the relation between the two measures, we correlated the proportional looking time at the social interaction scenes with the social engagement score by using Pearson's r correlation. We did not include gender in any of our analysis, as we did not find

any effect of gender in Experiment 1. As planned in the pre-registration, we repeated all analyses over a merged sample combining participants from both experiments.

2.3.2 Results

Attentional preference for others' social interactions

The mean proportional looking time to the social interaction stimuli increased with age (Beta = .04 ± SE = .01, $t(1,49)$ = 3.73, $p < .001$, η^2 = .22, Figure 2.3a). In addition to the pre-registered analysis, we repeated our main analysis after excluding trials in which infants exclusively looked at one stimulus, revealing the same pattern (Beta = .04 ± SE = .01, $t(1,48)$ = 2.61, $p = .01$, η^2 = .12). The average number of trials discarded in this way per infant was 2.5 (SD = 2.2). We found the same pattern when repeating our analysis over a merged sample including participants from both Experiments (Beta = .04 ± SE = .01, $t(1,89)$ = 5.09, $p < .001$, η^2 = .23).

Figure 2.3

Scatterplots with individual data points including participants from both experiments
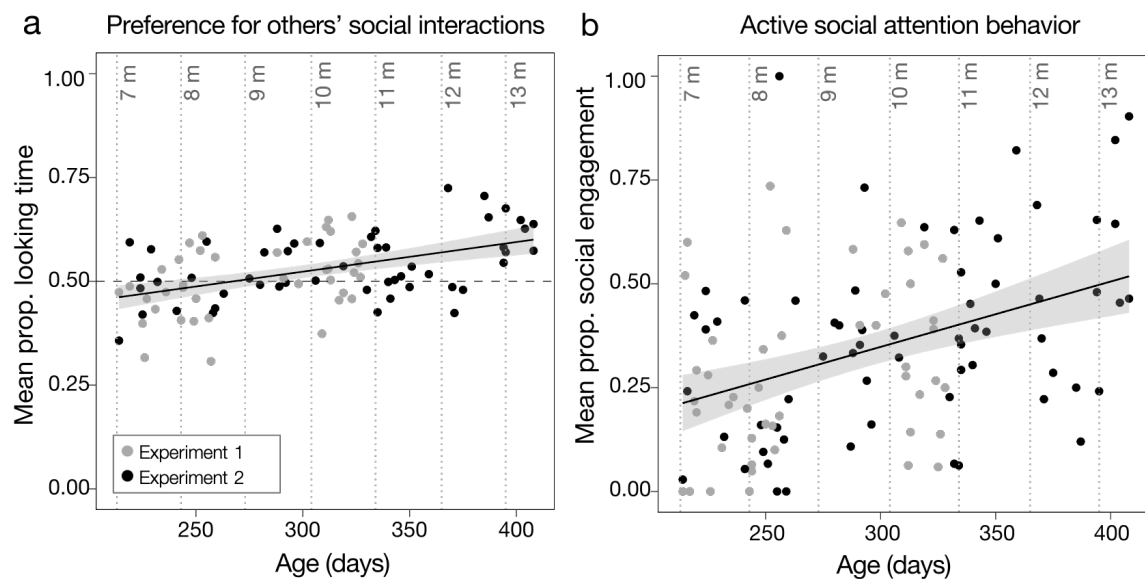
Notes. (a) Effect of age on mean proportional looking time to social interaction videos ($p < .001$). The dashed line at .50 represents chance level. (b) Effect of age on mean proportion score of active social engagement ($p < .001$). The vertically dashed lines indicate age in months. The linear regression lines with confidence ribbons fit to the overall data of the plots. The lower variance in infants' preference for social interactions compared to their active social behavior represents a methodological artefact, no systematic developmental difference between the constructs.

Active social attention behavior

The social engagement score increased with age both in the separate sample (Beta = .09 ± SE = .03, $t(1,62)$ = 3.35, $p < .001$, η^2 = .15) and in the merged sample (Beta = .09 ± SE = .02, $t(1,109)$ = 4.49, $p < .001$, η^2 = .16, Figure 3b). Exploratory analyses of the separate looking behaviors revealed that

infants with increasing age produced more joint attention looks (Beta = 2.80 ± SE = .85, $t(1,62)$ = 3.28, $p < .001$, η^2 = .15), and fewer general looks at their parent (Beta = −2.71 ± SE = 1.0, $t(1,62)$ = −2.71, $p = .01$, η^2 = .11, see Table 2.3a). Additional analyses including infants from both experiments revealed a similar pattern (see Table 2.3b). The proportional looking time at others' social interaction did not correlate with infant' social behavior scores at the individual level ($N = 50$; $r(48)$ = .23, $p = .11$).

Table 2.3

Results from exploratory linear models for the effect of age (days) on the absolute frequency of the four infant behaviors (a) in Experiment 2, and (b) for a merged sample including participants from Experiment 1 and 2

(a)	Experiment 2 (Total N = 64)				
Looking behavior	N	Beta	SE	t	p
General look at parent	63	−2.71	1.00	−2.71	.01**
Look at parent's face	8	−0.03	0.04	−0.71	.48
Eye contact	49	0.03	0.33	0.08	.93
Joint attention look	58	2.80	0.85	3.28	<.01***
(b)	Experiment 1 & 2 (Total N = 111)				
Looking behavior	N	Beta	SE	t	p
General look at parent	110	−2.20	0.84	−2.62	.01**
Look at parent's face	13	0.04	0.04	0.97	.33
Eye contact	87	0.06	0.29	0.22	.82
Joint attention look	97	2.93	0.56	5.20	.00***

Notes. N = Number of participants showing the behavior at all. Figure A3 in the supplementary materials provides visualizations of the data for all four infant behaviors. **$p < .01$ ***$p < .001$

Overall analysis of individual differences across both experiments

Probing the relation between infants' visual preference for others' interactions and their active social engagement score with a higher-powered analysis revealed a statistically relevant relation ($N = 90$; $r(88)$ = .24, $p = .03$, R^2 = .06, see Figure 2.4). These findings complement the pattern of group-level differences for both measures by showing an increase in social attention and behavior on the individual level. To explore the impact of age on this relation, we ran a linear model for the active social engagement score, including the interaction between proportional looking time to the social interaction stimuli and age (in days). The interaction did not reveal a significant effect (Beta = −.36 ± SE = .23, $t(3,86)$ = −1.61, $p = .11$, η^2 = .03) and was therefore dropped from the model. The same model including proportional looking time and age as main effects revealed a significant

effect of age (Beta = .08 ± *SE* = .02, *t*(2,87) = 3.54, *p* < .001, η^2 = .12), not proportional looking time (Beta = .13 ± *SE* = .28, *t*(2,87) = 0.48, *p* = .63, η^2 = .003).

Figure 2.4

Scatterplot illustrating the mean proportional looking time to social interaction stimuli plotted against the mean proportional social engagement score for a merged sample of Experiment 1 (N = 40) and Experiment 2 (N = 50)
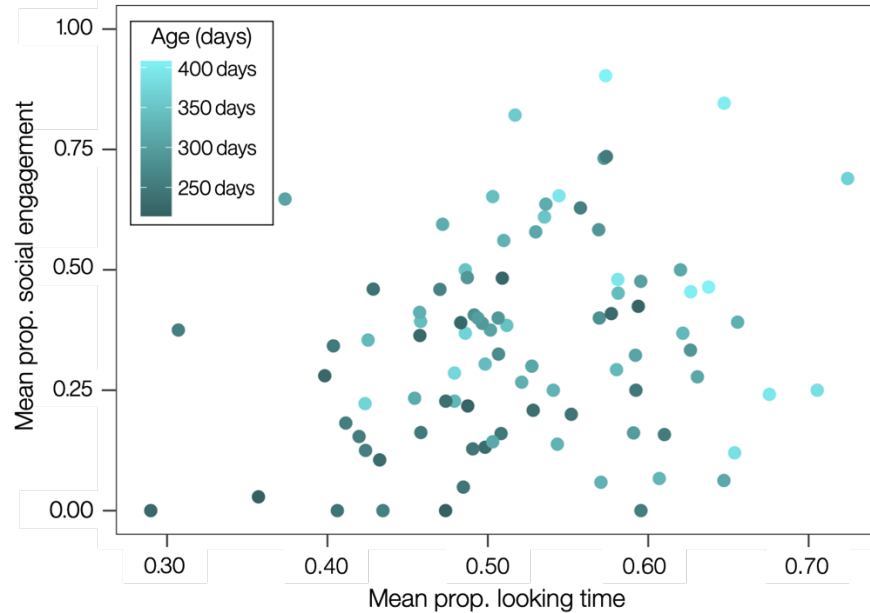
Notes. The dots represent individual data points of participants from both experiments. The color gradient represents the participant's age ranging from 7 months, 0 days (darkest color) to 13 months, 13 days (brightest color). In a merged sample over both experiments, the two measures were spuriously correlated through infants' age (*N* = 90, *p* = .03).

2.3.3 Discussion

We found a continuous increase in both infants' preferential orienting toward third-party social interactions and their social attention during active social interaction (especially joint attention looks). We found analogous patterns when repeating our analyses over a merged sample of Experiment 1 and 2. Infants' proportional looking time at others' interactions was not correlated with their social engagement scores in Experiment 2, but the two measures were correlated in a merged sample including infants from both experiments. Further analyses suggested that this correlational relation was predominantly driven by age, indicating that the two measures were not directly related at the individual level.

In contrast to the increase in infants' overall social engagement score and joint attention, infants' general looks at their parent decreased with age. This finding is in line with our observation, that infants' "general looks" were mainly looks at toys in their parent's hand (see "Data analysis and coding", Experiment 1). Based on our impressions during video coding before running statistical analyses, we speculate that a decrease in parent-toy interaction may have

caused the age-related decrease in infants' general looks at their parent. With increasing age of their child, parents appeared to make fewer attempts to engage their child with the toys as children began exploring the toys by themselves.

2.4 General Discussion

Previous work on infants' social attention did not assess whether infants' preferential orienting toward third-party social interactions coincides with their social attention during active social engagement, which undergoes a significant development during the second half of the first year of life. In Experiment 1 we found that, in contrast to 7- to 8.5-month-olds, older infants at 9.5 to 11 months of age (a) show an increasing preference to watch social interactions over individual actions, and (b) show a higher attentional interest in an interaction partner during active participation in social interaction. In Experiment 2, we could replicate this increase at both levels, and show that it develops in a continuous manner from 7 to 13 months of age. In a merged sample over both experiments, infants' orienting toward others' interactions was positively correlated with their social attention during own social engagement, but this correlation was mainly driven by infants' age. Our findings suggest that infants' social attention is driven toward social interactions toward the end of the first year of life.

The increase that we found in infants' active social attention, specifically their joint attention, aligns with prior work suggesting changes in infants' social interaction behavior towards the end of the first year of life (Carpenter, Nagell, et al., 1998; Striano & Reid, 2006). More specifically, the continuous trajectory supports previous studies suggesting that infants' social engagement skills develop gradually rather than changing abruptly from before to after 9 months of age (see also Bakeman & Adamson, 1984; Striano & Bertin, 2005). This is further in line with the assumption that triadic attention results from multiple continuous developments unfolding over time and in interaction with the environment, rather than being the result of one isolated emerging social skill causing a sudden change (de Barbaro, Johnson, & Deak, 2013). The age-related increase, that we found in both experiments, does not imply that younger infants did not show any joint attention behaviors at all. Indeed, we found infants at all ages, including 7-month-olds, engaging in at least one joint attention episode with their parent (97 out of 111 infants over both Experiments). This finding corresponds with previous studies demonstrating early joint attention behaviors emerging already before 9 months of age (Bakeman & Adamson, 1984; Striano & Bertin, 2005).

Our findings regarding infants' attentional preference for others' interactions extend previous studies by revealing a continuous increase throughout the second half of the first year of life. When probing infants' preferential looking score against chance level in Experiment 1, only infants older than 9 months showed a statistically relevant preference for the face-to-face

interaction videos (even though individual infants in the younger age group showed a preference as well). Our finding that older infants preferred attending to third-party interactions corresponds with prior work demonstrating a preference for face-to-face interactions in 9-month-old infants or older (Beier & Spelke, 2012; Galazka et al., 2014; Handl et al., 2013). Based on previous findings, it is rather unlikely that the absence of preferential orienting in 7- to 8.5-month-old infants resulted from a lacking ability to differentiate between the two scenarios. Already 6-month-old infants use others' body orientation to infer an interactive relationship between two people (Augusti et al., 2010). Accordingly, we suggest that the younger participants in the current studies did identify a difference between the two scenarios, but did not have a *preference* for one over the other scenario. Another possibility would be that our preferential-looking task was too demanding for the younger participants, as two videos were shown at the same time (in contrast to studies using a one-by-one stimulus presentation, e.g., Augusti et al., 2010). However, even if the higher complexity of our procedure had undermined the onset age of infants' above-chance preference for social interactions, the findings from both experiments point to an increasing orienting toward the end of the first postnatal year. During the same period, previous studies with geometrical shape agents have shown that infants develop representations of different kinds of third-party social relations, and that they use these representations to make inferences about the future. Seven-month-old infants represent affiliative relationships and expect social group members to perform similar actions (Powell & Spelke, 2013), 9-month-old infants use intergroup representations to make moral evaluations about others (Hamlin, Mahajan, Liberman, & Wynn, 2013), and 10-month-olds represent dominance-relationships and use this information to predict competition outcomes (Thomsen, Frankenhuis, Ingold-Smith, & Carey, 2011).

Comparing the developmental pathways of both modalities at the group level suggests that infants' attentional orienting to others' interactions indeed follows a similar increase as their active social attention. In addition, we found a spurious relation at the individual level in a merged sample over both experiments, in that infants with higher attentional preference for others' interactions showed more social attention behaviors during interaction with their parent. However and importantly, additional analyses revealed that this effect was driven by an underlying effect of age. While the current findings show that both modalities are related in terms of concurrent developmental trajectories, future studies will need to examine the specific underlying processes and mechanisms explaining this relation. It is likely that an interplay of multiple mechanisms is involved, as the absence of a correlational relation speaks against the notion that one single construct underlies the development of both social attention behaviors (see also Slaughter & McConnell, 2003).

One possible interpretation of our findings would be that social behavior and perception are both driven by motivational systems guiding infants to situations in which they can engage in or observe others' interactions. One specific mechanism could be an intrinsic motivation to acquire

knowledge and to learn from others (e.g., Litman, 2005). Even though our study did not focus on learning per se, our finding that infants increasingly engage in coordinated attention and increasingly prefer attending to face-to-face interactions raises the possibility that information-seeking motivations steer infants' attention toward situations in which they can gather knowledge. Both situations provide opportunities to acquire culturally relevant knowledge, including knowledge about content in the environment (e.g., information about novel objects, Csibra & Shamsudheen, 2015) or knowledge embodied in interpersonal interaction (e.g., coordinated action rituals, Legare & Nielsen, 2020). Another candidate mechanism could be infants' intrinsic social motivation (e.g., Chevallier et al., 2012). Previous studies have shown that infants find it intrinsically rewarding to engage in social interactions, and that this social motivation increases during the second half of the first year of life (Striano & Bertin, 2005; Venezia et al., 2004). Considering the possibility that infants find it also intrinsically valuable to observe others' interactions (e.g., Thiele et al., 2021), it would be possible that social reward-seeking mechanisms underlie the parallel increase in infants' attention to direct interaction partners and to others' interactions. Another factor that may influence infants' behavior at the broader level is a more general motivation to establish and foster social bonds with social group members. Early affiliative motives may not only modulate infants' behavior in direct interactions but also increase their sensitivity to social relations between others (for related evidence with 18-month-olds, see Over & Carpenter, 2009).

In addition to motivational mechanisms, it would be possible that the two modalities are causally or reciprocally related to one another. For example, infants' practical experience and active exploration in social interactions may influence the detection and understanding of others' interactions (Gredebäck & Melinder, 2010; Henderson, Wang, Matz, & Woodward, 2013). Vice versa, infants' experience and knowledge gained through observation of others may have an impact on their own behavior as well (Matheson et al., 2013). Importantly, this study was not designed to detect and disentangle such immediate relations between the two levels. This would require longitudinal study designs, together with a closer matching between infants' own natural interactions and the observed interactions, for example, by including touch as an interactive behavior, or by matching the knowledge that can be potentially learned from the interaction (e.g., object-related information, coordinated action rituals). Moreover, to assess infants' *understanding* of others' interactions, it would require different measures, such as predictive saccades (e.g., Bakker et al., 2011), or outcome measures of learning (e.g., object encoding, Cleveland & Striano, 2007; or manual actions, Matheson et al., 2013).

2.4.1 Limitations

The findings obtained from this study need to be considered against some limitations. First, since we investigated infants' social orienting capacities under controlled experimental conditions, the

videos depicted third-party interactions in a very simplified and in some sense restricted way. This was intended given that our primary goal was to match the videos from both conditions with regards to perceptual salience (e.g., motion, synchronicity, luminosity). Since the control scenes were created by mirroring the social interaction scenes, the rational meaning of the actions in the control videos was lower compared to the social interaction videos (e.g., performing the clapping movements without a social partner). Our findings suggest that infants with increasing age were not distracted by this issue, as they preferentially looked at the social interaction videos. For younger infants, however, we cannot rule out the possibility that they were distracted by the lower rational meaning in the control videos (Houston-Price & Nakai, 2004). Another difference between the videos was that only the social interaction scenes contained movement toward the center of the screen. Based on our data we cannot completely exclude the possibility that infants' looking behavior was influenced by a preference for perceptually grouped content. However, based on the finding by Galazka and colleagues (2014), that infants' preferential looking at social interactions disappeared in a control condition with inverted stimuli, we consider it rather unlikely that infants' looking pattern in the present study has been driven by low-level perceptual features. Future studies should systematically disentangle what visible features of social interactions underlie infants' visual preference (e.g., eye contact, face-to-face orientation, proximity, touch, rationality). Moreover, additional measures should be used to examine whether infants' attentional preference for others' social interactions is driven by affective-motivational mechanisms. Looking times alone do not provide direct information about motivational processes and should be complemented with measures of emotional arousal and valence, for example, by measuring infants' facial expressions (Steckler et al., 2018) or pupil dilation (Hepach & Westermann, 2016). In contrast to the current study design, this would require a one-by-one presentation order of stimuli. In addition, neuroimaging methods could complement the current findings regarding possible cortical specialization processes with regard to social interaction processing (Isik et al., 2017). Additionally and more generally, future studies are required to complement laboratory findings with infants' natural orienting in their everyday environment.

Another limitation is that we did not directly control for the impact of parental activity on infants' behavior during free play. To investigate reciprocal dependencies between interaction behaviors of infants and their parents, it would require a correspondingly detailed coding procedure considering the specific duration of behaviors and a setup with multiple cameras, allowing to record both interaction partners from different perspectives. In addition, a more advanced setup would allow to account for social attention behaviors going beyond the eye contact based behaviors measured in this study. Mobile eye tracking studies, for example, have demonstrated that one-year-olds' joint attention behaviors are not restricted to eye contact and gaze following. Infants and parents increasingly coordinate their attention by mutually following manual actions on objects, without necessarily looking at each other's eyes (Yu & Smith, 2013).

Another limitation is that our findings are restricted to interactions between infants and their primary caregiver. Since we did not investigate differences between mothers and fathers as primary caregivers, we cannot draw inferences regarding the influence of parental gender (Lewis et al., 2009). Moreover, our findings cannot not account for systematic differences in infants' behavior toward other interaction partners such as siblings (Teti, Bond, & Gibbs, 1988), peers (Bakeman & Adamson, 1984), or strangers (Dixon et al., 1981). In addition, we did not control for possible third factors that may have an effect on either of the two measures or mediate their relation. For example, an infant with an insecure attachment style may avoid eye contact with their mother during free play, but show preferential orienting to interactions between strangers (Claussen, Mundy, Mallik, & Willoughby, 2002). Other possible influential factors could be, for example, infant temperament (Todd & Dixon, 2010), motor ability and activity level (Clearfield et al., 2008), or own previous experience with an observed interaction (e.g., Gredebäck & Melinder, 2010). Furthermore, infants' developing receptive language abilities (Frank, Braginsky, & Marchman, 2021) may increase their interest in self-experienced and observed social turn taking. Moreover, given the great number of studies suggesting impairments in social attention and motivation in children with autism spectrum disorders (Chevallier et al., 2012), it would be important to investigate the visual preference for social interactions in a high-risk sample. Finally, our findings are restricted to typically developing infants growing up in a Western, industrialized context. Given the substantial variation in the extent to which children in different cultural contexts rely on direct pedagogy and observational learning (e.g., Callaghan et al., 2011; Mesman et al., 2018), it would be interesting to investigate cross-cultural differences in the development of infants orienting toward both situations. This would enable conclusions regarding the evolution of infants' attentional preference for social interactions.

2.4.2 Conclusions

In summary, we could show that infants' social behavior and attention are increasingly driven toward social interactions throughout the second half of the first year of life. From 7 to 13 months of age, infants do not only show increased active social engagement, but are additionally increasingly biased to attend to third-party interactions. Our findings suggest that infants develop capacities and preferences enabling them to approach social interactions through multiple pathways, including first-hand experience and third-party observation. This indicates that, toward the end of the first year of life, infants take an increasingly active role in maximizing the availability of situations in which they can potentially learn from others. Thus, at a broader level, infants' increasing orienting toward own and others' social interactions represents an important development on their way to becoming a competent member of their cultural community.

3

Study II

Social Interaction Targets Enhance 13-month-old Infants' Associative Learning

Abstract

Infants are attentive to third-party interactions, but the underlying mechanisms of this preference remain understudied. This study examined whether 13-month-old infants ($N = 32$) selectively learn cue-target associations guiding them to videos depicting a social interaction scene. In a visual learning task, two geometrical shapes were repeatedly paired with two kinds of target videos: two adults interacting with one another (social interaction) or the same adults acting individually (non-interactive control). Infants performed faster saccadic latencies and more predictive gaze shifts toward the cued target region during social interaction trials. These findings suggest that social interaction targets can serve as primary reinforcers in an associative learning task, supporting the view that infants find it intrinsically valuable to observe others' interactions.

Keywords: associative learning, social attention, infant development, social interaction

3.1 Introduction

Social interactions provide an essential source of learning opportunities for infants, through both active participation as well as observation of others' interactions (Paradise & Rogoff, 2009; Tomasello, 2016). To maximize such learning opportunities, infants are equipped with capacities and mechanisms that navigate them to social interactions (e.g., Reid & Striano, 2007).

Infants preferentially attend to communicative signals from birth. Newborns look longer at faces with opened than closed eyes (Batki et al., 2000), show enhanced neural processing of direct over averted gaze (Farroni et al., 2002), and orient toward infant-directed over adult-directed speech (Cooper & Aslin, 1990). In the second half of the first year of life, infants are also attentive to social interactions between others. Six-month-olds perform more saccadic gaze shifts in accordance with the reciprocal flow of a conversation when they see two people facing each other as opposed to people standing back-to-back while talking (Augusti et al., 2010), 9-month-olds look longer at face-to-face interactions when simultaneously presented with two people standing back-to-back (Handl et al., 2013; see also Beier & Spelke, 2012), 12-month-olds look longer at social over non-social turn-taking events (Bakker et al., 2011), and 14-month-olds look longer at biological motion of face-to-face interactions as compared to horizontally mirrored point-light displays (Galazka et al., 2014). Together, these findings suggest that, toward the end of the first year of life, typically developing infants are not only attentive to signals of direct interaction opportunities, but also to situations in which they can observe social interactions between others.

The emerging preference for others' social interactions is highly adaptive as it ultimately guides infants to potential observational learning opportunities. In their everyday life, however, infants' social attention and behavior is driven by proximal causes (Tamir & Hughes, 2018). One possible proximal driver of infants' preference for social interactions is the incentive value of social interactions (Anderson, 2016). Studies on active social engagement support the idea that direct social interactions are intrinsically valuable to typically developing infants. In the second half of the first year of life, infants show signs of seeking and liking social engagement (Striano & Bertin, 2005; Striano & Rochat, 1999; Venezia, Messinger, Thorp, & Mundy, 2004; Venezia Parlade et al., 2009). It remains unclear whether it is also intrinsically rewarding for infants to observe *others'* interactions. One way to examine this possibility is through reinforcement learning (Berridge et al., 2009). If it is rewarding for infants to observe third-party interactions, they should (based on reward learning) learn the association between an arbitrary shape cue and a target video more effectively if the target shows a social interaction as compared to a non-interactive control scene (Berridge & Robinson, 2003; Vernetti, Smith, & Senju, 2017; Tamir & Hughes, 2018). Previous research has shown that targets with high social-emotional value can enhance infants' associative learning and motivate behavior to acquire the valued stimulus. For example, when seeing a non-social cue (arbitrary shape) repeatedly preceding a target video showing the face of their mother,

7-month-old infants show faster decreasing saccadic latencies to the target region across trials compared to a target video showing the face of a stranger (Tummeltshammer et al., 2019). To what extent third-party social interaction targets can serve as primary reinforcer remains unclear.

In the current study, we investigated the incentive value of third-party social interactions in an associative visual learning paradigm. Infants saw one of two different non-social cues (circle or triangle) presented in the center of a screen, repeatedly paired with one of two target videos appearing left or right of the cue (target regions). One video showed two adults turning toward one another and engaging in a social interaction (touching hands, leaning toward one another). In contrast, the other video showed the same two adults performing identical movements as in the social interaction video while standing back-to-back (non-interactive control). To assess infants' learning performance, we measured their saccadic latencies from the central cue to the correct target region across trials and explored the occurrence of anticipatory gaze shifts (see also Reuter, Emberson, Romberg, & Lew-Williams, 2018; Wang et al., 2012).

Based on previous research, we expected that associative learning would be reflected in a decrease in saccadic latencies across trials. We hypothesized that if social interaction targets increase infants' learning performance in the visual learning task, infants' saccadic latencies should decrease relatively faster in the social interaction condition than in the non-interactive condition (interaction effect of condition and trial). To examine whether infants transferred the value from the social interaction target to the social interaction-predictive cue, we compared their proportional looking times at the cue before and after the visual learning task (preferential-looking task), and coded their first-touch behavior while presented with touchable plush versions of the cue shapes (manual forced-choice task). We assumed that if infants successfully learned the associative meaning of the cue shapes, and if they preferred the social interaction target, they should choose the social interaction predictive cue in both tasks. We tested infants between 13 and 14.5 months of age. Even though infants have been found to learn statistical regularities among central cues and peripheral targets at a younger age (e.g., Wu & Kirkham, 2010), we decided to test older infants based on piloting and because our paradigm was relatively complex compared to these studies.

3.2 Methods

We pre-registered the hypotheses, methods, procedures, and the data analysis plan on AsPredicted (https://aspredicted.org/zt975.pdf). As mentioned in the pre-registration, we assessed infants' gaze-following abilities in addition to our main research question to explore possible relations to their performance in the visual learning task. As there was no relation between infants' performance in gaze following and visual learning, we report the procedure, analyses, and results in the supplementary materials.

3.2.1 Participants

Thirty-two infants between 13 months, 0 days, and 14 months, 15 days were included in the final sample of the study (n = 16 female; M = 416.8 days, SD = 15.4 days). Data from 6 additional infants were excluded due to calibration error (n = 3) or because the infant did not complete the study (n = 3). We excluded 5 of the 32 infants from the manual forced-choice task because they did not look at both shapes before making a choice (n = 1), did not touch a shape within 2 min (n = 3), or because they touched both shapes at the same time (n = 1). All participants were born full-term. They were recruited from the database of the Uppsala Child and Baby Lab at Uppsala University and came from Uppsala (Sweden) or surrounding areas, an urban, industrialized context. We did not collect individual data regarding the participants' socioeconomic background, but families in this database typically come from mixed, mainly mid to high socioeconomic backgrounds, with the parents having a university degree. The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from both caregivers for each infant before any assessment or data collection. All procedures involving human subjects in this study were approved by the Uppsala Local Ethical Review Board at Uppsala University.

3.2.2 Procedure

The testing took place at the Uppsala Child and Baby Lab (Uppsala University) between October and November 2018. Each testing session started with a 10-min eye tracking phase, including the visual learning and preferential looking task (Video 1 in the online article shows the gaze replay of one exemplary infant participating in both tasks of the eye tracking phase). During the tasks, the infants sat in front of a screen on their parent's lap. We used a 23″ monitor with 96 dpi and 1920 × 1080 screen resolution with an integrated Tobii TX300 eye tracker (Tobii Technology, Stockholm, Sweden). We used a five-point calibration procedure to calibrate the eye tracker to the participant's eyes. The experiment was run by using E-Prime (version 3.0; Psychology Software Tools, Pittsburgh, PA, USA) and E-Prime Extension for Tobii (version 3.1), interfacing with the Tobii eye tracking hard- and software (Tobii Studio version 3.4.8.1348) via TET and Clearview PackageCalls. We used E-Prime for randomization and counterbalancing. Data were recorded separately for the left and the right eyes at a sampling frequency of 120 Hz. Following the eye tracking phase, we conducted the manual forced-choice task. The procedure is described in the supplementary materials. All manual choice sessions were video recorded.

3.2.3 Stimuli and Design
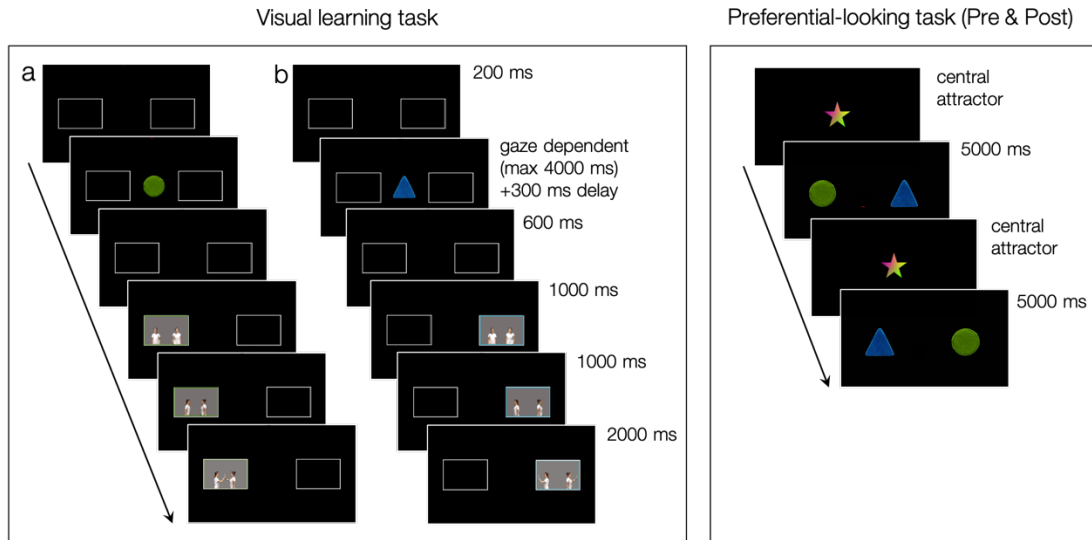
Visual learning task

Every participant saw a total number of 24 gaze-contingent trials (12 per condition). On each trial, one of two non-social cues (blue triangle or green circle) appeared between two white frames in the center of a black screen. Once the infant fixated on the cue for 150 ms, a "ping" sound appeared, and the cue remained on screen for another 300 ms before it disappeared (Video 2 in the online article shows four exemplary trials with sound). After a 600 ms delay, one of two different target stimuli appeared within one of the two white frames (4000 ms): a video depicting a social interaction scene or a non-interactive control scene. The social interaction video showed two women facing forward before they turned toward one another and engaged in a social interaction (leaning towards one another or touching hands). The control video showed the same two women turning away from one another, performing the identical movements as in the social interaction video while standing back-to-back. The rule of target appearance (right or left of the cue) remained consistent for both videos throughout the experiment. The videos were framed in the color of the corresponding cue shape to highlight the associative relation between cue and target. If the infant did not fixate on the cue within two seconds, it disappeared before returning for another two seconds. If the infant did not look at the cue during this second two-second-interval, no target video was displayed, and the next trial began (see also Tummeltshammer et al., 2014). The timing of the learning task is illustrated in Figure 3.1 and explained in detail in Table B1 in the supplementary materials. Infants were presented with a four-second kaleidoscope video every four trials to maintain their attention (see also Reuter et al., 2018). We counterbalanced the location of the social interaction target (right or left from the central cue), the shape cueing the social interaction target (triangle or circle), and the order of the first two trials (social interaction first or non-interactive control first) across participants. The order of the remaining trials was pseudo-randomized, with the same cue never appearing more than two times in a row. Each video covered an approximate area of 13.1° width × 8.7° height (at a screen distance of 60 cm). The cue covered an area of 5° × 5°. The distance between the center of the cue and the outer edge of the target AOIs was 6.7°.

Preferential-looking task

Before and after the visual learning task, we assessed infants' visual preference by presenting the cue shapes side-by-side on the screen (see Figure 3.1). The shapes were shown in both possible left-right arrangements during two successive trials (at 5 seconds each). The position of the shapes on the first trial was counterbalanced across participants.

Figure 3.1

Exemplary sequence and timing of the visual learning task and the preferential-looking task (presented pre- and post-learning)

Notes. The visual learning task shows two exemplary trials with (a) the circle cueing the social interaction video on the left side, and (b) the triangle cueing the non-interactive control video on the right side (touching hands interaction). The target onset was at 900 ms after the infant had fixated the cue (300 ms delay plus 600 ms gap).

Manual forced-choice task

In addition to the screen-based preferential-looking task, we assessed infants' manual choice preference for touchable plush versions of the cue shapes, by coding their first touch behavior when presented with both shapes at the same time (see supplementary materials for procedural details). The left-right positioning of the shapes on the choice board was counterbalanced across participants.

3.2.4 Coding and Data Analysis

The following pre-processing and analysis of the data was planned in the pre-registration. All additional analyses are explicitly labeled as exploratory. We used the Tobii Velocity-Threshold Identification (I-VT) fixation classification filter to define fixations and saccades. Data for both the left and the right eyes of each participant were averaged. When one eye could not be measured, we used the data from the other eye. Blinks and saccades were excluded for all measures of looking times. All areas of interest (AOIs) were defined as 1° visual angle larger than the maximal dimensions of the stimulus (Gredebäck et al., 2009). We used R software environment (RStudio version 1.2.1335) for setting AOIs, as well as for pre-processing and analyzing the data. All general linear mixed models (GLMMs) were conducted using R package "lme4" (Bates, Maechler, Bolker, & Walker, 2020). All R scripts and the eye tracking raw data are openly accessible on the Open Science Framework (https://osf.io/4a9b6/).

Visual learning task

We defined three areas of interest: one square-shaped AOI covering the central cue area and two rectangular-shaped AOIs covering the target regions. We assessed the saccadic latencies toward the cued region for each trial and participant. Saccadic latency was defined as the time difference between the first fixation arriving within the cue AOI and the first fixation arriving within the correct target AOI. As pre-registered, we excluded a trial from the analysis if (a) the infant did not look at the cue during four seconds, (b) the infant's gaze did not arrive in the correct AOI until video offset, (c) the saccadic latency was longer than 2000 ms (assuming that the infant had looked away from the screen after looking at the cue), or (d) the saccadic latency deviated more than $+/- 3$ SD from the individual mean saccadic latency within conditions. After excluding trials due to these criteria, all participants contributed at least 8 valid social interaction trials ($M = 11.03$, $SD = 1.20$) and 8 control trials ($M = 11.03$, $SD = 1.12$) to the final dataset. To compare the change in saccadic latency to the cued target region between conditions, we conducted a GLMM (Gaussian error distribution) for saccadic latency, including the interaction between trial (24 trials) and condition (social interaction, non-interactive control) as a fixed effect. As random effects, we included subject as an intercept, as well as random slopes for trial on subject, condition, and the interaction between trial and condition. P-values for the individual fixed effects were based on likelihood ratio tests comparing the full models with respective reduced models using the *drop1*-function in R with an alpha-level of .05.

We ran four analyses in addition to the pre-registered plan. First, following visual inspection of the average latency data, we explored condition differences in infants' saccadic latencies on the first trial. Second, we repeated our main analysis for saccadic latency over a subset of data including the first six trials only, to account for the possibility that infants had habituated to the target videos throughout the second half of the task (see analysis of looking times in section "Additional Analysis and Results" in the supplementary materials). Third, we compared the number of anticipatory gaze shifts between conditions. We assumed that if infants selectively learned about the appearance of the social interaction target video based on the associative meaning of the corresponding cue shape, they should show more anticipatory gaze shifts in the social as compared to the non-social condition. We defined a gaze shift as anticipatory when it was initiated before target onset. To account for the processing lag of the infant oculomotor system (Gredebäck et al., 2009), we expanded the actual time of target onset (900 ms after the infant had fixated the cue) by one-year-olds' reactive saccade latency (300 ms in one-year-olds; Reznick, Chawarska, & Betts, 2000; see also Canfield et al., 1997). The resulting threshold of 1200 ms was used to compute the number of trials for each infant during which they had performed a predictive gaze shift (latencies < 1200 ms), and the number of trials during which they had performed a reactive gaze shift (latencies > 1200 ms). To compare the number of anticipatory eye movements

to the cued target region between conditions, we conducted a GLMM for predictive and reactive gaze shifts (binomial error structure), including condition as fixed effect, subject as random intercept, as well as a random slope of trial on subject. Fourth, we explored infants' first look pattern to rule out that infants sought the social interaction target *region* rather than learning and responding to the meaning of the cue. We assumed that if infants preferred the social interaction target *region*, they should always look at this region first, independent of the meaning of the cue. For the analysis, we calculated the mean proportional number of first looks to the correct target region by dividing the number of first looks to the cued target AOI by the total number of first looks to both AOIs.

Preferential-looking task

To assess infants' looking preferences, we calculated the proportional looking time at the social interaction cue by dividing the duration of fixations on the social interaction cue shape by the total duration of fixations on both shapes. The resulting proportion scores were averaged over both five-second trials. We compared pre- and post-responses to the shape by using a paired t-test and ran two one-sample t-tests against .50 to determine whether the relative looking time to the social interaction shape differed from chance level before and after the learning task. We explored the post-test shape preferences further by comparing the mean proportional looking times between infants who showed enhanced learning in the social interaction condition and less enhanced learners. We divided the sample based on a median split of a latency difference score. The score was calculated for each individual by subtracting the mean saccadic latencies during control trials from latencies during social interaction trials (Tummeltshammer et al., 2019). In the supplementary materials, we report the pre-registered exploratory analysis based on a median split of a difference score from each individual's learning function beta-coefficients. We decided to use the mean difference score since the descriptive pattern of infants' saccadic latencies suggested a non-linear learning curve.

Manual forced-choice task

We assessed infants' choice behavior by coding which of the two shapes they touched first. A valid choice required that the infant had looked at both shapes and at the experimenter before or immediately preceding the touch. Moreover, choices were coded as invalid if the infant touched both shapes at the same time. We conducted a binomial test to determine the participant's choice. A second, naive coder coded a random 25% of the manual choice sessions (Cohen's kappa = 1).

3.3 Results

3.3.1 Visual Learning Task

Infants' learning between the two conditions did not change over time ($\chi^2(1)$ = .87, p = .35, estimate = 18.42, SE = 20.07). Instead and overall, infants looked faster to the location of the social target videos (M = 1107.35 ms, SD = 331.68) compared to the non-interactive control videos (M = 1284.79 ms, SD = 198.04; $\chi^2(1)$ = 9.53, p = .002, estimate = −172.18, SE = 52.50; see Figure 3.2). We did not find a main effect of trial ($\chi^2(1)$ = .08, p = .78, estimate = 2.35, SE = 8.48).

Figure 3.2

Change in saccadic latencies over social interaction trials and non-interactive control trials
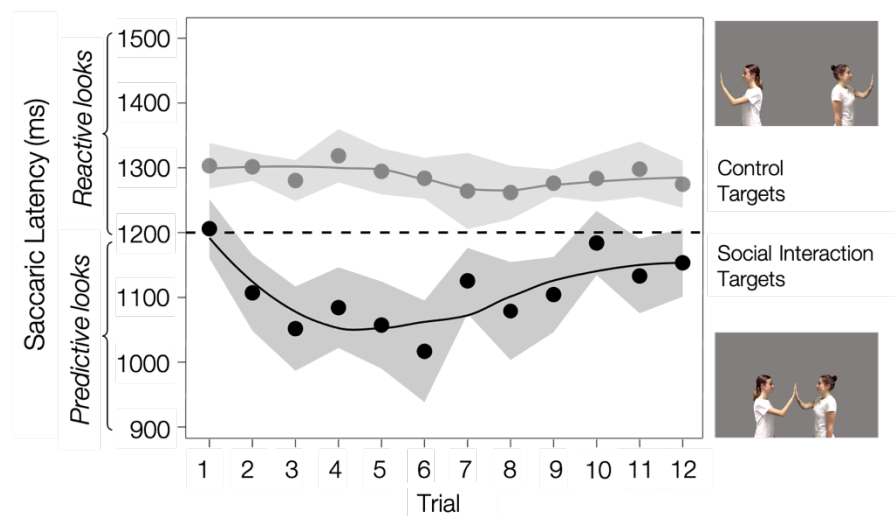
Notes. The dots represent the means overall individuals for each trial, the shaded areas the standard errors. The smooth curves loess fits to the data of the plot. The dashed line at 1200 ms represents the threshold for prediction, calculated by expanding the timepoint of target onset (900 ms after the infant had fixated the cue) by one-year-olds' reactive saccade latency (300 ms). Values below this line correspond to gaze shifts initiated before target onset, values above this line to gaze shifts initiated after target onset.

Exploratory analyses revealed that the main effect of condition was not present on the first trial ($\chi^2(1)$ = 3.47, p = .06, estimate = −97.94, SE = 51.88). Moreover, in a subset including data from the first six trials only, the interaction between condition and trial revealed a significant effect on infants' saccadic latency ($\chi^2(1)$ = 4.65, p = .03, estimate = −89.72, SE = 41.58). Overall, infants performed more predictive eye-movements during social interaction trials (mean proportion = .45, SD = .35) compared to control trials (mean proportion = .18, SD = .20; $\chi^2(1)$ = 8.78, p = .003, estimate = 1.90, SE = .62). In addition, exploratory analyses of infants' first looks indicated that infants learned the association between cue shape and target video rather than seeking the target region. Two one-sample t-tests revealed that the proportional number of first looks at the correct target region was greater than chance in both conditions (social interaction condition: M = .72, SD = .32; $t(31)$ = 3.90, p = <.001, d = .69; control condition: M = .72, SD = .31; $t(31)$ = 4.14, p = <.001, d = .73), with no difference between conditions (paired t-test, $t(31)$ = −.02, p = .98, d = .006). Table B2

in the supplementary materials shows the total number of first looks at the two target regions for both conditions.

3.3.2 Preferential-Looking Task

The mean proportional looking time at the social interaction shape did not differ from chance level—neither before ($M = .48$, $SD = .14$; $t(30) = -.65$; $p = .52$, $d = -.12$), nor after the learning task ($M = .46$, $SD = .15$; $t(31) = -1.58$; $p = .12$, $d = -.28$). There was no difference between pre- and post-test ($t(30) = 1.18$; $p = .25$, $d = .19$). The group comparison revealed that infants with enhanced performance in the learning task looked relatively longer at the social interaction shape ($M = .52$; $SD = .12$) compared to less enhanced learners ($M = .39$; $SD = .17$, $t(27) = 2.33$; $p = .03$, $d = .82$). However, the mean proportion score of the enhanced learners did not exceed chance level ($t(15) = .56$; $p = .58$, $d = .14$).

3.3.3 Manual Forced-Choice Task

Infants did not prefer one shape over the other. Thirteen out of 27 infants (48%) touched the social shape first ($p = 1$).

3.4 Discussion

Previous work showed that infants are attentive to third-party social interactions. The present study extends this finding by revealing that this bias goes beyond a preferential orienting in the here and now. In a visual learning task, we found that 13-month-old infants learned a cue-target association guiding them to videos showing two people engaging in social interactions. In contrast, we did not find such a learning effect for target videos displaying a non-interactive control scene. Our findings suggest an early emerging motivation in infants to recognize and seek out opportunities to observe others' social interactions.

Infants' learning of the interaction-predictive association was manifested in relatively faster latencies and more predictive gaze shifts in the social interaction (versus control) condition. In contrast to our hypothesis (i.e., decreasing saccadic latencies across trials), we found no effect of trial in our overall data. However, based on additional analyses, the current data nevertheless indicate that selective learning took place in the social interaction condition. First, the main effect of condition on latency was not present on the first trial, suggesting that it emerged during the learning task. Moreover, the faster latencies to the social interaction target region could not be explained by a general seeking of the interaction target *region* across conditions, suggesting that infants learned to anticipate the social interaction target based on the associative meaning of the cue. Additional support for learning of the interaction predictive cue-target association comes from our finding that infants showed more predictive gaze shifts toward the correct target region

in the social interaction condition compared to the non-interactive control condition. Visual inspections of the average latencies (see Figure 3.2) revealed that infants discovered the cue-target contingency rapidly, as the latencies to the social interaction target region decreased within the first few trials (see also Wang et al., 2012). Furthermore, the pattern of results raises the possibility that infants had habituated to the target videos throughout the second half of the learning task (see supplementary materials for supporting analyses). Considering this possibility, we carried out our main analysis for the first half of the learning task, revealing support for the idea that infants' learning between the two conditions changed over time during the first six trials. In contrast to the social interaction condition, infants' latencies remained unchanged at a reactive level throughout the control trials. The suggested absence of learning in the non-interactive condition contrasts with previous findings that already 6-month-olds are sensitive to statistical regularities in the visual domain (Tummeltshammer et al., 2014; Wang et al., 2012). One possible explanation for the discrepancy is that infants' motivation to watch the social interaction scenes had a suppressing effect on their responsiveness towards the control target, as the non-interactive scene was less meaningful to them (e.g., regarding social salience, potential learning opportunities, or intrinsic value). Given the similarities between the videos in both conditions (e.g., presence of two adults, equal amount and synchronicity of motion), the absence of learning in the control condition suggests rather that infants invested their resources selectively in favor of detecting and approaching the more meaningful interactive scenarios.

We did not find any looking preferences for the social interaction predictive cue shape itself, neither before nor after the learning phase (cf. Tummeltshammer et al., 2014). Even though enhanced learners showed higher proportions of looking times to the social interaction cue than less enhanced learners, their mean proportional looking time did not exceed chance level. One possible methodological explanation for the absence of a looking preference is that we presented our cues one-by-one in the center of the screen during the learning trials, but simultaneously and side-by-side during the preferential-looking task (in contrast to studies presenting the cues at identical positions in both phases, e.g., Tummeltshammer et al., 2014). Given this novel arrangement, infants may have needed more time to express a preference in their looking behavior. Thus, it is possible that the social interaction predictive cue had acquired value through reward learning, even if this was not reflected in infants' looking time at the cue in the post-test. Support for this assumption comes from the previously mentioned study by Tummeltshammer and colleagues (2018). Similar to our findings, infants in this study did not show increasing looking times to a reward-predictive cue from pre- to post-test. Infants' pupil size, however, increased in response to the cue, suggesting that the cue had acquired value without modulating infants' looking time. In contrast to the current study design, measuring such pupillary response would require a one-by-one presentation of the cues in the pre- and post-test. Similar to our findings in the preferential-looking task, we did not find infants to show a touching preference in the manual

choice task, possibly because they did not transfer the meaning of the cue shapes on the screen to the plush versions off-screen.

Together, our findings contribute to the idea that infants find it intrinsically rewarding to observe others' social interactions. Without receiving any external reward, infants recognized, learned, and responded to the associative meaning of an arbitrary shape cue allowing them to anticipate and approach third-party interaction scenes. The decrease in saccadic latencies toward the social interaction targets suggests that the infant reward system labels social interactions as subjectively valuable, causing superior associative learning and "reward-seeking" behavior. Both reinforcement learning and future-directed seeking have previously been considered indicators of social reward (Berridge & Robinson, 2003; Chevallier et al., 2012; Haith, Hazan, & Goodman, 1988; Vernetti et al., 2017). Based on these results, our findings extend the previous research on social reward by suggesting that not only direct social interactions but also third-party interactions have the potential to serve as a proximal motivator, increasing infants' attention, learning, and memory capacities (Tamir & Hughes, 2018). To gain a comprehensive understanding of possible reward mechanisms, future studies should investigate the affective component of reward. Additional measures such as pupil dilation or coding of facial expression could be used to measure infants' hedonic response to third-party interactions (Ariel & Castel, 2014; Hirshberg & Svejda, 1990; Tummeltshammer et al., 2019).

The enhanced learning performance for social interaction targets that we found in the visual learning task also aligns with prior work showing that attention-grabbing social cues modulate infants' responsivity (de Bordes et al., 2013). Following this interpretation, the mere prospect of observing a social interaction scene may have put infants in a state of heightened responsiveness, increasing their readiness to identify and learn the interaction-predictive cue-target association. Future studies are required to examine whether this increased responsiveness also promotes infants' processing of learnable content presented within the context of an observed social interaction (e.g., Cleveland & Striano, 2007). Importantly, the focus of this study was on processes *guiding* infants to situations in which they can observe others' interactions. We did not focus on processes enhancing infants' learning during the actual observation itself. Therefore, the logic of our study design was reversed (i.e., non-social stimulus cueing a social target) compared to previous studies investigating how infants use other people as social cue aiding them in detecting and learning about relevant content in their environment (e.g., Tummeltshammer, Wu, & Kirkham, 2013; Wu & Kirkham, 2011). In direct interactions with others, communicative signals such as direct gaze increase infants' ability to follow referential cues (Del Bianco et al., 2019; Senju & Csibra, 2008), support infants' learning from other novel attention cues (Wu, Tummeltshammer, Gliga, & Kirkham, 2014), and facilitate their encoding of cued target objects (Michel, Wronski, Pauen, Daum, & Hoehl, 2019; Parise, Reid, Stets, & Striano, 2008). Moreover, joint attentional engagement with others facilitates 9-month-olds' object processing (Cleveland & Striano, 2007),

as well as 18-month-olds' action imitation (Nielsen, 2006) and word learning (Hirotani et al., 2009). It would be a crucial next step to investigate whether similar factors (e.g., third-party ostension) organize infants' attention and increase their learning during ongoing observation of others' interactions (for related studies with older children see, e.g., Fitch et al., 2020; Gräfenhain et al., 2009). In addition, considering the complexity and diversity of human interactions, future studies are required to assess which specific features of social interactions have an impact on infants' visual preference (e.g., proximity, touch, mutual gaze, face-to-face orientation). Given the previous research on impairments in social attention and social motivation in children with autism spectrum disorders (Chawarska & Shic, 2009; Vivanti, 2017; Chevallier et al., 2012), it would be highly relevant to investigate the incentive value of third-party interactions in a high-risk sample. Moreover, to get a more comprehensive insight into developmental trajectories, it would be important to test infants longitudinally at different ages and from different cultural backgrounds (Nielsen & Haun, 2016).

In summary, we could show that 13-month-olds' attention and learning is biased toward situations in which they can observe others' interactions. Our findings extend previous research on infants' preferential orienting to third-party interactions by showing that this preference goes beyond currently available situations. We could demonstrate that infants can detect and use initially meaningless cues in their environment to predict future opportunities to observe third-party interactions. At a broader level, this finding has significant implications for the understanding of how infants learn about their world. Given the importance of third-party social interactive settings for early learning, infants' ability to detect, anticipate, and approach social interactions even if they are not immediately visible can serve as an adaptive goal as it provides infants with possible learning opportunities.

4

Study III

Observing Others' Joint Attention Increases 9-Month-Old Infants' Object Encoding

Abstract

In direct interactions with others, 9-month-old infants' learning about objects is facilitated when the interaction partner addresses the infant through eye contact before looking toward an object. In this study we investigated whether similar factors promote infants' observational learning from third-party interactions. In Experiment 1, 9-month-old typically developing infants from mixed socioeconomic backgrounds from urban Germany (N = 32, 13 female) were presented with four types of videos showing one object and two adults. The scenarios varied systematically regarding the eye contact between the adults (eye contact or no eye contact), and the adults' object-directed gaze (looking toward or away from the object). To assess infants' encoding performance we measured their looking times when seeing the familiarized object subsequently next to a novel object, interpreting an enhanced novelty preference as reversely indicating greater encoding of the familiarized object. Infants showed an increased novelty preference, but only after observing a joint attentional setting during which two adults attended to the familiarized object together (following eye contact). In Experiment 2, we found an identical pattern of results in a matched first-party design in which 9-month-old infants (N = 32, 16 female) were directly addressed by one single adult on screen. Infants' encoding was only enhanced when the adult made eye contact with the infant before looking at an object. Together, this suggests that the capacity to learn through observing others' interactions emerges already in the first postnatal year, and that it may depend on similar factors as infants' learning through direct social engagement.

Keywords: object encoding, social attention, social interaction, infancy, ostension

4.1 Introduction

Social interactions represent an essential source of learning opportunities in infancy. Both active participation in interactions and the observation of others' interactions help infants to acquire knowledge about their environment (Paradise & Rogoff, 2009; Rogoff et al., 2003). During the second half of the first postnatal year, infants' social behavior undergoes a crucial development as they begin to engage in triadic interactions. The transition from purely interpersonal dyadic (infant-adult) interactions to triadic (infant-adult-object) interactions represents a milestone in the first year of life, because it enables infants to incorporate external objects in their interactions and thereby communicate and learn about their environment (Tomasello et al., 2005). In triadic interactions with others, infants' learning about novel objects is strongly promoted when they see their interaction partner gazing in their direction before moving their visual attention toward an object (in the following referred to as "object-directed gaze"). Both factors are vital components of joint attention interactions during which infant and adult coordinate their attention to an object of mutual interest (Carpenter, Nagell, et al., 1998). Underlying factors of infants' observational learning from *others'* interactions remain less clear. In the current study we investigated whether infants' learning through observation of others' interactions depends on similar factors as infants' learning in direct interactions.

Infants begin to engage in coordinated joint attention with others between 6 and 9 months of age (Carpenter, Nagell, et al., 1998; Striano & Bertin, 2005). In contrast to triadic parallel attention episodes in which infant and adult look to an object individually, triadic joint attention episodes include mutual awareness of the shared attentional experience. Early joint attention behavior is typically defined as the alternation of eye-gaze between two interaction partners and an object of mutual interest (e.g., Bakeman & Adamson, 1984). Breaking it down into its separate components, the joint attentional triangle includes three types of communicative looks (Carpenter & Liebal, 2011): An initiation look during which the initiator looks in the direction of the recipient to make eye contact, a reference look during which both partners gaze toward the object (object-directed gaze), and a sharing look during which both partners re-engage in eye contact, indicating their mutual awareness of the joint attentional experience.

During the same period as infants develop a joint attentional awareness, they begin to make functional use of joint attention in terms of learning about new objects. Studies with real-interactive settings have revealed that 7- and 9-month-old infants show increased object encoding when they are familiarized with a novel object in a joint attention situation as compared to a non-interactive control condition (Cleveland et al., 2007; Cleveland & Striano, 2007). In the joint attention scenario, an experimenter established initial eye contact with the infant before alternating their gaze between a visible object and the infant. In the non-interactive "object-only" condition, the experimenter alternated their gaze between the object and a spot at the ceiling

without looking in the direction of the infant at any point. As a measure of encoding, previous studies have used infants' response to novelty (Fantz, 1964; Rose et al., 1982). When seeing the familiarized object next to a novel object in a subsequent paired-preference test, 7- and 9-month-old infants (but not 4- and 5-month-olds) looked relatively longer to the novel object in the joint attention condition as compared to the object-only condition (Cleveland et al., 2007; Cleveland & Striano, 2007; Striano et al., 2006). Following the logic of the paradigm, this novelty preference reversely indicates that the familiarized object had been previously encoded, causing higher attention to the novel, yet unprocessed object (for a review on infant visual attention and object recognition see Reynolds, 2015). In line with these findings, studies with slightly older children have shown that joint attention represents an important role (not a fundamental requirement, Scofield & Behrend, 2011; Tomasello, Strosberg, & Akhtar, 1996) for successful word learning around 2 years of age (Baldwin, 1995; Hirotani et al., 2009; Tomasello & Farrar, 1986; see also Scofield & Behrend, 2011).

In focusing more specifically on initial eye contact and object-directed gaze as two components of the joint attentional triangle, screen-based eye tracking studies have revealed that the *interplay* between both factors has an important impact on infants' object encoding. Already at the early age of 4 months, infants show increased encoding of an object if they have seen an adult shifting their gaze toward this object. This is indicated by an increased response to a previously non-cued object (increased looking times in behavioral measures: Hoehl, Wahl, & Pauen, 2014; Reid & Striano, 2005; Wahl et al., 2013; enhanced neural response in physiological measures: Hoehl, Reid, et al., 2008; Hoehl et al., 2012; Reid et al., 2004; Wahl et al., 2013). In addition to others' object-directed gaze, infants' object encoding is influenced by ostensive signals in direct face-to-face interactions (Csibra & Gergely, 2006, 2009). At 9 months, infants' object encoding is only enhanced if an adult has addressed the infant before performing the object-directed gaze shift, for example, through direct gaze or infant directed speech. If the adult does not address the infant sufficiently before looking at the object, 9-month-olds can follow the adult's gaze to a target object (Gredebäck, Astor, & Fawcett, 2018; Szufnarowska et al., 2015), but they do not show signs of encoding it (Okumura et al., 2020). The other way around, if the adult keeps looking in the direction of the infant without shifting their gaze to a visible object, 9-month-old infants do not process the object, even if they fixate on it when the object location is highlighted non-socially by illumination (Okumura, Kanakogi, Kobayashi, & Itakura, 2017). Together this suggests that, in direct triadic interactions with others, 9-month-old infants' learning about an object is enhanced by the interplay between others' object-directed gaze and previous eye contact. This is in line with the Natural Pedagogy account, which suggests that ostensive signals like eye contact (but also infant-directed speech, calling of the infants' name, or contingent responses) increase infants' attention and responsiveness to learn-worthy content, and scaffold their referential learning in direct face-to-face interactions (Csibra & Gergely, 2006, 2009).

At least by the end of the second year of life, children can learn about objects by merely observing others' triadic interactions. Two-year-old toddlers, for example, learn novel object labels equally well when they are directly addressed as when they overhear a conversation between others (Akhtar, 2005; Akhtar et al., 2001; Floor & Akhtar, 2006; O'Doherty et al., 2011). There is some indication from previous studies suggesting that this observational learning may be facilitated by similar processes as infants' learning from direct interactions. One recent study has, for example, shown that 2-year-old toddlers learned a novel object label only when observing two actors attending to an object together (i.e., following previous eye contact), but not when one of the actors was engaged with another activity (Fitch et al., 2020). Moreover, 18-month-olds have been found to learn the location of a hidden object only if an adult had used ostensive cues to indicate the location for another adult (alternating their gaze between interaction partner and object location while pointing), not when the adult had used non-ostensive similar behaviors (absent-minded gazing and extended index finger, Gräfenhain et al., 2009). These findings suggest that, at least by the age of 2 years, the sensitivity to ostensive cues may extend to third-party settings and that observed joint attention may promote observational learning about objects (Böckler, Knoblich, & Sebanz, 2011; see also Meng et al., 2017).

Infants' sensitivity to third-party interactions develops well before their second birthday: Around 6 months of age, infants perform more saccadic gaze shifts in accordance with the flow of a conversation when two people are facing each other as compared to two people standing back-to-back while talking (Augusti et al., 2010), they show increased pupil dilation in response to others' irrational compared to rational feeding interactions (Gredebäck & Melinder, 2010), and prefer to touch animated geometric-shape-agents who previously helped another agent over hindering characters (Hamlin, Wynn, & Bloom, 2007). Nine-month-old infants preferentially attend to face-to-face interactions over scenes in which the same two people turn away from one another (Handl et al., 2013; see also Galazka et al., 2014), and 10-month-olds expect a talking person to look in the direction of a human interaction partner (rather than an animated toy truck, Beier & Spelke, 2012; see also Bakker et al., 2011). At 12 months of age, infants can predict others' action goals in observed feeding interactions (Gredebäck & Melinder, 2010), and they anticipate a person to respond to another person's action only if this action serves a communicative goal (e.g., speech not coughing, Yamashiro & Vouloumanos, 2018; see also Thorgrimsson et al., 2015, 2014; von Hofsten et al., 2009).

Together, these findings suggest that already in the second half of the first year of life infants can recognize and selectively attend to situations in which two people engage in a social interaction, keenly observe and visually examine these scenes, and begin to understand the communicative function of others' gaze and gestures without being addressed themselves. What remains unclear, however, is whether infants at this age can already make use of others' interactions to gather knowledge about their environment and, if so, whether similar processes as

in direct interactions facilitate this observational learning. To directly compare the underlying processes of infants' learning from others' interactions with their learning from direct interactions, it requires a systematic investigation of the relative impact of third-party eye contact and third-party object-directed gaze on early observational learning.

4.1.1 The Current Study

The aim of this study was twofold: first, to investigate whether 9-month-old infants can learn about objects through merely observing social interactions between third parties and, second, to examine whether this learning depends on similar factors as infants' referential learning in direct interactions (i.e., the interplay between eye contact and object-directed gaze as present during joint attention). We tested 9-month-old infants because infants at this age typically engage systematically in joint attention behaviors themselves (e.g., Carpenter, Nagell, et al., 1998) and can make functional use of these situations in terms of object learning (e.g., Cleveland & Striano, 2007).

We conducted two experiments allowing us to directly compare infants' object encoding in a third-party observational context (Experiment 1) with their encoding in a first-party context (Experiment 2). In Experiment 1, infants were presented with an object-encoding task during which they saw four types of videos each showing one object together with two adults (encoding phase). Based on a 2 × 2 design, we systematically varied the content of the videos with respect to the eye contact between the two actors (eye contact or no eye contact) and the actors' gazing at the object (looking at the object or looking away from the object). After each video, infants were presented with a preferential-looking phase during which they saw the object they had just seen (familiarized object) next to a novel object. As a measure of infants' previous object encoding performance, we calculated their proportional looking times to the novel object, interpreting an increased novelty preference as an indicator of increased processing of the familiar object (Reid & Striano, 2005).

Based on previous studies, we hypothesized that if similar processes contribute to infants' learning in observational contexts as in direct interactions, infants' object encoding should be facilitated by observed joint attention, that is, when two actors gaze at an object together after previous eye contact. In this case, infants' proportional looking time to the novel object should be significantly higher in the third-party "eye contact – looking at object" condition compared to all other three conditions. If infants process information in observational contexts differently than in first-party contexts, the study design would allow to generate alternative explanations related to the relative impact of observed eye contact and others' gaze cues to an object. A main effect of one or both factors would, for example, indicate that infants' object encoding is influenced independently by observed eye contact and/or others' direct gazing at an object.

To compare infants' object encoding during third-party observation with a situation in which they were directly addressed, we ran a second experiment with a methodologically matched

first-party design. In contrast to Experiment 1, the videos in Experiment 2 showed one single adult gazing in the direction of the infant (eye contact) or looking away from the infant (no eye contact) before looking toward (or away from) an object. Based on previous research, we expected that the interplay between eye contact and subsequent object-directed gaze would promote infants' object encoding (e.g., Cleveland & Striano, 2007; Okumura et al., 2020). More specifically, we hypothesized that infants' proportional looking time to the novel object should be significantly higher in the first-party "eye contact – looking at object" condition compared to the other three conditions.

4.2   Experiment I

4.2.1   Methods

Ethical approval for the design and procedure of the study "Observing others' joint attention increases 9-month-old infants' object encoding" was provided by the Child Subjects Committee of the Max Planck Institute for Evolutionary Anthropology (no protocol number). We pre-registered the hypotheses, methods, procedures, and the data analysis plan for this experiment on the Open Science Framework (OSF; https://osf.io/yfegm/). Video examples, eye tracking raw data, and R scripts for pre-processing and analyzing the data are available at the same link on the OSF.

Participants

Thirty-two infants between 9 months, 0 days and 10 months, 0 days of age were included in the final sample of Experiment 1 (n = 13 female; M = 291.0 days, SD = 10.13 days). Data from three additional infants were excluded because they did not provide the minimum amount of one valid trial per condition (see section "Coding and Data Analysis" for valid trial criteria). All participants were born full-term (> 37 weeks). They were recruited on a voluntary basis via phone from the database of the Max Planck Institute for Evolutionary Anthropology in Leipzig. Children in this data base come from Leipzig (Germany) or surrounding areas, an urban Central-European, industrialized context. We did not collect individual data regarding the participants' socioeconomic background, but families in this database typically come from mixed, mainly mid to high socioeconomic backgrounds. Written informed consent was obtained from one parent of each infant prior to testing. The participants received a small gift as thank you for their participation in the study. The sample size was planned based on an a priori power analysis simulation including data from a pilot study (for details see section C3 in the supplemental materials).

Procedure

The testing took place at the Max Planck Institute for Evolutionary Anthropology in Leipzig (Germany). All participants were presented with an eye tracking task, during which they sat in front of a screen on their parent's lap. We used a 23.8″ monitor with 93 dpi and 1920 × 1080 screen resolution. To run the experiment and to record infants' gaze movements, a Tobii eye tracker (Tobii X120, Tobii Technology, Stockholm, Sweden) and Tobii eye tracking software (Tobii Studio version 3.4.8.1348) was used. Data was recorded separately for the left and the right eye at a sampling frequency of 120 Hz. We used a five-point calibration procedure to calibrate the eye tracker to the participant's eyes. The total duration of the Experiment was approximately 10 minutes, the whole visit at the lab around 20 minutes.

Stimuli and design

All participants were presented with 16 trials of an object-encoding task. Each trial consisted of an object encoding phase and a preferential-looking phase (see also Wahl et al., 2013). In the beginning of each phase, a blinking star was presented in the center of the screen together with an attention-grabbing sound to attract the infant's attention (2 seconds duration before the object encoding phase, and 1second duration before the preferential-looking phase). The overall duration of each trial was 24 seconds (see Figure 4.1).

Object Encoding Phase. Infants saw a video showing an object positioned between two adults. As objects, we used 32 pictures of abstract toys from an object stimuli collection applied in a study by Wahl and colleagues (2013). Based on a 2 × 2 design, the content of the videos was manipulated with regard to (a) the presence of initial eye contact between the two actors (eye contact or no eye contact), and (b) whether or not the actors looked at the object (looking toward or away from the object). To manipulate the eye contact between the actors, we used the movements of their bodies (turning toward or away from one another), the relative positioning of their bodies (face-to-face or back-to-back), and their gaze direction (eye contact or looking in opposite directions). The actors kept a neutral facial expression and remained silent during the entire sequence (see also Meng et al., 2017). To clearly highlight the third-party context, the actors never faced in the direction of the participant. All videos were presented without sound. The four conditions resulting from the 2 × 2 design were tested within-subjects, with each child being presented with four trials of each condition (see also Szufnarowska et al., 2015).

The videos had a total duration of 11 seconds, with identical timing over all conditions and videos (see Figure 4.1): Initially, the actors were seen in back view (1 s) before they turned toward (or away from) one another (1 s), and remained in this face-to-face or back-to-back position (1 s). Then, both actors turned their heads and gaze simultaneously in the direction of (or away from) the object (1 s) and remained in this position (5 s). Finally, the actors turned their heads and gazes

back (1 s) and remained in the initial face-to-face or back-to-back position (1 s). The videos were consistent with previous studies regarding the total duration, the non-social baseline sequence in the beginning, the duration of the actor's fixation times on the object, and the overall duration of eye contact between the interaction partners (Meng et al., 2017; Okumura et al., 2013, 2017, 2020; Szufnarowska et al., 2015; Theuring et al., 2007). In contrast to the video stimuli used in previous studies, we split the two-second overall duration of the eye contact sequence in two one-second lasting phases. This way, the joint attention scenario ("eye contact – looking at object" condition) met the minimum requirement for a complete joint attentional triangle (Carpenter & Liebal, 2011), including initial eye contact, subsequent looking toward an object, as well as a "closing" eye contact sequence following the mutual look at the object (similar to the dynamic in previous studies with real-interactive settings, e.g., Cleveland et al., 2007). All videos were presented in full-screen view, covering an area of 48° width × 27° height (at a screen distance of 60 cm).

Figure 4.1

Exemplary sequence and timing of one trial of the object processing task for each of the four experimental conditions in Experiment 1
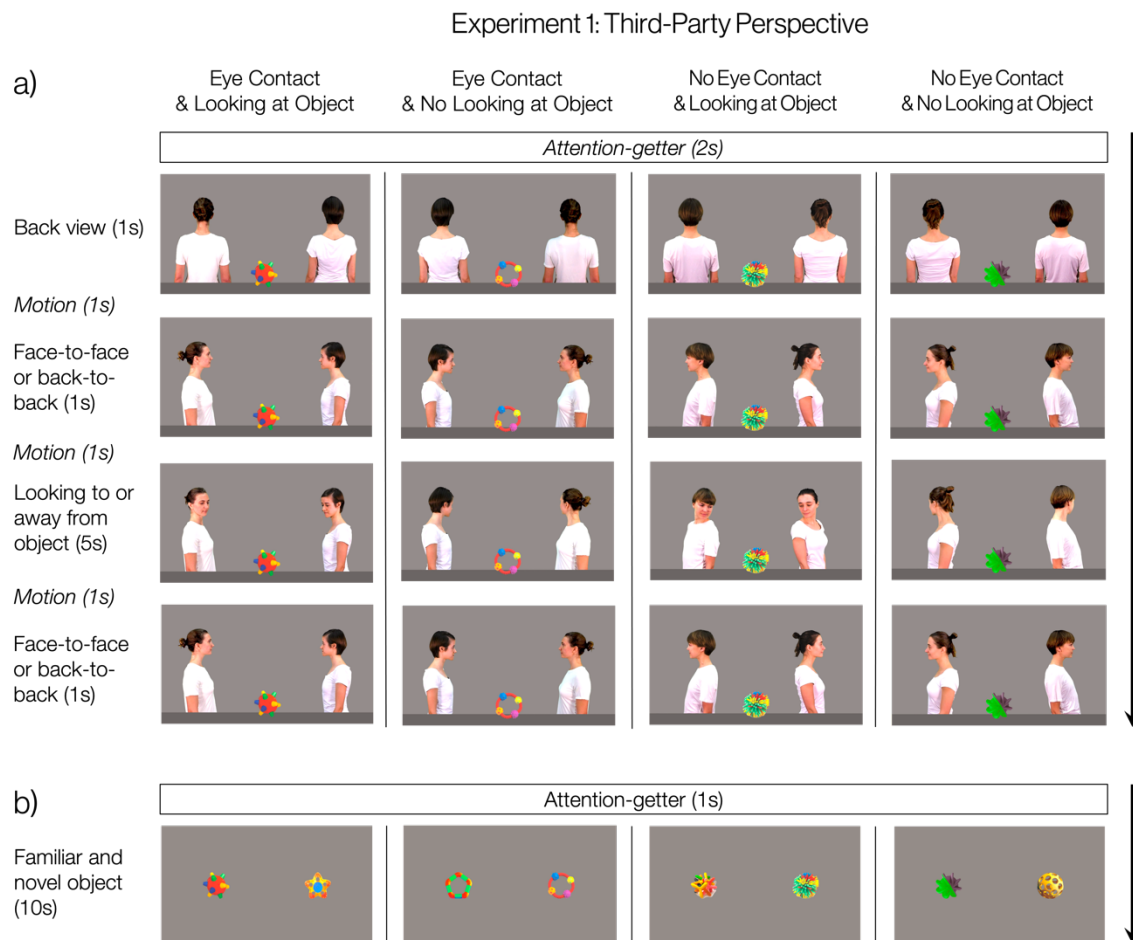
Notes. Every trial consisted of (a) an object encoding phase (11s), and (b) a preferential-looking phase (10s). Before each phase, an attention-getter (blinking star) was presented in the center of the screen on a black background.

Every child saw two different dyads of actors: one dyad performing in all trials of the two "eye contact" conditions, and the other dyad performing in all trials of the two "no eye contact" conditions. The left-right positioning of the actors within the dyads was reversed in the "looking at object" and "no looking at object" conditions (see Figure 4.1). The two dyads were seen equally often in all conditions. The actors were all female, wore white t-shirts and were visible from the waist up. All possible body and head movements from all actors in all conditions covered an area of 14.5° × 20.8° on both sides of the object, with the head movements covering an area of 11.2° × 8.8°. The objects covered an area of 6.5° × 6.5°. The minimum distance between each actor's face and the object was 9.3°. In section C1 in the supplemental materials we provide detailed information on how the video stimuli were created.

Preferential-Looking Phase. We presented infants with two objects at the same time: the object they had previously seen in the encoding phase (familiar object) and a novel object. The objects were presented side-by-side on a grey background for a duration of 10 seconds. The size of the objects was identical to the object size in the encoding phase. The distance between the outer edges of the objects was 18.2°. The positioning of the novel object (right or left) was counterbalanced within infants and condition (i.e., each infant saw the novel object in two trials per condition on the right side, and in two trials on the left side). The pairing of objects in the preferential-looking phase was randomized and consistent over participants. All infants saw the same 32 objects, with the same toy never occurring twice. Sixteen objects were presented as familiar objects in the object encoding phase, and 16 additional objects as novel objects in the preferential-looking phase. The role of each individual object (novel or familiar) was counterbalanced across participants and conditions, meaning that each object served equally often as novel and as familiar object in all four conditions over all participants.

The 16 experimental trials were presented in four blocks with four trials each. Within a given block, each trial presented a different condition (see also Szufnarowska et al., 2015). We counterbalanced the order of conditions during the first block across infants (in the way that an equal number of infants saw condition 1 first, condition 2 first, etc.). The order of conditions in the remaining three blocks was pseudo-randomized. No condition occurred more than twice in a row. After every block, infants were presented with a 4-second kaleidoscope video with a soothing melody to maintain their attention (see also Reuter et al., 2018; Szufnarowska et al., 2015).

Coding and data analysis

We used the R software environment (R version 3.6.3, RStudio version 1.2.1335) for pre-processing and analyzing the data, as well as for setting areas of interest (AOIs). As main dependent variable, we measured infants' mean proportional looking time to the novel object in the preferential-looking phase. For this purpose, we defined two square-shaped AOIs: one AOI covering the novel object, and one AOI covering the familiar object. To accommodate for inaccuracies in calibration,

all AOIs were defined 1° visual angle larger than the maximal dimensions of the stimuli (Gredebäck et al., 2009). We assessed the total duration of fixations in both AOIs for each trial and participant, including fixation data from the entire preferential-looking phase. Data for both the left and the right eye of each participant was averaged. When one eye could not be measured, we used the data from the other eye. To define fixations, we used the Tobii Velocity-Threshold Identification (I-VT) fixation filter with default settings (for details see section C2 in the supplemental materials). In a next step, we calculated the proportional looking time at the novel object ("novelty preference score") by dividing the duration of fixations to the novel object by the total duration of fixations to both objects. The novelty preference score could take values between 0 and 1, with values above .50 indicating a relatively longer looking time at the novel object. We only included a trial if infants had looked at least at one object during the preferential-looking phase for at least one fixation, and if they had paid visual attention to the central parts of the video for at least one fixation (see also Michel et al., 2017). In accordance with our pre-registered plan, we counted the initial face-to-face (or back-to-back) phase, and the looking-to-object (or away-from-object) phase as "central parts", excluding the motion sequences. Infants were only included in the analysis if they provided valid data in at least one trial per condition after being filtered according to these criteria (see also Wahl et al., 2013). On average, infants provided 3.46 valid trials (SD = .69) per condition (for detailed valid trial statistics see Table C1 in the supplemental materials).

To test our hypotheses, we fitted a generalized linear mixed model (GLMM) with a Gaussian error structure. All models were fitted using the R package "lme4" (Version 1.1-21, Bates et al., 2020). The dependent measure was the novelty preference score. To investigate infants' object encoding as a function of third-party eye contact and third-party object-directed gaze, we fitted a GLMM including the interaction between third-party eye contact (eye contact, no eye contact) and others' object-directed gaze (looking toward the object, not looking at object) as a fixed effect. To account for possible trial effects, trial (z-transformed) was included as control variable in the model. All factors were tested within-subjects. As random effects, we included subject as intercept, as well as random slopes on subject for trial, eye contact, object-directed gaze, and the interaction between eye contact and object-directed gaze. The significance of the individual fixed effects was based on likelihood ratio tests comparing the full models with respective reduced models excluding the individual fixed effects using the *drop1*-function in R with an alpha-level of .05.

We ran the following analyses in addition to the pre-registered plan. First, we conducted six pair-wise comparisons to compare the novelty preference score between all four conditions. All pairwise comparisons were based on the GLMM fitted for the main analysis, using the R package "emmeans" (Version 1.4.6, Lenth, 2020). To account for multiple comparisons, the alpha-level was adjusted via Bonferroni correction. Second, we calculated one-sample t-tests against .50 within each condition to determine whether the novelty preference score significantly differed from chance level. In addition, we explored whether infants' direct attention to the object in the

encoding phase could explain their looking preference in the subsequent preferential-looking phase. For this purpose, we fitted two additional GLMMs to investigate infants' novelty preference score (dependent variable). In one model, we included infants' looking time to the object in the encoding phase as fixed effect (within-subject factor). In the second model, we included a binary fixed effect variable indicating for each trial whether the child had looked at the object during encoding at all, that is, whether their fixation duration within the object AOI was greater than zero (within-subject factor: yes, no). Table C1 in the supplemental materials provides an overview of the corresponding valid trial statistics. In both models, we included subject as random intercept, as well as the random slope of the fixed effect on subject. Fixation data from the entire video sequence was included for the measure of looking time. We have conducted some further analyses to better understand the impact of infants' overt attention during the encoding phase on their encoding performance. We did not find any systematic condition differences in infants' looking times or gaze patterns. For conciseness, we present the corresponding results in section C2 in the supplemental materials.

4.2.2 Results

The comparison between the full model and the reduced model revealed a significant result indicating that at least one of the fixed effects had an impact on infants' mean proportional looking time to the novel object in the preferential-looking phase ($\chi^2(4) = 14.99$, $p = .005$). More specifically, the interaction between third-party eye contact and third-party looking at the familiarized object had a significant effect on infants' novelty preference ($\chi^2(1) = 4.03$, $p = .04$, estimate = −.08, $SE = .04$; see Figure 4.3a). We did not find an effect of trial ($\chi^2(1) = 1.22$, $p = .27$, estimate = .01, $SE = .01$).

Additional analyses revealed that infants' novelty preference score in the joint attention condition ("eye contact – looking at object") was significantly higher compared to the scores in all other three conditions (see Table 4.1). Moreover, the score was significantly higher than chance level in only this condition ($M = .62$, $SD = .13$, $t(31) = 5.28$, $p < .001$, $d = .93$, see Table 4.2). Infants' preferential orienting to the novel object did not depend on their attention to the familiarized object in the encoding phase: Neither did infants' looking time to the familiarized object predict their subsequent novelty preference ($\chi^2(1) = .49$, $p = .48$, estimate = .01, $SE = .01$), nor was the preference score systematically influenced by whether or not infants had looked at the object during encoding at all ($\chi^2(1) = .14$, $p = .71$, estimate = −.009, $SE = .02$). Table C3 in the supplemental materials shows the descriptive statistics of the looking times during the encoding phase in all four conditions.

Table 4.1

Results from post-hoc pair-wise comparisons of the novelty preference score between the four conditions of Experiment 1 (third-party perspective)

Compared conditions	estimate	SE	df	t	p
Eye Contact – Looking at Object vs. No Eye Contact – Looking at Object	−.10	.03	81.9	−3.28	.009**
Eye Contact – Looking at Object vs. Eye Contact – No Looking at Object	.08	.03	74.8	2.89	.03*
Eye Contact – Looking at Object vs. No Eye Contact – No Looking at Object	.10	.03	91.8	3.40	.006**
No Eye Contact – Looking at Object vs. Eye Contact – No Looking at Object	−.01	.03	89.0	−0.48	1.0
No Eye contact – Looking at Object vs. No Eye Contact – No Looking at Object	.002	.03	160.2	0.08	1.0
Eye Contact – No Looking at Object vs. No Eye Contact – No Looking at Object	−.02	.03	150.9	−0.56	1.0

Notes. *$p < .05$ **$p < .01$

Table 4.2

Results from post-hoc one-sample t-tests testing the novelty preference score within the four conditions of Experiment 1 against chance

Condition	M (SD)	t	df	p	d
Third-Party Eye Contact Looking at object	.62 (.13)	5.28	31	<.001***	.93
No Third-Party Eye Contact Looking at object	.52 (.11)	0.96	31	.34	.17
Third-Party Eye Contact No Looking at object	.53 (.11)	1.47	31	.15	.26
No Third-Party Eye Contact No Looking at Object	.51 (.11)	0.49	31	.62	.09

Notes. ***$p < .001$

4.2.3 Discussion

We found that 9-month-old infants showed an increased looking preference for a novel over a familiarized object after they had observed two adults attending to the familiarized object jointly (i.e., following mutual eye contact). This novelty preference was significantly higher compared to situations in which infants observed two adults looking at each other but away from the object (purely interpersonal attention), two adults looking at the object individually in a non-communicative back-to-back setting (non-coordinated parallel attention), or scenarios in which two adults looked neither at the object nor at each other (neither person- nor object-directed attention). Based on the assumption that novelty preference reflects better stimulus encoding (Cleveland et al., 2007) our findings suggest that, in a purely observational setting, others' joint

attention toward an object can increase 9-month-olds' encoding of this object. We did not find any evidence supporting the assumption that infants' increased encoding depended on infants' overt attention to the object during encoding. This is in line with previous studies showing that direct attention to an object is not required for its encoding (Cleveland & Striano, 2007; Wahl et al., 2013). In the general discussion we provide a more detailed discussion of this assumption.

To directly compare infants' object encoding performance in a third-party observational setting to a situation in which they were directly addressed, we tested an additional sample of infants in a second experiment investigating infants' object encoding in a matched first-party setting during which they saw one single adult on screen.

4.3 Experiment II

4.3.1 Methods

The experimental design, procedure, as well as data pre-processing and analysis procedures were identical to Experiment 1, but this time only one actor was visible in the videos in the encoding phase. We pre-registered the hypotheses, methods, procedures, and the data analysis plan for this experiment on the OSF (https://osf.io/dp5cg/). Video examples, eye tracking raw data, and R scripts for pre-processing and analyzing the data are available at the same link on the OSF. The design and procedure of this experiment was approved by the same Ethics Committee as in Experiment 1.

Participants

Thirty-two full-term infants between 9 months, 0 days and 10 months, 0 days of age were included in the final sample of Experiment 2 (n = 16 female; M = 282.69 days, SD = 8.4 days). Data from four additional infants were excluded because they did not provide the minimum amount one valid trial per condition. The criteria for data inclusion were the same as in Experiment 1. The participants were recruited from the same data base as described in Experiment 1. Each child participated only in one of the two Experiments.

Stimuli and design

The videos in the encoding phase were as similar as possible to previous first-party studies (e.g., Okumura et al., 2013, 2020), while keeping methodological and visual details consistent with Experiment 1 (see section C1 in the supplemental materials for more details regarding the video stimuli). Analogous to the 2 × 2 design in Experiment 1, the content of the videos in the encoding phase was manipulated with regard to (a) the gaze direction of the adult in relation to the infant (eye contact, or no eye contact), and (b) whether or not the actor looked at the object (looking toward or away from the object). To manipulate the actor's gaze direction in relation to the infant,

we used the movement of the actor's body (turning toward the child or to the side), the relative positioning of the body (facing the child or averted), and the actor's gaze direction (looking toward the child or to the side).

The videos in this experiment were edited in a way that they had exactly the same timing and degree of motion as the videos in Experiment 1, except that they contained one and not two actors (see Figure 4.2): Initially, the actor was seen in back view (1 s) before turning in the direction of the child or to the side (1 s), and remaining in this position (1 s). Then, the actor turned her head and gaze simultaneously in the direction of (or away from) the object (1 s), and remained in this position (5 s). Finally, the actor turned her head and gaze back (1 s), and remained again in this initial position (1 s).

Figure 4.2

Exemplary sequence and timing of one trial of the object processing task for each of the four experimental conditions in Experiment 2
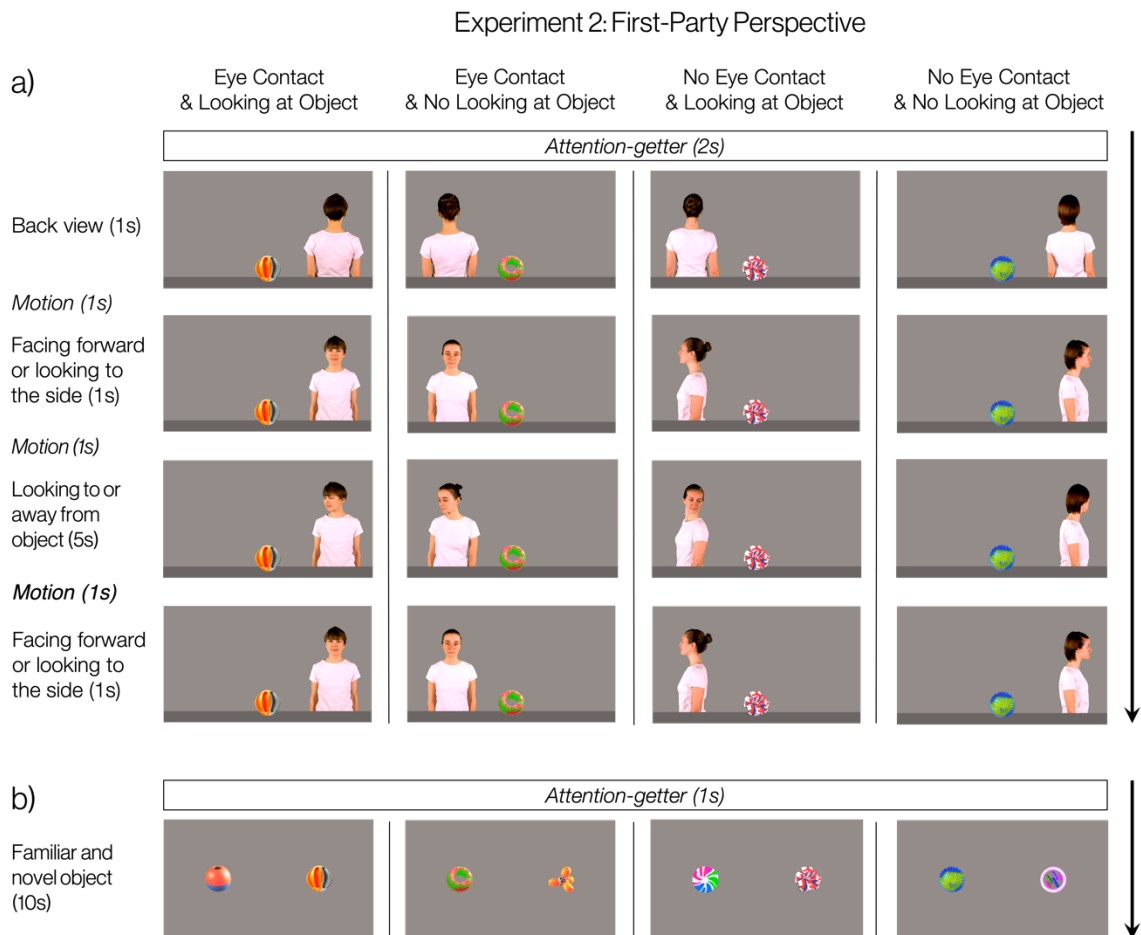
Notes. Every trial consisted of (a) an object encoding phase (11s) and (b) a preferential-looking phase (10s). Before each phase, an attention-getter (blinking star) was presented in the center of the screen on a black background.

All possible body and head movements from all actors in all conditions covered an area of 15° × 20.8° on both sides of the object, with the head movements covering an area of 12.3° × 8.8°. The minimum distance between each actor's face and the object was 8.4° (see Figure C2 in the supplemental materials for an illustration of the visual arrangement on screen). As in Experiment 1, each child was presented with four trials of each condition, presented in four blocks. Every child saw each of the 4 conditions performed by a different actor. The actors were the same as in the first Experiment. They were shown equally often in all four conditions, counterbalanced between infants. The positioning of the actor left or right from the central object was counterbalanced within condition, child, and block (i.e., each child saw the actor in each block during two trials on the right side and during two trials on the left side; over the whole experiment, the actor occurred two times on the right side and two times on left side within each condition).

Coding and data analysis

Data pre-processing and trial inclusion criteria were the same as in Experiment 1. Importantly for the manipulation of eye contact in Experiment 2, this ensured that only trials were included in which the infant had paid attention to the initial direct-gazing phase during which the adult looked in the direction of the infant. We conducted the same main analyses and post-hoc tests. Analogous to Experiment 1, our main dependent variable was infants' proportional looking time to the novel object in the preferential looking phase. In the main model, we included the interaction between eye contact (eye contact, no eye contact) and object-directed gaze (looking toward or away from the object) as main fixed effect. The same control variables and random effects were included as in Experiment 1. On average, infants provided 3.42 valid trials (SD = .81) per condition (see Table C3 in the supplemental materials for the detailed valid trial statistics).

4.3.2 Results

The full-null model comparison revealed a significant result ($\chi^2(4)$ = 15.64, p = .004). Specifically, the interaction between eye contact and others' object-directed gaze had a significant impact on infants' novelty preference score ($\chi^2(1)$ = 5.77, p = .02, estimate = −.11, SE = .04; see Figure 4.3b). We did not find an effect of trial ($\chi^2(1)$ = 1.47, p = .22, estimate = −.01, SE = .01).

Additional analyses revealed that infants' novelty preference in the "eye contact – looking at object" condition was significantly higher compared to all other three conditions (see Table 4.3). Moreover, the score was significantly higher than chance level in only this condition (M = .61, SD = .13, $t(31)$ = 4.62, p < . 001, d = .82, see Table 4.4). Infants' preference for the novel object did not depend on the time they had looked at the object during the encoding phase ($\chi^2(1)$ = .003, p = .95, estimate = −.001, SE = .01), or on whether they had looked at the object during this phase

at all ($\chi^2(1) = 1.03$, $p = .30$, estimate = −.04, $SE = .04$). Table C4 in the supplemental materials shows the descriptive statistics of the looking times during the encoding phase in all four conditions.

Table 4.3

Results from post-hoc pair-wise comparisons of the novelty preference score between the four conditions of Experiment 2 (first-party perspective)

Compared conditions	*estimate*	*SE*	*df*	*t*	*p*
Eye Contact – Looking at object vs. No Eye Contact – Looking at object	−.10	.03	79.3	−3.17	.01*
Eye Contact – Looking at object vs. Eye Contact – No looking at object	.10	.03	79.3	3.23	.01*
Eye Contact – Looking at object vs. No Eye Contact – No looking at object	.10	.03	92.9	3.05	.02*
No Eye Contact – Looking at object vs. Eye Contact – No looking at object	.002	.03	91.4	0.07	1.0
No Eye Contact – Looking at object vs. No Eye Contact – No looking at object	−.004	.03	167.2	−0.14	1.0
Eye Contact – No looking at object vs. No Eye Contact – No looking at object	.006	.03	166.4	0.21	1.0

Notes. Results are based on the pairwise contrasts between the estimated marginal means of all conditions, inferred from the main GLMM fitted to the data by using the R-package "emmeans". P-values are adjusted via Bonferroni correction for six tests. *$p < .05$

Table 4.4

Results from post-hoc one-sample t-tests testing the novelty preference score within the four conditions of Experiment 2 against chance

| Condition	M *(SD)*	*t*	*df*	*p*	*d*
First-Party Eye Contact Looking at object	.61 *(.13)*	4.62	31	<.001***	.82
No First-Party Eye Contact Looking at object	.49 *(.11)*	−0.44	31	.66	−.07
First-Party Eye Contact No Looking at object	.48 *(.12)*	−0.79	31	.43	−.14
No First-Party Eye Contact No Looking at Object	.50 *(.12)*	0.12	31	.91	.02

Notes. ***$p < .001$

To explore possible differences between Experiment 1 and 2, we repeated our main analysis over a merged sample including infants from both experiments (*N* = 64). In addition to the fixed effects included in the main analysis, we included Experiment (1 or 2) as fixed effect. As in the separate analyses of the two experiments, we found that infants' novelty preference varied as a function of eye contact and looking at the familiar object (effect of the interaction: $\chi^2(1) = 8.35$, $p = .004$, estimate = −0.09, $SE = 0.03$). We did not find any effect of trial ($\chi^2(1) = 0.0001$, $p = .99$,

estimate = 0.00, *SE* = 0.008) or experiment ($\chi^2(1) = 1.97$, *p* = .16, estimate = −0.02, *SE* = 0.02). Identical to the separate analyses of each experiment, infants preferred the novel over the familiar object only in the "eye contact – looking at object" condition when combining data of both experiments.

Figure 4.3

Boxplots with individual data points illustrating the mean proportional looking time to the novel object in the preferential-looking phase for the four conditions of (a) Experiment 1 and (b) Experiment 2
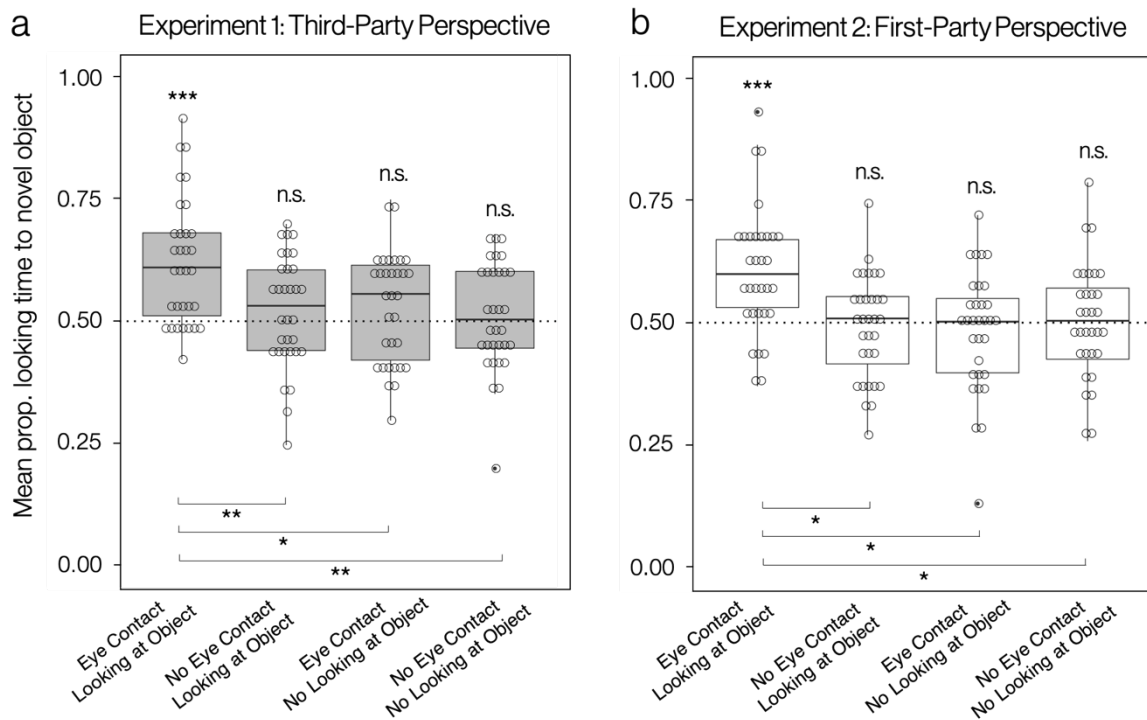
Notes. The dashed line at .50 represents chance level. The results of the post-hoc one-sample t-tests against chance level are depicted above the boxplots, the results of the pair-wise comparisons below. P-values of the pair-wise comparisons are adjusted via Bonferroni correction. *p < .05 **p < .01 ***p < .001

4.3.3 Discussion

We found that 9-month-old infants showed an increased looking preference for a novel over a familiarized object if an adult had established eye contact with the infant before looking at the familiar object. This novelty preference was relatively higher compared to situations in which infants had seen an adult looking in their direction but away from the object, an adult looking at the object but not at the infant, or an adult looking neither at the object nor the infant. Infants' increased novelty preference could not be explained by increased overt attention to the object during encoding. Our findings suggest that the interplay between others' object-directed gaze and previous eye contact promotes object learning in 9-month-old infants, as the two factors alone were not sufficient to elicit this effect. This represents a conceptual replication of previous studies showing that joint visual attention affects infants' subsequent looking behavior and object encoding (e.g., Cleveland et al., 2007; Itakura, 2001; Okumura et al., 2020; Wahl et al., 2013).

Additional analyses over a merged sample including participants from Experiment 1 and 2 suggested that infants' encoding performance in a third-party observational setting was identical to their performance in a first-party setting.

4.4   General Discussion

Previous studies have shown that, by the end of the second year of life, infants can learn about their environment by merely observing others' social interactions. In the present study we show that this ability emerges already in the second half of the first year after birth, and that early observational learning may be influenced by similar factors as infants' referential learning through direct interactions around this age. In Experiment 1, we found that 9-month-old infants showed an increased preference for a novel object compared to a familiarized object if they had previously seen two adults looking at the familiarized object in a joint attentional setting (i.e., following previous eye contact). This novelty preference was greater than chance level and higher compared to all other conditions, reversely indicating that observed joint attention to an object selectively increased infants' encoding of this object. In Experiment 2, we found the corresponding result pattern in a matched first-party design during which infants were directly addressed by one single adult on screen. Infants' novelty preference was highest if they had seen an adult addressing them through direct gaze before looking at the familiar object.

Our finding that the *interplay* between eye contact and others' object-directed gaze selectively increased infants' object encoding when being directly addressed, aligns with the theoretical assumption that communicative cues (Csibra & Gergely, 2006, 2009) and coordinated attention with others (Tomasello et al., 2005) provide an important framework for young infants' learning. Our finding that also *observed* joint attention can increase infants' object encoding, provides a crucial extension of this view by demonstrating that the promoting effect of joint attention goes beyond first-hand experience in direct pedagogical settings (see also Fitch et al., 2020 for supporting evidence with older children). Moreover and importantly, our findings imply that infants' successful learning through observation depended on active monitoring of others' interactions, not on passive reception (Oudeyer & Smith, 2016; Rogoff et al., 2003): When two adults established a communicative context by making eye contact, but then turned their gaze away from a visible object, infants did not show any signs of increased encoding of this object, suggesting that they invested their resources selectively in presumably meaningful triadic interactions. To enable selective learning, infants had to detect, selectively and intently observe, and evaluate the relevant learning opportunity. It remains to be studied what other factors besides observed coordination of visual attention define a meaningful social interaction. Following evidence from first-party studies, other potential factors could relate to characteristics of the interaction partners, such as parents versus strangers (e.g., Hoehl et al., 2012), neutral versus

emotional affective states (e.g., Hoehl, Wiese, & Striano, 2008), or real versus artificial interaction partners (e.g., O'Doherty et al., 2011).

From a developmental perspective, the strikingly similar result pattern in the first- and third-party settings of our experiments is particularly noteworthy considering the narrow age range we tested in this study. The age around 9 months has been previously shown to represent a period during which infants' social development and learning undergo crucial developments as they develop competencies for triadic interactions (Carpenter, Nagell, et al., 1998; Striano & Rochat, 1999). By 9 months of age, infants can reliably coordinate their attention with others in joint attention episodes, and make functional use of these situations to acquire knowledge about external objects in their environment (Cleveland & Striano, 2007; Okumura et al., 2020). Our finding that, at the same age, the promoting effect of joint attention extends to merely observational settings, provides a novel perspective and important contribution to our understanding of the diverse pathways through which infants can learn from social interactions. Together with the previous literature on infants' active social interaction behavior, our findings raise diverse possibilities regarding the developmental relationship between the first- and third-party level. One possibility would be a causal relation. For example, infants' increasing own triadic skills and experiences during the second half of the first year of life may modulate their perception, understanding, and learning from others' joint attention or vice versa (see also Gredebäck & Melinder, 2010; Henderson et al., 2013; Herold & Akhtar, 2008). Alternatively, it would be possible that infants' sensitivity to own and others' joint attentional interactions develops in a parallel manner, driven by underlying social motivational processes (Dawson et al., 2004). To disentangle the relationship and possible causality of the two levels, longitudinal study designs would be needed.

Importantly, the main focus in this study was on the *outcome* of infants' encoding, measured by infants' visual recognition memory after the processing situation itself. To gain an additional insight into concrete processes during the actual encoding, we explored infants' looking pattern during the encoding phase further (see also section C2 in the supplemental materials for a discussion of our additional analyses). Overall, we did not find any indication that infants' object encoding would depend on their own direct attention to the object. This is in line with previous studies suggesting that overt attention to an object is not required for its processing (neurological measures: Hoehl et al., 2014; Reid & Striano, 2005; Reid et al., 2004; behavioral measures: Wahl et al., 2013; Okumura et al., 2013; Cleveland & Striano, 2007; Cleveland et al., 2007). Instead, it is likely that increased covert attentional orienting (i.e., shifts of visual attention occurring independently of eye movements) modulated infants' processing of the target object (Johnson, 1994; Posner, 1980).

The exact processes of infants' covert attentional orienting during third-party joint attention remain unclear. One possibility would be that observed communicative cues such as eye

contact highlight the meaningfulness of an upcoming referential gaze shift, causing a covert attentional shift in the direction of the referenced target object (Wahl et al., 2013; see also Daum & Gredebäck, 2011). Alternatively, the opportunity to observe a meaningful interaction may increase infants' responsiveness more generally, providing them with the necessary attentional activation to process everything within the range of the covert attentional field ("socially aware mode", Puce & Bertenthal, 2015). To disentangle these possibilities and to fully capture the mechanisms responsible for facilitated object encoding, it requires neural measures capturing attentional phenomena over and above direct visual attention. Moreover, infants' neural processes during the encoding situation need to be studied in relation to their subsequent behavioral and neuronal response to the object.

The implications of our findings are particularly profound with regard to cross-cultural differences in infants' everyday learning environments (Akhtar & Jaswal, 2020). Children across cultural contexts participate to highly variable extents in direct face-to-face interactions (e.g., Mesman et al., 2018). Infants in many cultures are thus much more accustomed to observational learning than to direct teaching (Paradise & Rogoff, 2009; Rogoff et al., 2007). From this perspective, our findings support the idea that joint attention per se, that is, independent of whether experienced through active participation or via observation of others, may represent a culturally universal communicative context in which generic knowledge can be transmitted (see also Correa-Chávez & Roberts, 2012). To investigate this possibility directly, it would require studies systematically comparing infants from different cultural backgrounds regarding their sensitivity and responsiveness to ostensive signals in direct interactions (e.g., Hernik & Broesch, 2019) as well as in third-party settings (e.g., Correa-Chávez & Rogoff, 2009).

4.4.1 Limitations

The findings of this study need to be considered against some limitations. First, our definition of joint attention (and, thus, the manipulation in our study design) was based on gaze cues. This was intended because mutual eye contact and others' object-related gaze have been previously identified as important influential factors on infants' object encoding. Moreover, gaze alternation between an object and an interaction partner has been used as an early indicator of 9-month-olds' joint attentional awareness in studies on infants' active social behavior (e.g., Bakeman & Adamson, 1984). To account for the broader range of social cues that infants encounter and produce in real interactions, future studies should test the generalizability of our findings to other ostensive signals (e.g., infant directed speech, calling the infants' name, contingent responsivity; Csibra, 2010) and other referential social cues such as pointing or showing gestures (especially once infants begin to use these gestures themselves to initiate joint attention, Carpenter, Nagell, et al., 1998; Mundy et al., 2007). This relates to another restriction of our findings, namely the highly controlled stimuli and testing environment. This limitation is particularly relevant regarding the

interpretation of the similar result pattern in the first-party and third-party context. To investigate whether, in the real world, third-party joint attention is equally powerful on infants' learning as actively shared attention, our finding needs to be probed in a more natural environment. Moreover, looking-time based measures alone cannot directly inform us about the immediate underlying mechanisms that promote infants' learning in active interactive and observational settings (Aslin, 2007). Future studies are required using additional measures, such as measures of neural or physiological activity to capture signs of underlying mechanisms like emotional arousal (see, e.g., Hepach & Westermann, 2016 for a review on pupillometry in infancy research).

Importantly, our focus was on influential factors within the *social* part of the joint attentional triangle, not on the *kind* of information that infants learn about novel objects within different contexts. In focusing more specifically on the latter aspect, previous studies have suggested that 9-month-old infants may retain qualitatively different information about novel objects in direct communicative as compared to non-communicative contexts (Okumura, Kobayashi, & Itakura, 2016; Yoon, Johnson, & Csibra, 2008). While communicative context induced the representation of information about object identity (relevant for generalization) at the cost of learning information about object location, infants preferentially encoded the opposite information in non-communicative settings (Yoon et al., 2008; but see Silverstein, Gliga, Westermann, & Parise, 2019 for failed replication attempts). Investigating the influence of such a communication-induced memory bias in third-party settings would deepen our understanding of the kind of information infants learn from observing others' interactions, and how this compares to their learning in direct pedagogical settings.

4.4.2 Conclusion

In summary, we could show that observing others' interactions can increase 9-month-old infants' object encoding, and that this early observational learning is influenced by similar factors as infants' learning from direct social interactions at the same age. In both situations, the interplay between eye contact and object-directed gazing selectively enhanced infants' encoding of novel information. These findings suggest that not only active participation in joint attention, but also the mere observation of others' joint attention has a profound impact on early social learning. At a broader level, this has significant implications for the understanding of the multifaceted ways in which human infants learn about their world.

5
General Discussion

During the last ten years, a growing number of studies have investigated how infants represent social relations and interactions between third parties and how they come to use third-party interactions to learn about their environment. In contributing to this literature, this dissertation investigated the underlying processes and mechanisms involved in infants' observational learning from third-party social interactions. Throughout this thesis, I have argued that a thorough account of infants' learning from third-party interactions should not only focus on the actual learning situation but also consider how infants gain access to potential learning opportunities. Following this assumption, this thesis investigated (a) social attentional developments and underlying motivational drivers helping infants to detect and approach third-party interactions (Study 1&2) and (b) factors that promote infants' memory and learning while they observe third-party interactions (Study 3). In this chapter, I will discuss and integrate the empirical findings of this thesis, point out central limitations of the current approaches, and derive open questions that need to be studied in future research.

5.1   Summary of Results

Study 1: How does infants' orienting to third-party interactions develop during the second half of the first year of life, and how does this development coincide with changes in infants' active social attention behavior?

The results from Study 1 suggest that infants increasingly orient their attention toward third-party interactions throughout the second half of the first year of life. In Experiment 1, 9.5- to 11-month-old infants ($N = 20$) showed a stronger looking time preference than 7- to 8.5-month-olds ($N = 20$) for videos showing dyadic face-to-face interactions over videos showing non-interactive back-to-back scenes. Moreover, during active social engagement with their parent, 9.5- to 11-month-olds showed more social looking behaviors (especially joint attention) than 7- to 8.5-month-olds (see also Carpenter, Nagell, et al., 1998; Striano & Rochat, 1999). In Experiment 2, we could replicate our findings from Experiment 1 and extend them by showing that the increase in

both modalities follows a continuous and linear trajectory from 7 to 13 months of age ($N = 50$). In a merged sample over both experiments, the two measures were spuriously correlated through infants' age ($N = 90$). Together, these findings suggest that infants' attentional orienting toward others' interactions coincides with developments in infants' social attention during active social engagement.

Study 2: Is it intrinsically valuable for infants to observe third-party interactions?

The results from Study 2 extend the findings from Study 1 in that they provide support for the idea that proximate reward mechanisms may account for infants' attentiveness to third-party social interactions. In a gaze-contingent associative learning task, 13-month-old infants ($N = 32$) showed superior learning of a cue-target association guiding them to videos depicting a dyadic face-to-face interaction scene compared to videos showing two people acting individually while standing back-to-back. More specifically, infants showed faster saccadic latencies and more predictive gaze shifts from a centrally displayed geometrical shape cue to a laterally displaced target region when having the opportunity to observe a social interaction target. These findings suggest that, without any additional reinforcement, face-to-face interaction targets can selectively power infants' associative learning. This indicates that third-party interactions can serve as primary reinforcers in 13-month-old infants.

Study 3: Can infants in the first year of life use third-party interactions to learn about objects? If so, do similar processes contribute to infants' observational learning from third-party interactions as to their referential learning in direct social interactions?

The results from Study 3 demonstrate that already 9-month-old infants can learn about the perceptual features of novel objects through merely observing triadic interactions between others. Moreover, the findings suggest that similar factors may influence this early observational learning as infants' referential learning through direct interactions at this age. In a looking-time based object encoding paradigm, 9-month-old infants showed increased object encoding when observing a joint attention scenario during which two adults shared eye contact before shifting their gaze mutually toward a visible object ($N = 32$, Experiment 1). More specifically, infants' novelty preference was significantly higher compared to conditions showing two adults looking at each other but away from an object (purely interpersonal attention), two adults looking at an object individually in a back-to-back setting (non-coordinated parallel attention), or two adults looking neither at a visible object nor at each other (neither person- nor object-directed attention). Experiment 2 replicated the previous finding that, in direct interactions, infants' object encoding depends on the interplay between eye contact and object-directed gaze as well (Cleveland & Striano, 2007; Okumura et al., 2020). Like in Experiment 1, infants' object encoding was only

enhanced when an adult made eye contact with them before shifting their gaze toward a visible object (N = 32, Experiment 2).

5.2 Research Contributions

Altogether, the empirical work in this thesis adds three substantial contributions to our understanding of infants' attention to and learning from third-party social interactions: The findings (1) provide a novel perspective on motivational mechanisms as proximal drivers of infants' attention to third-party interactions, (2) uncover factors that promote infants' learning during the observational learning situation, and (3) enhance our understanding of the ontogenetic origins of infants' attention to and learning from others' social interactions. I will elaborate more on each of these three contributions in the following sections.

5.2.1 Considering Infants' Intrinsic Motivations As Proximal Drivers Toward Third-Party Social Interactions

The influence of infants' intrinsic motivations on their attention and learning has received little attention in the third-party interaction literature, especially compared with research focusing on the infant's active role in social interactions. As outlined in more detail in chapter 1, previous first-party studies have revealed that infants' behavior in the second half of the first year of life is marked by an increased motivation to orient to social partners (social orienting), to seek and like social engagement with others (social reward), and to establish social relationships (social maintaining). All three dimensions have been considered indicators of social motivation (Chevallier et al., 2012). The empirical work of this thesis extends this research by suggesting that not only infants' active interaction behavior but also their attentional orienting toward others' interactions is influenced by social motivational factors. In the following two sub-sections, I will discuss the findings from Study 1 and Study 2 as supporting evidence for two psychological components of social motivation: social orienting and social reward. In the third sub-section, I will provide an alternative interpretation of the findings by discussing them in the light of infants' intrinsic motivation to learn from others.

Evidence for an attentional preference for third-party social interactions ("social orienting")

In the social motivation framework by Chevallier and colleagues (2012), the social orienting component of social motivation is defined as prioritized attention to social signals. Most previous studies on infants' attention to third-party interactions have relied on paradigms with one-by-one stimulus presentation, such as violation-of-expectation paradigms (Beier & Spelke, 2012; Gustafsson et al., 2016) or paradigms focusing on infants' saccadic eye movement patterns during real-time observation (Augusti et al., 2010; Handl et al., 2013). While these studies reveal important

insights into infants' perception and understanding of third-party interactions, they do not allow for conclusions about whether infants *prioritize* attending to one over the other scenario. Study 1 addressed this shortcoming by using a forced-choice preferential-looking approach during which two videos were presented simultaneously, prompting infants to divide their attention between an interactive face-to-face scenario and a non-interactive back-to-back scene (see also Galazka et al., 2014).

From a developmental perspective, the findings from Study 1 suggest that infants' attentional orienting to face-to-face interactions over back-to-back scenes increases continuously from 7 to 13 months of age and that this preference exceeds chance level at around 9 months of age. Assuming that prioritized attention to social interaction scenes reflects a greater interest in these stimuli, this suggests that infants become increasingly motivated to observe others' interactions. Together with our finding that, during the same period, infants' orienting toward direct interaction partners undergoes a continuous increase, Study 1 reveals parallel trajectories in two conceptually different approaches to social attention: social orienting as prioritized visual attention in a screen-based eye tracking task, and social orienting as social attention behavior during naturalistic interaction. The parallel increase in both modalities at the population level speaks to the idea that the two measures captured developments of the same underlying construct, possibly an increasing motivation to orient toward own and others' social interactions.

Evidence for the intrinsic reward value of third-party social interactions ("social reward")

According to Chevallier and colleagues (2012), the social reward component of social motivation is manifested in seeking and liking of social engagement. In addition, Berridge and Robinson (2009) have proposed reinforcement learning as a third manifestation of social reward. At this definitory background, the findings from Study 2 address two out of three components of social reward. First, the results suggest that 13-month-old infants seek out situations in which they can observe a third-party interaction. This was achieved by restricting infants' immediate access to the target videos. In order to approach and anticipate the desired target video, infants had to fixate the associated cue shape (otherwise no video was displayed), recognize the perceptual difference between face-to-face and back-to-back target, encode the location of the face-to-face interaction target, detect and encode the association between cue shape and target location, and then recall all this information when seeing the social interaction predictive cue again. Infants' selective learning of cue-target associations guiding them to face-to-face interaction targets but not non-interactive back-to-back targets indicates attentional foraging of social interaction scenes (Manohar & Husain, 2013).

In addition to the seeking component of social reward, the results from Study 2 suggest that infants are biased to organize reinforcement learning around third-party interaction stimuli. In contrast to most previous studies on reinforcement learning in social interactions, the social

stimulus was used as the rewarding *outcome*, not as a cue guiding infants toward a rewarding non-social outcome (as, e.g., in Michel, Kayhan, Pauen, & Hoehl, 2021; Moore & Corkum, 1998; Vernetti, Senju, Charman, Johnson, & Gliga, 2018). This way, the findings from Study 2 present new evidence that the observation of a social interaction scene can work as a primary reinforcer in itself. The absence of learning for the non-interactive but social back-to-back targets emphasizes that this effect is specific to social relations in an interactive context. The mere presence of two human agents was not sufficient to reinforce infants' learning.

Taken together, the findings from Study 1 and Study 2 complement each other in showing that infants' attention and associative learning are driven toward seeking and approaching third-party interactions. This adds to the previous finding that 12-month-old infants show increased pupil dilation while observing third-party interactions (Gustafsson et al., 2016). To gain a more comprehensive understanding of social reward mechanisms, future studies should investigate whether infants experience hedonic pleasure while observing third-party interactions (Venezia et al., 2004). This could be achieved by measuring infants' facial expressions via Electromyography as a measure of emotional valence (Addabbo, Vacaru, Meyer, & Hunnius, 2020) or by measuring physiological responses aside from pupillary changes, for example, via heart rate or electrodermal activity (Wass, Clackson, & Leong, 2018).

Indication for infants' intrinsic motivation to learn

According to curiosity-driven learning theories, another motivational mechanism that may account for the findings of this thesis is infants' intrinsic motivation to acquire knowledge. As already mentioned in chapter 1, curiosity is also thought to elicit approach behavior and experience of intrinsic reward (Litman, 2005; Berlyne, 1966; Loewenstein, 1994). From this perspective, it would be possible that infants prefer attending to face-to-face over back-to-back scenarios (Study 1) and organize their learning to predict the occurrence of face-to-face targets (Study 2), because information-seeking motives may bias their attention toward situations with the highest available potential to impart information. Compared to a back-to-back arrangement, information transfer is more efficient, and the learning potential higher, when two agents face each other (Augusti et al., 2010). The idea that infants seek out situations in which they can gather knowledge is further supported by the results from Study 3, revealing that 9-month-old infants invested their memory and learning resources selectively in situations where two people engaged in a joint attention interaction, not in any of the other three conditions. This indicates that in the first year of life infants can evaluate the learning potential of an observed triadic situation, and they expect the highest learning benefit from observed joint attention.

Even though the findings in Study 3 speak to the idea that knowledge-seeking motives influence infants' attention and memory in observational learning contexts, it needs to be pointed out again that none of the studies in this thesis were explicitly designed to measure infants'

information-seeking motives. Instead, the focus was on the social-relational aspect of third-party interactions. To directly investigate the influence of information-seeking motives would require a systematic manipulation of the informative value of a learning opportunity. This could be achieved, for example, by varying characteristics of the interaction partners, such as their reliability as informants (Zmyj, Buttelmann, Carpenter, & Daum, 2010) or their competence and expertise (Begus & Southgate, 2012). Another future avenue would be to disentangle the relative impact of information-seeking motives and social motives on infants' visual attention behavior. Based on previous findings from the first-party interaction domain, it is likely that infants' attention is driven by an interplay between both factors (see also Manohar & Husain, 2013). For example, a study by Begus, Gliga, and Southgate (2016) has shown that 11-month-olds' preference for native speakers, which is often interpreted as a socially motivated preference, can be better explained by infants' underlying desire to acquire knowledge. Another study has shown that 12-month-olds' initiation of pointing gestures is primarily motivated by the expectancy to receive novel information rather than seeking a mutually shared experience (Kovács et al., 2014; see also Begus, Gliga, & Southgate, 2014).

To tease apart the relative influence of social motives and information-seeking motives on infants' attention and learning in third-party interaction contexts, one could set up a study design that tests the two motivational mechanisms against each other. One way to do so would be to systematically manipulate a third-party interaction scenario regarding its' social-emotional value on the one hand and its' informative value on the other hand. Would infants prefer attending to and learning from interactions with high informative value over interactions with high social-emotional value or vice versa? Or would the interplay between both factors represent the most preferred and beneficial learning option? Another possibility would be to manipulate the incentive value of the learning content, like the perceptual features or cultural relevance of an object, and compare it to the incentive value of the social context. Would infants prefer a social interactive scenario including a less attractive object over a non-interactive scenario including a highly attractive object? Or would a learning content with high incentive saliency overwrite infants' social preference?

5.2.2 Revealing Influential Factors That Promote Infants' Learning About Objects During Observing Third-Party Interactions

In addition to the novel perspective on motivational factors affording potential learning opportunities, this thesis provides novel insights into factors that promote infants' attention and memory during the actual learning situation. The main focus in this thesis, specifically in Study 3, was on influential factors within the observed social situation. In addition, the study findings provide insights into possible internal processes that may influence infants' attention and memory at the observer level. I will discuss both perspectives in the following two subsections.

Influential factors within the observed situation: Third-party joint attention promotes infants'
perceptual learning about objects

By systematically varying the gaze direction of two adults toward one another and toward a visible object, the findings from Study 3 demonstrate that observed joint attention provides a supporting context for 9-month-old infants' learning about novel objects. But what information did infants actually learn about objects embedded within an observed joint attention situation?

Since the objects were introduced in a silent and controlled environment with a neutral emotional context, they were not charged with additional meaning when infants saw them again in the preferential-looking phase (Becchio, Bertone, & Castiello, 2008). More specifically, there was no auditory or visual information systematically mapped to the target objects during the encoding phase, such as object-related emotional attitudes, actions, or word labels. Therefore, it is reasonable to assume that infants' novelty preference reflected their "pure" recognition of the visual object features such as its color and shape. Moreover, since the target objects were presented at different locations in the preferential-looking phase than in the encoding phase, the findings from Study 3 imply that infants encoded the object's identity rather than its location. As discussed in the corresponding study chapter, this speaks to the idea of a third-party communication-induced memory bias (Okumura et al., 2016; Yoon et al., 2008). However, future studies are needed to investigate this possibility systematically by examining whether infants learn different kinds of information from third-party communicative compared to non-communicative contexts.

As explained in more detail in chapter 1, the encoding and recognition of visual object features is foundational for more complex and later emerging forms of object-related learning in observed joint attention contexts, such as the learning of novel object labels (Fitch et al., 2020) or the imitation of object-directed action routines (Herold & Akhtar, 2008). In this light, the findings from Study 3 indicate that before infants typically possess the behavioral or language skills to engage in more complex forms of object-related learning in third-party settings, they possess fundamental abilities required to learn about objects through observing others' joint attention. As discussed in more detail in Study 3, this is particularly noteworthy since, around the same age, infants increasingly engage in and learn from direct joint attention in first-party settings (Experiment 2 in Study 3; Cleveland & Striano, 2007).

Influential factors at the observer level: Indication for a learning-promoting "face-to-face mode"

What internal processes increase infants' object encoding during observed joint attention? One possibility would be that the infant looks longer at an object when seeing two mutual gaze-cues pointing toward it and enhancing its perceptual saliency. However, as discussed in more detail in chapter 4, the data from Study 3 speak against the possibility that infants' encoding outcomes relied on their overt visual attention to the object. Most strikingly, infants' novelty preference

score was independent of whether they had looked at the object at all during the encoding phase. This aligns with findings from previous object encoding studies, such as studies in which the duration of the encoding phase was too short to even allow infants to initiate and perform an overt gaze movement toward a gaze-cued target object. For example, Reid and colleagues (2004) found that infants kept fixating a centrally displayed actor throughout a 3-second encoding phase. Nevertheless, infants showed an increased novelty preference in a subsequent preferential-looking test (see also Wahl et al., 2013). Additional evidence comes from studies with live-interactive setups, during which the experimenter explicitly controlled that each individual infant had looked at the object for a pre-defined duration of 20 seconds before moving on to the subsequent paired-preference test (Cleveland & Striano, 2007, 2007). Despite the identical duration of direct attentional exposure during familiarization, infants' object encoding was only enhanced in these studies if they had seen an adult looking at the object after directly gazing in the direction of the infant. In summary, the above-outlined findings rule out the possibility that differences in overt visual attention cause the effect of joint attention on infants' object encoding.

Given that infants must perceive an object in order to encode its' visual features, it is conceivable that observed joint attention shifts infants' *covert* attentional focus toward it. Since the focus in this thesis was exclusively on measures of overt visual attention, the following argumentation is merely speculative and needs to be validated with measures of covert visual attention in future studies (Christodoulou, Leland, & Moore, 2018). One possibility would be that the inferred social relationship from an observed face-to-face setting puts infants in a state of heightened responsiveness—similar to when they are directly addressed via eye contact (Grossmann, Johnson, Farroni, & Csibra, 2007). This would be in line with a model by Puce and colleagues, according to which the (adult) brain switches between a "default mode" and a "socially aware mode", depending on the social saliency of a given input (Puce et al., 2015). In the non-social default mode, information is assumed to be processed automatically and without conscious awareness. During the socially aware mode, in contrast, top-down processes are thought to enable more conscious information processing. Inspired by Puce and colleagues' idea of a socially aware mode, it would be possible that an observed face-to-face arrangement between two people switches on a "face-to-face mode", which sets infants in a state of heightened responsiveness toward learnable content. This would further align with the *social gating hypothesis*, which claims that the infant learning system is particularly sensitive and responsive to information presented within a social context (Kuhl, 2007; Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Sage & Baldwin, 2011). In contrast to direct pedagogy, social gating is considered a more general learning-enhancing mechanism that enhances memory and learning in social contexts without others' instructive intentions required (for a review, see Sage & Baldwin, 2010).

There is some evidence from previous studies supporting the idea that general learning-enhancing mechanisms like social gating or a socially aware mode may modulate infants'

responsiveness to learnable content when observing a third-party interaction. For example, the previously discussed finding of 12-month-olds' increased pupillary response toward face-to-face versus back-to-back interactions could also indicate a state of focused attention (Gustafsson et al., 2016). Moreover, observed face-to-face arrangements have been shown to promote implicit learning mechanisms, which play an important role in scaffolding early learning: 12-month-old infants can extract statistical regularities of gesture sequences in the presence of face-to-face but not back-to-back contexts (Quadrelli, Monacò, Turati, & Bulf, 2020), and 13-month-old infants can form cue-target associations in face-to-face but not back-to-back contexts (Study 2 in this thesis).

According to the here proposed face-to-face processing mode, observing others' interactions may actuate an internal attentional stance that facilitates infants' encoding and learning of novel information. Importantly, observed face-to-face context does not increase infants' general responsiveness toward everything available in their visual range. In Study 3, infants only showed increased object encoding when two people looked at the object while facing each other, not when two people looked away from the object in a face-to-face arrangement. If observed face-to-face context would elicit a generally increased responsiveness, infants should have shown increased object encoding in both face-to-face conditions. Instead, the face-to-face mode seems to canalize infants' covert attention selectively to information worth learning by organizing or "gating" their responsiveness toward informative cues.

5.2.3 Deepening Our Understanding of the Ontogenetic Origins of Infants' Attention to and Learning From Third-Party Interactions

In addition to the novel perspective on motivational mechanisms and on factors promoting infants' learning about objects during third-party observation, this thesis contributes to our understanding of the ontogenetic origins of infants' attention to and learning from others' social interactions.

Onset age of infants' memory and learning abilities in the context of third-party interactions

The first contribution relates to the onset age of infants' ability to learn from observing others' interactions. As outlined in more detail in chapter 1, the previous research suggested the following developmental progression: From 6 months onwards, infants can use others' body, head, and gaze orientation toward one another to recognize the presence of a social relationship between two human agents (Augusti et al., 2010; Handl et al., 2013). During the following 6 months, infants develop an increasing understanding of the turn-taking dynamic of interpersonal interactions, with an increasing specialization for animate interaction partners (Beier & Spelke, 2012; Molina et al., 2004). By their first birthday, infants begin to infer others' communicative intent within an observed interaction (Yamashiro & Vouloumanos, 2018), and they develop an understanding of others' collaborative action goals (Elsner et al., 2014; Thorgrimsson et al., 2015). Evidence for

object-related learning through observing third-party interactions had been earliest demonstrated in the second year of life, at around 18 months of age (Akhtar, 2005; Matheson et al., 2013). The findings from Study 3 substantially extend this research by demonstrating that the ability to learn about novel objects through observing third-party interactions starts to emerge in the first year of postnatal life. Already at 9 months of age, infants can use observations of third-party interactions to organize their attention and memory in favor of encoding, storing, and recalling the visual perceptual features of novel objects.

Similarities between first- and third-party perspective on social interactions

Furthermore, this thesis addresses parallels between the first- and third-party perspective on social interactions. Table 5.1 summarizes the central findings of this thesis in direct comparison to the corresponding findings from the previous first-party interaction literature, which have been partly replicated in this thesis.

Table 5.1

Summary of the findings from this thesis (focusing on the third-party perspective) in direct comparison to corresponding findings from the previous first-party interaction literature

	First-Party Interactions *Previous Findings*	Third-Party Interactions *Findings in this Thesis*	Possible Underlying Influential Factors
Study 1	During the second half of the first year of life, infants become increasingly motivated to engage in social interactions with others *(Carpenter et al., 2017; Striano & Rochat, 1999).**	During the second half of the first year of life, infants increasingly orient their attention toward third-party interactions.	Social orienting and social reward as motivational mechanisms driving infants toward situations in which they can engage in or observe others' social interactions.
Study 2	Toward the end of the first year of life, infants find it intrinsically rewarding to engage in social interactions with others *(Venezia et al., 2004; Clearfield et al., 2008).*	Proximate reward mechanisms may account for infants' attentional orienting toward third-party interactions as well.	
Study 3	9-month-old infants coordinate their attention with others in joint attention episodes, and they can make functional use of these situations to about novel objects *(Cleveland & Striano, 2007).**	Around the same age, infants can learn about visual object features through merely observing joint attention interactions between others.	Joint attention increases infants' learning about objects in the context of first- and third-party social interactions.

Notes. The cited studies in column "First-Party Interactions" represent study examples outlined in more detail together with further evidence in the general introduction (chapter 1), as well as in the corresponding study chapters (chapters 2–4). * Previous findings replicated in this thesis.

The overall comparison reveals that, at the population level, developments in the third- and first-party interaction domain share striking similarities. This raises the possibility that similar factors may underpin infants' attention and learning in observational and self-experienced social interactive settings. Furthermore, the overall picture supports the view that the second half of the first year of life represents a critical period in infants' social development. In combination with

previous findings, this thesis suggests that infants in the second half of the first postnatal year develop abilities and preferences (a) guiding them toward own and others' social interactions and (b) enabling them to use both situations to acquire knowledge about their environment. At a broader level, this enables infants to flexibly take the most out of their social learning environment, which largely varies regarding the extent of direct teaching and observational learning opportunities. I will elaborate more on this variation in the following sections.

5.3 Limitations and Future Directions

The findings obtained from this thesis need to be considered against some limitations. Methodological constraints resulting from the paradigm or the stimuli are discussed in the discussion sections of the individual study chapters. The current section focuses on empirical limitations concerning the aim of this thesis. What needs to be studied in the future to further improve our understanding of infants' attention to and learning from third-party social interactions? In the following sections, I will point out five central limitations of this thesis and use them as a basis to derive follow-up questions that should be addressed in future research.

5.3.1 Investigating Immediate Relations Between First- and Third-Party Perspective

The overall findings of this thesis reveal parallel developments in infants' social attention and learning in direct social interactions and third-party observational contexts (see Table 5.1). Throughout this thesis, I have interpreted this pattern in terms of broader underlying mechanisms that may influence developments in both modalities. However, it remains unclear whether and, if yes, how the two perspectives are immediately related to one another. Does infants' own practical experience with a particular interaction influence how infants perceive and learn from similar interactions between other people and vice versa? Or is self-experience possibly *required* for infants to make inferences about an observed interaction? Even though we compared infants' third-party perspective with their first-party perspective in Study 1 and Study 3, neither study was conceptualized to investigate these questions. First of all, both studies were based on cross-sectional study designs. Moreover, in Study 3, the first- and third-party conditions were tested between subjects. Even though this was not the case in Study 1, conclusions about immediate dependencies are restricted in this study as well because the focus was on conceptually different interactions in the first- and third-party task. While in the eye tracking task infants observed third-party *dyadic* interactions, the free play focused on first-party *triadic* interaction. In addition, and more generally, we did not examine whether infants had own prior experience with any of the presented third-party interactions.

Only a few previous studies have investigated the specific underlying mechanisms of infants' perception and understanding of their own and others' interactions. A study by Gredebäck

and Melinder (2010), for example, has demonstrated that infants' perception of a third-party feeding interaction is related to infants' own experience with being fed. In the study, 6- and 12-month-old infants observed an adult feeder moving a spoon directly to the mouth of another adult. Twelve-month-old infants, but not 6-month-olds, fixated the recipient before the spoon arrived, indicating that 12-moth-olds can infer and anticipate others' action goals within an observed interaction. At the individual level, this ability was positively correlated with infants' own experience being fed, in that infants with more feeding experience performed more anticipatory gaze shifts in accordance with the observed feeding interaction. In the same study, both 6- and 12-month-old infants showed an increased pupil dilation response while observing a non-rational feeding interaction during which the feeder placed food on the back of the recipient's hand, compared to a rational feeding interaction during which the feeder placed the food in the recipient's mouth. The authors interpreted the increased pupil size as an indicator of surprise and concluded that already 6-month-old infants could evaluate the rationality of an observed interaction. As infants' pupillary response was not related to their own feeding experience, the authors further concluded that rationality evaluation might require less experience than the anticipation of others' action goals.

The findings by Gredebäck and Melinder (2010) suggest a correlational link between infants' first-hand experience with a particular interaction and their understanding of observed similar interactions. However, the direction of this relation remains unclear based on their findings. To gain an insight into causal relations, Henderson, Wang, Matz, and Woodward (2013) used a training study approach during which they systematically manipulated infants' own experience via training. The study revealed that 10-month-old infants only inferred others' collaborative action goals in a third-party interaction after receiving own active training with the same collaborative activity (i.e., retrieving a toy collaboratively out of a box). By training infants at an age at which they typically do not engage in goal-directed joint activities themselves, the authors could demonstrate that infants' understanding of third-party collaborative action goals causally depends on their own manual experience with the observed activity. This finding speaks to the broader assumption of embodied accounts, which propose that first-hand experience with a manual action directly contributes to infants' perception and understanding of this action from an observer perspective (Flanagan & Johansson, 2003; Sommerville, Woodward, & Needham, 2005; Gredebäck, Gottwald, & Daum, 2021). Adapting this assumption to action sequences in a coordinated interactive context, it would be possible that infants' understanding of third-party coordinated action goals develops as infants produce coordinated and goal-directed actions themselves. Indirect support for this assumption comes from previous studies showing that the capacity to attribute shared goals to an observed interaction emerges during the same period as children start to engage in goal-directed joint activities themselves (Fawcett & Gredebäck, 2013, 2015; Krogh-Jespersen, 2020).

A recent study by Begus, Curioni, Knoblich, and Csibra (2020) challenged the view that active motor experience is necessarily required for infants to infer others' shared action goals (see also "the teleological stance"; Gergely & Csibra, 2003). Instead of using complex action sequences between two human agents, 9-month-old infants were presented with simple motion actions of geometrical shape agents. Infants showed increased cortical activity in the posterior superior temporal sulcus (pSTS) in response to a joint action scenario (two geometrical-shape agents moving toward the same target), but not in response to an individual action scenario (only one of two agents moving toward a target). Referring to previous studies showing that pSTS activity is involved in coordinated social interaction processing versus independent action processing (Isik et al., 2017), the authors interpreted their finding as showing that infants attributed a shared goal to the observed joint action. Based on the assumption that infants do not engage in goal-directed joint activities before the second year of life, the authors further proposed that infants can infer shared goals before they typically engage in such activities themselves. At a broader level, this raises the possibility that infants may be equipped to gather information through observing third-party interactions before having incorporated the observed activity within the own behavioral repertoire.

In addition to their own practical experience, observational learning from third-party interactions is influenced by developments in the social-cognitive domain, such as infants' emerging ability to imagine themselves as a participant in an observed interaction ("self-other equivalence", Moore, 2007; Matheson et al., 2013; Herold & Akhtar, 2008). It remains unclear to which extent developments in infants' imaginative ability interact with their practical experience in terms of influencing the immediate relation between own interaction behavior and perception of others' interactions. Future studies are required to systematically investigate this question and to explore the influence and interrelatedness of other influential third factors. This could be most comprehensively achieved by using longitudinal study designs. In addition to deepening our understanding of the immediate relation between first- and third-party perspective, longitudinal study designs would furthermore allow us to investigate relations between attentional, motivational, and cognitive developments within the third-party domain—for example, the relation between infants' early sensitivity to social interactions and their later observational learning abilities.

Another interesting future avenue would be the application of hyperscanning methods in third-party observational contexts. Hyperscanning involves the simultaneous recording of the brain activity of multiple persons. Previous first-party studies using this method have revealed that communicative rhythms increase interpersonal synchronization of brain activity during parent-infant interactions (for a review, see Nguyen, Bánki, Markova, & Hoehl, 2020). Moreover, stronger interpersonal synchrony has been found to predict infants' learning about novel objects during social interactions with their parent (Leong & Schilbach, 2019). It would be interesting to

explore whether infants show similar effects in an observational setting—that is, increased neural synchronization with third parties who engage in a contingent social interaction over third parties who engage in a social situation with a disrupted communicative rhythm.

5.3.2   Considering the Infant's Natural Learning Environment

Another limitation is that all studies in this thesis were conducted in a highly controlled lab environment. Except for the free-play task in Study 1, all measures were based on stationary eye tracking tasks during which the infant watched pre-recorded videos of social interaction scenes on a screen. Moreover, the room setup prevented infants from any visual or auditory distraction. The high internal validity resulting from this setup came at the expense of external validity, meaning that the current findings cannot directly inform us about infants' social attention and learning in their natural environment.

One central limitation regarding the ecological validity of the included studies concerns the selection of third-party interactions presented to the participants. All interactions were silent, precisely timed, and perceptually highly controlled. While these features increased the internal validity of the stimuli, they diverge from the interactions infants typically encounter in their environment. Real interactions are much more dynamic and complex, contain emotion expressions, perceptual stimulation in multiple modalities, and are embedded within a cluttered environment. Therefore, an important future avenue would be to examine whether the findings of this thesis generalize to infants' natural learning environment (see also Hoehl & Bertenthal, 2021). One way to do so would be to measure infants' response toward situations in which two people interact naturally. Naturalistic interactions could be either pre-recoded and then presented on a screen, or they could be performed live in front of the infant. A similar approach has been used by Correa-Chávez and Rogoff (2009) with older children. In their study, 5- to 11-year-old U.S. children and Guatemalan Mayan children observed an interaction between their older sibling and an experimenter. During the interaction, the experimenter demonstrated to the sibling how to construct a novel toy. The authors measured the attentional interest of the observer child toward this third-party demonstration by coding their looks in the direction of the interaction. The observational learning success of the observer child was measured by letting them create the toy they had previously observed in the third-party demonstration toward their sibling.

The study design by Correa-Chávez and Rogoff represents a practical example of how to measure attention and learning while observing a semi-naturalistic interaction. However, what remains unclear from their design is whether the interaction matches with the kinds of interactions an individual actually encounters in their daily environment. One way to answer this question would be to ask the parents to rate their child's experience with a certain interaction (Gredebäck & Melinder, 2010). A more direct approach would be to measure the infant's real-life experience in their home environment, for example, by using manual observations via parent

diaries or experimenter visits. Alternatively, technological solutions could be used, such as stationary cameras recording the overall home environment, or technically more sophisticated devices like audio recorders with an integrated body camera attachable to the infant's clothes (Bergelson et al., 2019; Casillas, Brown, & Levinson, 2019). The huge advantage of using advanced technological methods like this is that it would allow collecting data about the infant's environment without any observer bias. More specifically, one could gather information about multiple domains of the infant's objective learning situation, including the frequency of potential learning opportunities, the kinds of interactions infants encounter, or the interaction partners they observe.

To examine infants' perception of their social and physical environment, mobile eye tracking systems could be used (Abney, Suanda, Smith, & Yu, 2020; Franchak, Kretch, Soska, & Adolph, 2011; Yu & Smith, 2013, 2017). Video recordings of mobile eye trackers would document *what* the infant sees from an egocentric view and provide insights into *how* they look at certain target events like third-party interactions. By using a longitudinal study design with a multi-method approach, it would also be possible to examine infants' learning from third-party interactions. For example, one could use mobile-eye tracking recordings to identify objects embedded within third-party interactions. Based on this information, one could investigate how the infant reacts to these objects after observing them in a third-party interaction. If the infant has learned something about a certain object, they may attend to it more frequently and longer, or imitate actions that another person has previously performed on it. In addition, object-directed vocalizations measured via voice recordings could be used as an indicator of infants' enhanced responsiveness toward an object (Goldstein, Schwade, Briesch, & Syal, 2010). Naturalistic observations like these would furthermore carry the potential of informing us about longer lasting effects on infants' memory and learning.

One crucial reason why it is relevant to study infants' attention and learning in their everyday environment is the significant inter-individual variation in infants' daily social experience (Rogoff et al., 2018). For example, a child growing up in an extended family with multiple siblings will have more opportunities to observe interactions between others than an infant growing up in a single-parent household with no siblings. Moreover, an infant who is frequently involved in face-to-face interactions will be more accustomed to direct pedagogical settings compared with an infant receiving less direct attention from adults in their environment.

As also mentioned in the empirical part of this thesis, the relative extent to which infants are accustomed to direct interaction and intent observation shows systematic variation across cultural contexts (Gaskins & Paradise, 2010; Rogoff et al., 2018). More specifically, infants' learning environment is strongly shaped by their caregivers' parenting goals and socialization practices, which are influenced by the culturally shared beliefs about childhood and rearing (Keller, 2018; Keller, Yovsi, et al., 2004). In many communities, like in Mayan communities, children are strongly

encouraged and expected to rely on observational learning (Mesman et al., 2018). As a consequence, infants spend much time intently observing their environment (Lancy, 2010; Odden & Rochat, 2004). This contrasts with the daily experiences infants make when growing up in child-centered societies, in which direct face-to-face interactions are valued as the primary source of learning (Tamis-Lemonda & Song, 2012). Relatedly, in cultures focusing on the welfare of the group rather than on the welfare of individuals, parents show less sensitive responsiveness toward their child and engage them less in direct face-to-face interactions as compared to cultures where the infant is seen as an autonomous being whose wishes and goals need to be satisfied (Keller, 2013). Gaskins and Paradise (2010) provide a comprehensive overview of cross-cultural differences in infant's daily experiences. It would be a critical direction for future research to examine explicitly how determinants of early social experience (e.g., socialization goals and practices, parenting goals and practices) shape infants' attention to and learning from third-party interactions.

5.3.3 Comparing the Efficiency of Infants' Learning From First- and Third-Party Interactive Contexts

Another limitation of this thesis is that it does not provide a comprehensive insight into how infants' observational learning from others' interactions compares to their learning in direct teaching contexts. According to the results in Study 3, 9-month-old infants' perceptual learning about objects is equally efficient when the infant observes a joint attention interaction between two adults as when the infant is directly addressed by an adult. However, the implications of this finding need to be put into perspective with restrictions resulting from the study paradigm. One restriction is that the screen-based first-party scenario was not representative of real infant-adult interactions, during which both partners engage in reciprocal social and emotional exchange. It is possible that a live teaching context with a sensitive and responsive interaction partner elicits superior learning compared to an observational setting. Another restriction is the focus on object encoding as the dependent measure. While object encoding represents an age-appropriate measure of memory and a necessary basis for more complex forms of learning, it does not inform us about infants' full-fletched learning of culturally relevant information in their environment. Even if infants encode and later recognize objects with similar efficiency in both contexts, it would be possible that the learning of more advanced information profits more extensively from direct teaching compared to observational learning.

As outlined in more detail in chapter 4, infant-directed interactions scaffold infants' learning about objects, especially joint attention. However, studies examining the isolated impact of child-directed cues on early word learning suggest that the learning of novel object labels is equally efficient in observational contexts: 18-month-old toddlers can learn novel object labels equally well when they are directly addressed by an adult as when they overhear a third-party conversation (Floor & Akhtar, 2006; Gampe et al., 2012). In contrast to word learning, previous

studies on action imitation have revealed that toddlers around the same age profit from pedagogical input when learning an object-related action. Even though 18-month-old toddlers are effectively able to learn novel object-related actions by observing a third-party demonstration, they are more likely to imitate actions and do so with higher accuracy if the demonstrator addresses them directly in a pedagogical context (Matheson et al., 2013; Shneidman, Todd, & Woodward, 2014).

To better understand *how* pedagogical context facilitates the efficiency of learning action sequences, Sage and Baldwin (2011) conducted a study in which they tested two alternative hypotheses: one being that pedagogical context shapes the processing of an observed event and the other one being that pedagogical context facilitates the learner's ability to incorporate the previously processed information in their behavior. To tease these two possibilities apart, Sage and Baldwin measured 13-month-old infants' processing and their own subsequent reproduction of a simple tool use sequence during which a hook tool was used to pull an out-of-reach object in reaching distance. The experimenter demonstrated the tool use sequence in either a social-pedagogical context (including child-directed speech, eye contact, and referential gaze) or a non-pedagogical but social context (including equivalent enthusiastic verbalizations and emotions but no communicative cues in the direction of the child). The study revealed that 13-month-old infants processed the tool use sequence equally well in both conditions, indicated by longer looking times to a causally possible over a causally impossible version of the observed action sequence in a subsequent preferential looking test. Infants' reproduction behavior, in contrast, differed between the pedagogical and non-pedagogical social condition. Infants showed a superior reproduction of the tool use sequence when learning it through a pedagogical as compared to a non-pedagogical demonstration.

Based on these findings, Sage and Baldwin (2011) concluded that pedagogical context shapes infants' ability to *utilize* encoded information rather than their *processing* of the event. In line with the *Natural Pedagogy* account, the authors argued that direct communicative cues induce a special stance, during which infants are increasingly motivated to incorporate the processed information into their own behavior because they perceive the communication as a "special opportunity to acquire generalizable knowledge" (p. 149, Csibra & Gergely, 2009). Furthermore, the study by Sage and Baldwin (2011) provides evidence for the social gating hypothesis: Infants' processing of the tool use sequence was only enhanced in the social conditions (pedagogical and non-pedagogical condition), not in an additional "non-social" condition during which only the hand of the demonstrator was visible. This indicates a higher receptiveness to novel input in social contexts. Even though the study by Sage and Baldwin (2011) did not include a third-party interaction condition, the findings highlight the importance of complementing visual encoding measures with additional outcome measures to systematically compare the learning outcome from first- and

third-party interactive settings. Moreover, the findings speak to the mechanisms that may underlie higher action imitation in direct pedagogical versus observation-based contexts.

From a third-party perspective, it would be interesting to follow-up on the findings by Sage and Baldwin by comparing infants' observational learning in a non-pedagogical social context with a third-party interactive context. Would infants learn "better" from a third-party interaction compared to a non-pedagogical social context? There is reason to assume that observed social interactions represent a richer learning opportunity compared to observed individual social contexts. The contingent and coordinated exchange of attention and actions that constitutes social interaction provides a much higher density of social cues compared to situations in which a single individual performs a target action alone, without addressing anyone. As a consequence, social interactions may guide infants' attention and learning more efficiently than non-interactive social contexts (see also Matheson et al., 2013). Following this assumption, it is likely that infants would process the same learning content better when it is embedded within a third-party interaction. This would be further in line with the face-to-face processing mode proposed in section 5.2.2. Moreover, based on previous studies revealing that at least 18-month-olds can utilize information they have processed in an observed interactive context (Floor & Akhtar, 2006), it is reasonable to assume that a third-party interaction condition would elicit a similar learning response as the pedagogical context condition in the study by Sage and Baldwin—both in terms of increased processing and reproduction behavior. Another aspect that needs to be considered when comparing the learning efficiency in an interactive with a non-interactive context is the content to be learned. Some cultural knowledge, such as interactive greeting rituals, inherently depends on social exchange and can therefore only be imparted in a social *interactive* context. This raises the importance to consider the learning content as another relevant dimension.

Another essential finding from the action-imitation literature is that the supporting effect of direct pedagogical context decreases throughout the first two years of life. By 24 months of age, toddlers imitate actions equally likely when a demonstrator addresses them directly than when they observe a third-party demonstration toward another person (Matheson et al., 2013; Nielsen, 2006). This increasing independence of direct teaching cues cannot be explained by developments in attentional control, as toddler's action imitation has been found to be independent of the time toddlers had looked at the action demonstration—both at 18 and 24 months (Shneidman et al., 2014; see also Sage & Baldwin, 2011). Social cognitive approaches to observational learning propose that the effect of age originates from developing social cognitive skills, such as toddlers' emerging ability to imagine themself in an observed interaction or to understand others' intentions without self-experienced shared attention (Matheson et al., 2013).

Another assumption of social cognitive accounts is that infants first rely on shared attention to infer others' intentions, as the close matching between their own and others' attentional focus facilitates this inference. Supporting evidence comes from a study by Moll and Tomasello (2007).

In their study, 14-month-old infants played with two objects in the presence of an adult experimenter, who established a joint visual context with the infant and each of the two objects. Then, the experimenter left the room while the infant played with a third object (the target object) and a second experimenter (the "assistant"). When the first experimenter returned, looked in the direction of the three objects, and said excitedly, "Look! Can you give it to me?", 14-month-old infants gave her the target object, that is, the object she did *not* know. This indicates that infants considered the experimenter's previous experience. In a matched third-party context, in which the experimenter interacted with the assistant and the two familiarization objects, 14-month-old infants did not give the experimenter the target object when she returned, indicating that they did not consider the experimenter's previous experience. At 18 months, in contrast, toddlers gave the target object to the experimenter even after observing a solitary setting during which the experimenter inspected the two familiarization toys on her own without looking in the direction of the child (Moll & Tomasello, 2007). From a developmental perspective, this supports the idea that infants first become able to determine what other people know based on their direct joint interactions with them. In sum, the previously outlined findings suggest that the reliance on direct pedagogical context over third-party observational context decreases throughout development and that it differs between different domains of knowledge (e.g., word learning, action imitation). Tying this back to the findings of this thesis, it would be interesting to examine whether individual variation in infants' early sensitivity to and perceptual learning from third-party interactions predicts their later developing propensity to acquire more complex knowledge through observing others' interactions.

There is some empirical evidence indicating that children develop different attention and learning strategies depending on their environmental experience. For example, the ability to learn words through overhearing has been shown to be influenced by the amount of time toddlers are surrounded by multiple adults in their daily lives: The more time 20-month-old toddlers spend with multiple adults, the better they are at learning words through overhearing others' conversations (Shneidman, Buresh, Shimpi, Knight-Schwarz, & Woodward, 2009). This finding lends further support to the idea that in cultural environments where infants are typically surrounded by multiple adults in their daily lives, infants may develop abilities aiding them to efficiently acquire novel information via intently attending to others' interactions. This could be, for example, contexts in which a network of multiple caregivers is involved in infant care ("allomaternal care", Burkart, Hrdy, & Van Schaik, 2009) or contexts in which the mother carries her infant on her back while engaging in daily work routines amongst other adults.

Empirical evidence for cultural differences in attention and learning from third-party interactions comes from the above-described study by Correa-Chávez and Rogoff (2009), in which 5- to 11-year-old children observed a semi-naturalistic interaction between their older sibling and an adult experimenter. Compared to U.S. children, Guatemalan Mayan children showed more

sustained attention while observing the third-party interaction and demonstrated superior observational learning outcomes. Moreover, U.S. children made more attempts to attract the experimenters' attention during the observation. The authors discussed their findings at the background of ethnographic information about the children's socio-cultural environment, arguing that Mayan children, who are raised to rely on observational learning, may be better equipped to learn from third-party observation compared to U.S. children, who are much more accustomed to learning through direct teaching and face-to-face interactions (see also Shneidman, Gaskins, & Woodward, 2016; Shneidman & Woodward, 2016). It would be an important direction in future research to investigate the influence of cultural context on observational learning further, especially in the *first* year of life where evidence is lacking. It requires systematic comparisons from early on and throughout development to disentangle the relative influence and the interplay between biological predispositions and early social experiences on early attention and learning.

5.3.4   Considering Individual Characteristics of the Infant

Another limitation of this thesis is that it did not consider inter-individual variation in individual infant characteristics. One factor that may influence infants' attention and learning in observational contexts is infant attachment style. The influence of attachment style on social attention has been mainly studied in first-party contexts. For example, 12- to 15-month-old infants with a disorganized attachment style have been found to perform significantly less initiating joint attention behaviors than infants with a secure or insecure attachment style (Claussen et al., 2002). It remains unclear whether infants' attachment style also influences their social attention in observed interactive contexts. However, a previous study with geometrical shape agents supports the possibility that infants' socio-emotional experiences with their caregivers may influence their social information processing in third-party contexts (Biro, Alink, Huffmeijer, Bakermans-Kranenburg, & van IJzendoorn, 2015).

A further influential factor is infant shyness—an early developing form of shyness that typically emerges toward the end of the first year of life and shows a significant inter-individual variation (Putnam & Stifter, 2002; Schaffer, 1966). The previously outlined findings by Shimpi and colleagues (2013) raise the idea that infants' wariness of strangers, an integral component of infant shyness, may be related to their learning in third-party contexts. More specifically, it would be possible that the observation of others' interactions represents a particularly beneficial learning context for shy infants, as it may help them to satisfy two competing motivations: the desire to approach social interactions and the simultaneous tendency to avoid social situations ("approach-avoidance conflict", Coplan, Prakash, O'Neil, & Armer, 2004). Observing an interaction between two strangers may allow shy infants to pursue their intrinsic social motivation while avoiding the feared experience of engaging with a stranger. Supporting evidence comes from a study revealing that 7- to 13-month-old infants with high shyness scores show an attentional preference for faces

of strangers with averted gaze compared to faces of strangers with direct gaze. Infants with low shyness scores, in contrast, show the opposite preference for strangers gazing in their direction (Matsuda, Okanoya, & Myowa-Yamakoshi, 2013). It remains to be studied whether shy infants prefer attending to third-party over direct interactions and whether this preference is accompanied by preferential learning through observation. Furthermore, it would be interesting to examine whether shy infants adjust their willingness to learn from a stranger after observing them interacting with another person. It would be possible that observational experience with a stranger degreases infants' wariness of this person and, consequently, increases their readiness to engage with and learn from them in the future. Indication for such an indirect experience effect comes from a previous study showing that 15-month-old infants show an increased willingness to play with a stranger after observing this person in a positive interaction with the own mother or another stranger (Feiring, Lewis, & Starr, 1984).

While the focus in this thesis was on typically developing infants, another essential future path will be to study how infants with autism spectrum disorder (ASD) attend to and learn from third-party interactions. Investigating this question would be especially relevant given the large body of research demonstrating that already in the first year of life, infants later diagnosed with ASD show significant impairments in social attention and social motivation (Dawson et al., 2004; for a review, see Bruinsma, Koegel, & Koegel, 2004). Twelve-month-old infants with ASD are, for example, much less attracted by social stimuli compared with typically developing infants at this age (Osterling & Dawson, 1994). Moreover, studies focusing more specifically on social attention in social interactive settings have revealed that individuals with ASD show major deficits in essential components of joint attention: Compared to their typically developing counterparts, infants and toddlers between 6 and 24 months of age avoid eye contact (Adrien et al., 1993), look away from faces gazing in their direction (Chawarska & Shic, 2009), engage with lower frequencies in referential gaze shifts and intentional communicative gestures like pointing or showing (Wimpory, Hobson, Williams, & Nash, 2000), and engage less in behaviors indicating hedonic liking of joint play (Mundy, Sigman, Ungerer, & Sherman, 1986). Interpreting these findings within the overall framework of this thesis, infants with ADS lack in foundational requirements for learning through social interactions: they show impaired attentional orienting to social partners (depriving them of accessing social interaction situations) and, during the interaction itself, impaired social-communicative behaviors. Especially early impairments in joint attention skills have been discussed as a predictor for later impairments in cognitive development, including language acquisition (Mundy, Sigman, & Kasari, 1990) and theory of mind (Baron-Cohen, Leslie, & Frith, 1985).

It requires future studies to investigate how infants with ASD represent and learn from third-party social interactions. Some evidence comes from studies with toddlers and preschoolers with ASD, suggesting an impaired attentiveness to face-to-face interactions. For example, 18-

month-old toddlers with ASD have been found to perform more gaze shifts between two objects than between two human agents (Swettenham et al., 1998). Moreover, when observing a conversation between two people, 4-year-old children with ASD look shorter than typically developing 12-month-old infants at the speaking person and they do not anticipate the turn-taking interactive dynamic. In contrast, when seeing videos of two objects taking turns in making motor sounds, preschoolers with ASD show similar looking times and saccade patterns as typically developing infants (von Hofsten et al., 2009). Taken together, these findings indicate that infants with ASD may be attracted less by the social component of observed interactions than typically developing infants.

5.3.5 Focusing on Characteristics of the Interaction Partners and the Learning Content

Another limitation of this thesis is that the focus was exclusively on third-party interactions between adults, specifically unfamiliar adults. There are several reasons why it would be important to address this limitation in future research. One reason is that infants' environmental experience with others' interactions is not restricted to interactions between adults. It also includes peer-to-peer interactions or interactions between adults and peers. Moreover, infants in the first year of life presumably predominantly observe interactions between familiar individuals like their caregivers or siblings. Thus, to better account for the multiple facets of infants' social environment, it requires studies investigating infants' attention to and learning from interactions between different interaction partners. For example, it would be possible that infants with higher regular peer contact differ in their attention to and learning from observed peer interactions compared to infants with lower peer contact (Seehagen & Herbert, 2011).

The second reason is an empirical one. Previous studies on action imitation in first-party interactive contexts have revealed that both the age and familiarity of an interaction partner modulate toddlers' learning. Regarding the influence of familiarity, both short-term familiarity (induced via warm-up with the experimenter, Slaughter, Nielsen, & Enchelmaier, 2008) and long-term familiarity (mother versus stranger, Seehagen & Herbert, 2012; but see Devouche, 2004) have been found to increase imitation of arbitrary novel object actions. Moreover, Kinzler and colleagues (2012) have shown that 10-month-old infants preferentially approach objects introduced by an adult speaking their native language compared to a foreign speaker. Even though Kinzler and colleagues did not test learning directly, their findings lend support to the idea that infants prefer to learn about novel objects from culturally knowledgeable teachers. Regarding the influence of the demonstrator's age, previous studies have suggested that infants learn different *kinds* of information from peer and adult demonstrators. In a study by Zmyj, Daum, Prinz, Nielsen, and Aschersleben (2011), for example, 14-month-old infants were more likely to imitate novel and arbitrary actions like touching a lamp with the forehead from an adult model, while they learned

familiar actions like pulling objects apart preferentially from peers. The authors interpreted their findings in line with the assumption that two underlying motivations may influence infants' action imitation: the motivation to acquire knowledge (preferably derived from competent adults) and the motivation to communicate and affiliate with others (social motivation in peer interactions).

Shimpi, Akhtar, and Moore (2013) have systematically examined the influence of partner characteristics in the context of observational learning from third-party interactions. In a series of experiments, the authors presented 18-month-old toddlers with different scenarios showing an adult experimenter demonstrating a sequence of novel arbitrary actions to a third-party learner. In one experiment, the age and familiarity of the *learner* were manipulated in that the learner was either an unfamiliar adult, an unfamiliar child, the own mother, or the own sibling. In two other experiments, the familiarity of the *demonstrator* was manipulated by including a warm-up phase or not. Toddlers' learning was furthermore compared to a situation in which they were directly addressed by a demonstrator with whom they had prior warm-up experience or not. The most striking finding of the study by Shimpi and colleagues (2013) was that toddlers were able to learn novel action sequences via observing a third-party interaction between completely unfamiliar strangers, while they imitated significantly fewer actions when a stranger had addressed them directly. Across experiments, toddlers' action imitation following third-party demonstration was indeed independent of kinship and age of the model and the short-term familiarity of the demonstrator. This contrasts with previous findings from first-party settings, and it questions the assumption that toddlers imitate actions in third-party contexts because they identify with the model (Meltzoff, 2007; Moore, 2006). Based on the results by Shimpi and colleagues, it seems rather likely that toddlers interpret third-party interactions and relations between others independently from their own relationship with the interaction partners. Moreover, and in contrast to direct social interactions, the authors speculated that the observation of third-party interactions might decrease toddlers' wariness toward strangers, making prior affiliative interactions with a demonstrator unnecessary for third-party learning.

Various questions remain open and need to be studied in future studies. First of all, it remains unclear how infants in the *first* year of life attend to and learn from third-part interactions between peers and familiar individuals. This could be, for example, investigated by using similar study designs as in the studies included in this thesis while varying characteristics of the interaction partners. Moreover, infants' response to peer-to-peer interactions needs to be studied further. While Shimpi and colleagues (2013) varied the age of the observed learner, the demonstrator remained an adult across conditions. Based on the previous first-party literature suggesting that infants learn different kinds of information from a direct peer model than from an adult model, it would be interesting to examine how infants learn from observed interactions with a peer demonstrator. Would they have a similar bias for imitating novel actions from an adult demonstration and familiar actions from an observed peer demonstration?

In addition, it would be interesting to investigate the contrast between observed adult-child and observed adult-adult interactions further. Shimpi and colleagues' findings suggest that, at least at 18 months of age, imitative learning does not differ between the two contexts. However, the study only focused on the characteristics of the interacting individuals. The learning activity remained consistent across all conditions, meaning that the same actions were performed in the same way. In reality, however, adult-adult interactions are systematically different from adult-child interactions. Child-directed speech, for example, differs in various properties from adult-directed speech regarding phonology, morphology, syntax, and tempo (Soderstrom, 2007). Thus, it would be possible that the observation of a more naturalistic pedagogical setting between an adult and a same-aged peer facilitates infants' learning compared to an observed adult-adult or peer-to-peer interaction—not because the observer can better identify with the learner (Shimpi et al., 2013), but rather due to simplifications that come along with characteristics of the observed pedagogical setting. Another difference is the *kind* of knowledge transmitted through an observed adult-adult and adult-child interaction. Adult-directed communication conveys much more episodic information than child-directed communication, which in turn especially conveys generalizable knowledge (Gelman, Ware, Manczak, & Graham, 2013). Following the assumption that infants are biased to acquire generic knowledge in pedagogical contexts (Csibra & Gergely, 2009), infants may profit from observing pedagogical adult-child interactions compared to observing adult-adult interactions.

5.4 Overall Conclusion

In summary, the studies presented in this thesis suggest that, throughout the first year after birth, typically developing infants develop foundational abilities and preferences, enabling them to detect and approach third-party interactions and to use these situations to learn about their environment. Moreover, the overall results reveal remarkable similarities between infants' social attention and learning in the context of self-experienced and observed social interactions. Interpreting these findings at a broader level, this dissertation contributes to the growing body of research highlighting the relevance of observational learning as a universal learning strategy in human infants. Given the vast inter-individual and cross-cultural variability in the degree to which infants encounter direct social interactions in their daily lives, the presented findings support the idea that human infants are equipped with a highly efficient toolkit of motivations and abilities affording early cultural learning in different environmental contexts.

References

Abney, D. H., Suanda, S. H., Smith, L. B., & Yu, C. (2020). What are the building blocks of parent–infant coordinated attention in free-flowing interaction?. *Infancy, 25*, 871-887.

Adamson, L. B., & Bakeman, R. (1985). Affect and Attention: Infants Observed with Mothers and Peers. *Child Development, 56*, 582-593. https://doi.org/1_0.2307/1129748

Addabbo, M., Vacaru, S. V., Meyer, M., & Hunnius, S. (2020). 'Something in the way you move': Infants are sensitive to emotions conveyed in action kinematics. *Developmental science, 23*, e12873. https://doi.org/10.1111/desc.12873

Adrien, J. L., Lenoir, P., Martineau, J., Perrot, A., Hameury, L., Larmande, C., & Sauvage, D. (1993). Blind Ratings of Early Symptoms of Autism Based upon Family Home Movies. *Journal of the American Academy of Child & Adolescent Psychiatry, 32*, 617–626. https://doi.org/10.1097/00004583-199305000-00019

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science, 8*, 199–209. https://doi.org/10.1111/j.1467-7687.2005.00406.x

Akhtar, N., & Jaswal, V. K. (2020). Stretching the Social: Broadening the Behavioral Indicators of Sociality. *Child Development Perspectives, 14*, 28–33. https://doi.org/10.1111/cdep.12351

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning Words through Overhearing. *Child Development, 72*, 416–430. https://doi.org/10.1111/1467-8624.00287

Anderson, B. A. (2016). Social reward shapes attentional biases. *Cognitive Neuroscience, 7*, 30–36. https://doi.org/10.1080/17588928.2015.1047823

Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., & Lee, K. (2013). Developmental Origins of the Other-Race Effect. *Current Directions in Psychological Science, 22*, 173–178. https://doi.org/10.1177/0963721412474459

Ariel, R., & Castel, A. D. (2014). Eyes wide open: Enhanced pupil dilation when selectively studying important information. *Experimental Brain Research, 232*, 337–344. https://doi.org/10.1007/s00221-013-3744-5

Aslin, R. N. (2007). What's in a look?. *Developmental Science, 10*, 48-53. https://doi.org/10.1111/j.1467-7687.2007.00563.

Astor, K., & Gredebäck, G. (2019). Gaze following in 4.5- and 6-month-old infants: The impact of proximity on standard gaze following performance tests. *Infancy, 24*, 79–89. https://doi.org/10.1111/infa.12261

Augusti, E.-M., Melinder, A., & Gredebäck, G. (2010). Look who's talking: Pre-verbal infants' perception of face-to-face and back-to-back social interactions. *Frontiers in Psychology, 1*, 161. https://doi.org/10.3389/fpsyg.2010.00161

Aureli, T., Presaghi, F., & Garito, M. C. (2017). Mother-infant co-regulation in dyadic and triadic contexts at 4 and 6 months of age. *Infant and Child Development, 27*, e2072. https://doi.org/10.1002/icd.2072

Bakeman, R., & Adamson, L. B. (1984). Coordinating Attention to People and Objects in Mother-Infant and Peer-Infant Interaction. *Child Development*, *55*, 1278. https://doi.org/10.2307/1129997

Bakker, M., Kochukhova, O., & von Hofsten, C. (2011). Development of social perception: A conversation study of 6-, 12- and 36-month-old children. *Infant Behavior and Development*, *34*, 363–370. https://doi.org/10.1016/j.infbeh.2011.03.001

Baldwin, D. A. (1995). Understanding the link between joint attention and language. *Joint attention: Its origins and role in development, 131,* 158.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Barton, K. (2019). *Mu-MIn: Multi-model inference.* R Package version 1.40.0. https://cran.r-project.org/web/packages/MuMIn

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2020). *Lme4: Linear mixed-effects models using "Eigen" and S4 classes.* R package version 1.1.15. https://cran.r-project.org/web/packages/lme4/lme4.pdf

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development*, *23*, 223–229. https://doi.org/10.1016/S0163-6383(01)00037-6

Becchio, C., Bertone, C., & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences*, *12*, 254–258. https://doi.org/10.1016/j.tics.2008.04.005

Begus, K., Curioni, A., Knoblich, G., & Gergely, G. (2020). Infants understand collaboration: Neural evidence for 9-month-olds' attribution of shared goals to coordinated joint actions. *Social Neuroscience*, *15*, 655–667. https://doi.org/10.1080/17470919.2020.1847730

Begus, K., Gliga, T., & Southgate, V. (2014). Infants Learn What They Want to Learn: Responding to Infant Pointing Leads to Superior Learning. *PLoS ONE*, *9*, e108817. https://doi.org/10.1371/journal.pone.0108817

Begus, K., Gliga, T., & Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proceedings of the National Academy of Sciences*, *113*, 12397–12402. https://doi.org/10.1073/pnas.1603261113

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function: Infant pointing serves an interrogative function. *Developmental Science*, *15*, 611–617. https://doi.org/10.1111/j.1467-7687.2012.01160.x

Begus, K., & Southgate, V. (2018). Curious learners: How infants' motivation to learn shapes and is shaped by infants' interactions with the social world. In Saylor, M. M. & Ganea, P. A. (Eds.), *Active Learning from Infancy to Childhood: Social Motivation, Cognition, and Linguistic Mechanisms* (pp. 13–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-77182-3

Begus, K., Southgate, V., & Gliga, T. (2015). Neural mechanisms of infant learning: Differences in frontal theta activity during object exploration modulate subsequent object recognition. *Biology Letters*, *11*, 20150041. https://doi.org/10.1098/rsbl.2015.0041

Beier, J. S., & Spelke, E. S. (2012). Infants' Developing Understanding of Social Gaze: Infants' Developing Understanding of Social Gaze. *Child Development, 83,* 486-496. https://doi.org/10.1111/j.1467-8624.2011.01702.x

Bergelson, Bergmann, Byers-Heinlein, Cristia, Cusack, & Dyck. (2017). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science, 3*, 24-52. https://doi.org/10.1177/2515245919900809

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What Do North American Babies Hear? A large-scale cross-corpus analysis. *Developmental Science, 22*, e12724. https://doi.org/10.1111/desc.12724

Berlyne, D. E. (1966). Curiosity and exploration. *Science, 153*, 25-33. Retrieved June 9, 2021, from http://www.jstor.org/stable/1719694

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences, 26*, 507–513. https://doi.org/10.1016/S0166-2236(03)00233-9

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'Liking', 'wanting', and learning. *Current Opinion in Pharmacology, 9*, 65–73. https://doi.org/10.1016/j.coph.2008.12.014

Bertenthal, B. I., & Boyer, T. W. (2015). The Development of Social Attention in Human Infants. In Puce, A. & Bertenthal, B. I. (Eds.), *The Many Faces of Social Attention* (pp. 21–65). Springer International Publishing. Retrieved June 9, 2021, from http://link.springer.com/10.1007/978-3-319-21368-2_2

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

Bigelow, A. E., MacLean, K., & Proctor, J. (2004). The role of joint attention in the development of infants' play with objects. *Developmental Science, 7*, 518–526. https://doi.org/10.1111/j.1467-7687.2004.00375.x

Biro, S., Alink, L. R. A., Huffmeijer, R., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2015). Attachment and maternal sensitivity are related to infants' monitoring of animated social interactions. *Brain and Behavior, 5*. https://doi.org/10.1002/brb3.410

Bijleveld, E., Custers, R., & Aarts, H. (2009). The unconscious eye opener: Pupil dilation reveals strategic recruitment of resources upon presentation of subliminal reward cues. *Psychological science, 20*, 1313-1315. https://doi.org/10.1111/j.1467-9280.2009.02443.x

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Observing shared attention modulates gaze following. *Cognition, 120*, 292–298. https://doi.org/10.1016/j.cognition.2011.05.002

Bråten, S. (1988). Dialogic mind: The infant and the adult in protoconversation. In Carvallo, M. E. (Eds.), *Nature, Cognition and System I* (pp. 187-205). Springer, Dordrecht. Retrieved June 9, 2021, from https://link.springer.com/chapter/10.1007/978-94-009-2991-3_9

Brooker, I., & Poulin-Dubois, D. (2013). Is a Bird an Apple? The Effect of Speaker Labeling Accuracy on Infants' Word Learning, Imitation, and Helping Behaviors. *Infancy, 18*, E46–E68. https://doi.org/10.1111/infa.12027

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews, 10*, 169–175. https://doi.org/10.1002/mrdd.20036

Burkart, J. M., Hrdy, S. B., & Van Schaik, C. P. (2009). Cooperative Breeding and Human Cognitive Evolution. *Evolutionary Anthropology: Issues, News, and Reviews, 18*, 175–186. https://doi.org/10.1002/evan.20222

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., ... & Collins, W. A. (2011). Early social cognition in three cultural contexts. *Monographs of the society for research in child development, 76*, 1–142. Retrieved June 9, 2021, from http://www.jstor.org/stable/41261530

Canfield, R. L., Smith, E. G., Brezsnyak, M. P., Snow, K. L., Aslin, R. N., Haith, M. M., ... Adler, S. A. (1997). Information Processing Through the First Year of Life: A Longitudinal Study Using the Visual Expectation Paradigm. *Monographs of the Society for Research in Child Development, 62*, i–160. https://doi.org/10.2307/1166196

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant behavior and development, 21*, 315-330. https://doi.org/10.1016/S0163-6383(98)90009-1

Carpenter, M. (2010). Social Cognition and Social Motivations in Infancy. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 106–128). Wiley-Blackwell. https://doi.org/10.1002/9781444325485.ch4

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In Seemann, A. (Ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience* (pp. 159–181). MIT Press.

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, 63*, 1–174. https://doi.org/10.2307/1166214

Carpenter, M., & Tomasello, M. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Social Development, 4*, 217–237. https://doi.org/10.1111/j.1467-9507.1995.tb00063.x

Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences, 98*, 5363–5367. https://doi.org/10.1073/pnas.081074098

Casillas, M., Brown, P., & Levinson, S. C. (2019). Early Language Experience in a Tseltal Mayan Village. *Child Development, 91*, 1819–1835. https://doi.org/10.1111/cdev.13349

Castiello, U., Becchio, C., Zoia, S., Nelini, C., Sartori, L., Blason, L., … Gallese, V. (2010). Wired to Be Social: The Ontogeny of Human Interaction. *PLoS ONE, 5*, e13199. https://doi.org/10.1371/journal.pone.0013199

Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. J*ournal of autism and developmental disorders, 39*, 1663. https://doi.org/10.1007/s10803-009-0803-7

Cheng, Y., Liu, W., Yuan, X., & Jiang, Y. (2021). The Eyes Have It: Perception of Social Interaction Unfolds Through Pupil Dilation. *Neuroscience Bulletin,* 1-4. https://doi.org/10.1007/s12264-021-00739-z

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences, 16*, 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Christodoulou, J., Leland, D. S., & Moore, D. S. (2018). Overt and covert attention in infants revealed using steady-state visually evoked potentials. *Developmental Psychology, 54*, 803–815. https://doi.org/10.1037/dev0000486

Claussen, A. H., Mundy, P. C., Mallik, S. A., & Willoughby, J. C. (2002). Joint attention and disorganized attachment status in infants at risk. *Development and Psychopathology, 14*, 279-291. https://doi.org/10.1017/S0954579402002055

Clearfield, M. W., Osborne, C. N., & Mullen, M. (2008). Learning by looking: Infants' social looking behavior across the transition from crawling to walking. *Journal of Experimental Child Psychology, 100*, 297–307. https://doi.org/10.1016/j.jecp.2008.03.005

Cleveland, A., Schug, M., & Striano, T. (2007). Joint attention and object learning in 5- and 7-month-old infants. *Infant and Child Development, 16*, 295–306. https://doi.org/10.1002/icd.508

Cleveland, A., & Striano, T. (2007). The effects of joint attention on object processing in 4- and 9-month-old infants. *Infant Behavior and Development, 30*, 499–504. https://doi.org/10.1016/j.infbeh.2006.10.009

Cohn, J. F., & Tronick, E. Z. (1987). Mother–infant face-to-face interaction: The sequence of dyadic states at 3, 6, and 9 months. *Developmental psychology, 23*, 68–77. https://doi.org/10.1037/0012-1649.23.1.68

Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review of Psychology, 52*, 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child development, 61,* 1584-1595. https://doi.org/10.1111/j.1467-8624.1990.tb02885.x

Coplan, R. J., Prakash, K., O'Neil, K., & Armer, M. (2004). Do You "Want" to Play? Distinguishing Between Conflicted Shyness and Social Disinterest in Early Childhood. *Developmental Psychology, 40*, 244–258. https://doi.org/10.1037/0012-1649.40.2.244

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*, 201–215. https://doi.org/10.1038/nrn755

Correa-Chávez, M., & Roberts, A. L. D. (2012). A cultural analysis is necessary in understanding intersubjectivity. *Culture & Psychology, 18*, 99–108. https://doi.org/10.1177/1354067X11427471

Correa-Chávez, M., & Rogoff, B. (2009). Children's attention to interactions directed to others: Guatemalan mayan and european american patterns. *Developmental Psychology, 45*, 630–641. https://doi.org/10.1037/a0014144

Crivello, C., Phillips, S., & Poulin-Dubois, D. (2018). Selective social learning in infancy: Looking for mechanisms. *Developmental Science, 21*, e12592. https://doi.org/10.1111/desc.12592

Csibra, G. (2010). Recognizing Communicative Intentions in Infancy. *Mind & Language, 25*, 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Munakata, Y. & Johnson, M. H. (Eds.). *Processes of change in brain and cognitive development. Attention and performance XXI* (pp. 249-274). Oxford University Press. Retrieved June 9, 2021, from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.585.6428&rep=rep1&type=pdf

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*, 1149–1157. https://doi.org/10.1098/rstb.2010.0319

Csibra, G., & Shamsudheen, R. (2015). Nonverbal Generics: Human Infants Interpret Objects as Symbols of Object Kinds. *Annual Review of Psychology, 66*, 689–710. https://doi.org/10.1146/annurev-psych-010814-015232

D'Entremont, B., Hains, S. M., & Muir, D. W. (1997). A demonstration of gaze following in 3-to 6-month-olds. *Infant Behavior and Development, 20*, 569-572. https://doi.org/10.1016/S0163-6383(97)90048-5

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*, 297–307. https://doi.org/10.1007/s00221-010-2479-9

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress. *Developmental Psychology, 40*, 271–283. https://doi.org/10.1037/0012-1649.40.2.271

De Barbaro, K., Johnson, C. M., & Deák, G. O. (2013). Twelve-month ''social revolution''emerges from mother-infant sensorimotor coordination: A longitudinal investigation. H*uman Development, 56*, 223-248. https://doi.org/10.1159/000351313

de Bordes, P. F., Cox, R. F. A., Hasselman, F., & Cillessen, A. H. N. (2013). Toddlers' gaze following through attention modulation: Intention is in the eye of the beholder. *Journal of Experimental Child Psychology, 116*, 443–452. https://doi.org/10.1016/j.jecp.2012.09.008

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences, 14*, 441–447. https://doi.org/10.1016/j.tics.2010.06.009

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208*, 1174-1176. https://doi.org/10.1126/science.7375928

Del Bianco, T., Falck-Ytter, T., Thorup, E., & Gredebäck, G. (2019). The Developmental Origins of Gaze-Following in Human Infants. *Infancy, 24*, 433–454. https://doi.org/10.1111/infa.12276

Deligianni, F., Senju, A., Gergely, G., & Csibra, G. (2011). Automated gaze-contingent objects elicit orientation following in 8-month-old infants. *Developmental Psychology, 47*, 1499–1503. https://doi.org/10.1037/a0025659

Devouche, E. (2004). Mother versus stranger: A triadic situation of imitation at the end of the first year of life. *Infant and Child Development, 13*, 35–48. https://doi.org/10.1002/icd.334

Di Giorgio, E., Turati, C., Altoè, G., & Simion, F. (2012). Face detection in complex visual displays: An eye-tracking study with 3- and 6-month-old infants and adults. *Journal of Experimental Child Psychology, 113*, 66–77. https://doi.org/10.1016/j.jecp.2012.04.012

Ding, X., Gao, Z., & Shen, M. (2017). Two Equals One: Two Human Actions During Social Interaction Are Grouped as One Unit in Working Memory. *Psychological Science, 28*, 1311–1320. https://doi.org/10.1177/0956797617707318

Dixon, S. D., Yogman, M., Tronick, E., Adamson, L., Als, H., & Brazelton, T. B. (1981). Early Infant Social Interaction with Parents and Strangers. *Journal of the American Academy of Child Psychiatry, 20*, 32–52. https://doi.org/10.1016/S0002-7138(09)60715-9

Elsner, Bakker, M., Rohlfing, K., & Gredebäck, G. (2014). Infants' online perception of give-and-take interactions. *Journal of Experimental Child Psychology, 126*, 280–294. https://doi.org/10.1016/j.jecp.2014.05.007

Fantz, R. L. (1964). Visual Experience in Infants: Decreased Attention to Familiar Patterns Relative to Novel Ones. *Science, 146*, 668–670. https://doi.org/10.1126/science.146.3644.668

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences, 99*, 9602–9605. https://doi.org/10.1073/pnas.152159999

Fawcett, C., & Gredebäck, G. (2013). Infants use social context to bind actions into a collaborative sequence. *Developmental Science, 16*, 841-849. https://doi.org/10.1111/desc.12074

Fawcett, C., & Gredebäck, G. (2015). Eighteen-Month-Olds, but not 14-Month-Olds, Use Social Context to Bind Action Sequences. *Infancy, 20*, 115–125. https://doi.org/10.1111/infa.12062

Feiring, C., Lewis, M., & Starr, M. D. (1984). Indirect effects and infants' reaction to strangers. *Developmental Psychology, 20*, 485–491. https://doi.org/10.1037/0012-1649.20.3.485

Fitch, A., Lieberman, A. M., Luyster, R. J., & Arunachalam, S. (2020). Toddlers' word learning through overhearing: Others' attention matters. *Journal of experimental child psychology, 193*, 104793. https://doi.org/10.1016/j.jecp.2019.104793

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature, 424*, 769-771. https://doi.org/10.1016/j.jecp.2019.104793

Floor, P., & Akhtar, N. (2006). Can 18-Month-Old Infants Learn Words by Listening In on Conversations? *Infancy, 9*, 327–339. https://doi.org/10.1207/s15327078in0903_4

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-Mounted Eye Tracking: A New Method to Describe Infant Looking: Head-Mounted Eye Tracking. *Child Development, 82*, 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project.* Cambridge, MA: MIT Press.

Fu, G., & Lee, K. (2007). Social grooming in the kindergarten: The emergence of flattery behavior. *Developmental Science, 10*, 255–265. https://doi.org/10.1111/j.1467-7687.2007.00583.x

Galazka, M. A., Roché, L., Nyström, P., & Falck-Ytter, T. (2014). Human Infants Detect Other People's Interactions Based on Complex Patterns of Kinematic Information. *PLoS ONE, 9*, e112432. https://doi.org/10.1371/journal.pone.0112432

Gampe, A., Liebal, K., & Tomasello, M. (2012). Eighteen-month-olds learn novel words through overhearing. *First Language, 32*, 385–397. https://doi.org/10.1177/0142723711433584

Gaskins, S., & Paradise, R. (2010). Learning through observation in daily life. In Lancy, D. F., Bock, J., & Gaskins, S. (Eds.), *The anthropology of learning in childhood* (p. 85–117). AltaMira Press.

Gelman, S. A., Ware, E. A., Manczak, E. M., & Graham, S. A. (2013). Children's sensitivity to the knowledge expressed in pedagogical and nonpedagogical contexts. *Developmental Psychology, 49*, 491–504. https://doi.org/10.1037/a0027901

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7*, 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Gifford-Smith, M. E., & Brownell, C. A. (2003). Childhood peer relationships: Social acceptance, friendships, and peer networks. *Journal of School Psychology, 41*, 235–284. https://doi.org/10.1016/S0022-4405(03)00048-7

Gliga, T., & Dehaene-Lambertz, G. (2005). Structural encoding of body and face in human infants and adults. *Journal of Cognitive Neuroscience, 17*, 1328–1340. https://doi.org/10.1162/0898929055002481

Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning While Babbling: Prelinguistic Object-Directed Vocalizations Indicate a Readiness to Learn. *Infancy, 15*, 362–391. https://doi.org/10.1111/j.1532-7078.2009.00020.x

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*, 544–549.

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). One-year-olds' understanding of nonverbal gestures directed to a third person. *Cognitive Development, 24*, 23–33. https://doi.org/10.1016/j.cogdev.2008.10.001

Gredebäck, G., Astor, K., & Fawcett, C. (2018). Gaze Following Is Not Dependent on Ostensive Cues: A Critical Test of Natural Pedagogy. *Child Development, 89*, 2091–2098. https://doi.org/10.1111/cdev.13026

Gredebäck, G., Gottwald, J. M., & Daum, M. M. (2021). *How our hands shape our minds: Six developmental pathways.* PsyArXiv. https://doi.org/10.31234/osf.io/378rz

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye Tracking in Infancy Research. *Developmental Neuropsychology, 35*, 1–19. https://doi.org/10.1080/87565640903325758

Gredebäck, G., & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition, 114*, 197–206. https://doi.org/10.1016/j.cognition.2009.09.004

Gredebäck, G., Theuring, C., Hauf, P., & Kenward, B. (2008). The microstructure of infants' gaze as they view adult shifts in overt attention. *Infancy, 13*, 533-543. https://doi.org/10.1080/15250000802329529

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*, 493–498. https://doi.org/10.1111/2041-210X.12504

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

Grossmann, T., & Johnson, M. H. (2007). The development of the social brain in human infancy: The social brain in infancy. *European Journal of Neuroscience*, *25*, 909–919. https://doi.org/10.1111/j.1460-9568.2007.05379.x

Grossmann, T., Johnson, M. H., Farroni, T., & Csibra, G. (2007). Social perception in the infant brain: Gamma oscillatory activity in response to eye gaze. *Social Cognitive and Affective Neuroscience*, *2*, 284–291. https://doi.org/10.1093/scan/nsm025

Gustafsson, E., Brisson, J., Mailloux, D., Mainville, M., Beaulieu, C., & Sirois, S. (2016). Do Infants Recognize Engagement in Social Interactions? The Case of Face-to-Face Conversation. *Infancy*, *21*, 685–696. https://doi.org/10.1111/infa.12135

Hafri, A., & Firestone, C. (2021). The Perception of Relations. *Trends in Cognitive Sciences*, *25*, 475–492. https://doi.org/10.1016/j.tics.2021.01.006

Hains, S. M. J., & Muir, D. W. (1996). Infant Sensitivity to Adult Eye Direction. *Child Development*, *67*, 1940–1951. https://doi.org/10.2307/1131602

Haith, M. M., Hazan, C., & Goodman, G. S. (1988). Expectation and Anticipation of Dynamic Visual Events by 3.5-Month-Old Babies. *Child Development*, *59*, 467–479. https://doi.org/10.2307/1130325

Hamlin, J. K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not Like Me = Bad: Infants Prefer Those Who Harm Dissimilar Others. *Psychological Science*, *24*, 589–594. https://doi.org/10.1177/0956797612457785

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*, 30–39. https://doi.org/10.1016/j.cogdev.2010.09.001

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557–559. https://doi.org/10.1038/nature06288

Handl, A., Mahlberg, T., Norling, S., & Gredebäck, G. (2013). Facing still faces: What visual cues affect infants' observations of others? *Infant Behavior and Development*, *36*, 583–586. https://doi.org/10.1016/j.infbeh.2013.06.001

Hay, D. F., Payne, A., & Chadwick, A. (2004). Peer relations in childhood. *Journal of Child Psychology and Psychiatry*, *45*, 84–108. https://doi.org/10.1046/j.0021-9630.2003.00308.x

Henderson, A. M. E., Wang, Y., Matz, L. E., & Woodward, A. L. (2013). Active Experience Shapes 10-Month-Old Infants' Understanding of Collaborative Goals. *Infancy*, *18*, 10–39. https://doi.org/10.1111/j.1532-7078.2012.00126.x

Henderson, A. M., & Woodward, A. L. (2011). "Let's work together": What do infants understand about collaborative goals?. *Cognition, 121*, 12-21. https://doi.org/10.1016/j.cognition.2011.05.008

Hepach, R., & Westermann, G. (2016). Pupillometry in infancy research. *Journal of Cognition and Development*, *17*, 359–377. https://doi.org/10.1080/15248372.2015.1135801

Hepper, P. G., Scott, D., & Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, *11*, 147–153. https://doi.org/10.1080/02646839308403210

Hernik, M., & Broesch, T. (2019). Infant gaze following depends on communicative signals: An eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, *22*, e12779. https://doi.org/10.1111/desc.12779

Herold, K. H., & Akhtar, N. (2008). Imitative learning from a third-party interaction: Relations with self-recognition and perspective taking. *Journal of Experimental Child Psychology*, *101*, 114–123. https://doi.org/10.1016/j.jecp.2008.05.004

Heron-Delaney, M., Wirth, S., & Pascalis, O. (2011). Infants' knowledge of their own species. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*, 1753–1763. https://doi.org/10.1098/rstb.2010.0371

Hirotani, M., Stets, M., Striano, T., & Friederici, A. D. (2009). Joint attention helps infants learn new words: Event-related potential evidence: *NeuroReport, 20*, 600–605. https://doi.org/10.1097/WNR.0b013e32832a0a7c

Hirshberg, L. M., & Svejda, M. (1990). When Infants Look to Their Parents: I. Infants' Social Referencing of Mothers Compared to Fathers. *Child Development, 61*, 1175-1186. https://doi.org/10.2307/1130885

Hoehl, S., & Bertenthal, B. I. (2021). *An Interactionist Perspective on the Development of Coordinated Social Attention.* PsyArXiv. https://doi.org/10.31234/osf.io/s5tgz

Hoehl, S., & Peykarjou, S. (2012). The early development of face processing—What makes faces special? *Neuroscience Bulletin, 28*, 765–788. https://doi.org/10.1007/s12264-012-1280-0

Hoehl, S., Reid, V. M., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science, 11*, 10–16. https://doi.org/10.1111/j.1467-7687.2007.00643.x

Hoehl, S., & Striano, T. (2015). The development and brain mechanisms of joint attention. In Calkins, S. D. (Ed.), *Handbook of infant biopsychosocial development* (p. 97–121). The Guilford Press.

Hoehl, S., Wahl, S., Michel, C., & Striano, T. (2012). Effects of eye gaze cues provided by the caregiver compared to a stranger on infants' object processing. *Developmental Cognitive Neuroscience, 2*, 81–89. https://doi.org/10.1016/j.dcn.2011.07.015

Hoehl, S., Wahl, S., & Pauen, S. (2014). Disentangling the Effects of an Adult Model's Eye Gaze and Head Orientation on Young Infants' Processing of a Previously Attended Object. *Infancy, 19*, 53–64. https://doi.org/10.1111/infa.12035

Hoehl, S., Wiese, L., & Striano, T. (2008). Young Infants' Neural Processing of Objects Is Affected by Eye Gaze Direction and Emotional Expression. *PLoS ONE, 3*, e2389. https://doi.org/10.1371/journal.pone.0002389

Hood, B. M., Willen, J. D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*, 131–134. https://doi.org/10.1111/1467-9280.00024

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development, 13*, 341–348. https://doi.org/10.1002/icd.364

Hunnius, S., & Geuze, R. H. (2004). Developmental Changes in Visual Scanning of Dynamic Faces and Abstract Stimuli in Infants: A Longitudinal Study. *Infancy, 6*, 231–255. https://doi.org/10.1207/s15327078in0602_5

Hunt, A. R., & Kingstone, A. (2003). Covert and overt voluntary attention: Linked or independent? *Cognitive Brain Research, 18*, 102–105. https://doi.org/10.1016/j.cogbrainres.2003.08.006

Ishikawa, M., Yoshimura, M., Sato, H., & Itakura, S. (2019). Effects of attentional behaviours on infant visual preferences and object choice. *Cognitive Processing, 20*, 317–324. https://doi.org/10.1007/s10339-019-00918-x

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences, 114*, E9145–E9152. https://doi.org/10.1073/pnas.1714471114

Isik, L., Mynick, A., Pantazis, D., & Kanwisher, N. (2020). The speed of human social interaction perception. *NeuroImage, 215*, 116844. https://doi.org/10.1016/j.neuroimage.2020.116844

Itakura, S. (2001). Attention to repeated events in human infants (Homo sapiens): Effects of joint visual attention versus stimulus change. *Animal Cognition*, *4*, 281–284. https://doi.org/10.1007/s100710100087

Jirout, J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, *32*, 125–160. https://doi.org/10.1016/j.dr.2012.04.002

Johnson, M. H. (1994). Visual attention and the control of eye movements in early infancy. *Attention and performance XV: Conscious and nonconscious information processing, 15*, 291-310.

Keller, H. (2007). *Cultures of infancy*. Lawrence Erlbaum Associates.

Keller, H. (2018). Parenting and socioemotional development in infancy and early childhood. *Developmental Review*, *50*, 31–41. https://doi.org/10.1016/j.dr.2018.03.001

Keller, H., Hentschel, E., Yovsi, R. D., Lamm, B., Abels, M., & Haas, V. (2004). The Psycho-linguistic Embodiment of Parental Ethnotheories: A New Avenue to Understanding Cultural Processes in Parental Reasoning. *Culture & Psychology*, *10*, 293–330. https://doi.org/10.1177/1354067X04042890

Keller, H., Scholmerich, A., & Eibl-Eibesfeldt, I. (1988). Communication patterns in adult-infant interactions in Western and non-Western cultures. J*ournal of Cross-Cultural Psychology, 19*, 427-445. https://doi.org/10.1177/0022022188194003

Keller, H., Yovsi, R., Borke, J., Kartner, J., Jensen, H., & Papaligoura, Z. (2004). Developmental Consequences of Early Parenting Experiences: Self-Recognition and Self-Regulation in Three Cultural Communities. *Child Development*, *75*, 1745–1760. https://doi.org/10.1111/j.1467-8624.2004.00814.x

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The Other-Race Effect Develops During Infancy: Evidence of Perceptual Narrowing. *Psychological Science*, *18*, 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., … Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*, F31–F36. https://doi.org/10.1111/j.1467-7687.2005.0434a.x

Killen, M., & Uzgiris, I. C. (1981). Imitation of actions with objects: The role of social meaning. The *Journal of Genetic Psychology, 138*, 219-229. https://doi.org/10.1080/00221325.1981.10534136

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2012). 'Native' Objects and Collaborators: Infants' Object Choices and Acts of Giving Reflect Favor for Native Over Foreign Speakers. *Journal of Cognition and Development*, *13*, 67–81. https://doi.org/10.1080/15248372.2011.567200

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as Epistemic Request: 12-month-olds Point to Receive New Information. *Infancy*, *19*, 543–557. https://doi.org/10.1111/infa.12060

Krogh-Jespersen, S., Henderson, A. M., & Woodward, A. L. (2020). Let's get it together: Infants generate visual predictions based on collaborative goals. *Infant Behavior and Development, 59*, 101446. https://doi.org/10.1016/j.infbeh.2020.101446

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, *10*, 110–120. https://doi.org/10.1111/j.1467-7687.2007.00572.x

Lancy, D. F. (2010). Learning 'From Nobody': The Limited Role of Teaching in Folk Models of Children's Development. *Childhood in the Past*, *3*, 79–106. https://doi.org/10.1179/cip.2010.3.1.79

Lee, K., Quinn, P. C., & Pascalis, O. (2017). Face Race Processing and Racial Bias in Early Development: A Perceptual-Social Linkage. *Current Directions in Psychological Science*, *26*, 256–262. https://doi.org/10.1177/0963721417690276

Legare, C. H., & Nielsen, M. (2020). Ritual explained: Interdisciplinary answers to Tinbergen's four questions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*, 20190419. https://doi.org/10.1098/rstb.2019.0419

Legerstee, M., Barna, J., & DiAdamo, C. (2000). Precursors to the development of intention at 6 months: Understanding people and their actions. *Developmental Psychology*, *36*, 627–634. https://doi.org/10.1037/0012-1649.36.5.627

Legerstee, M., & Varghese, J. (2001). The Role of Maternal Affect Mirroring on Social Expectancies in Three-Month-Old Infants. *Child Development*, *72*, 1301–1313. https://doi.org/10.1111/1467-8624.00349

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *emmeans: estimated marginal means.* R package version 1.4. 4. https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

Leo, I., & Simion, F. (2009). Face processing at birth: A Thatcher illusion study. *Developmental Science*, *12*, 492–498. https://doi.org/10.1111/j.1467-7687.2008.00791.x

Leong, V., & Schilbach, L. (2019). The promise of two-person neuroscience for developmental psychiatry: Using interaction-based sociometrics to identify disorders of social interaction. *British Journal of Psychiatry, 215*, 636-638. doi:10.1192/bjp.2019.73

Lewis, S. N., West, A. F., Stein, A., Malmberg, L.-E., Bethell, K., Barnes, J., … The Families, Children and Child Care (FCCC) project team. (2009). A comparison of father-infant interaction between primary and non-primary care giving fathers. *Child: Care, Health and Development*, *35*, 199–207. https://doi.org/10.1111/j.1365-2214.2008.00913.x

Liberman, Z., Kinzler, K. D., & Woodward, A. L. (2021). Origins of homophily: Infants expect people with shared preferences to affiliate. *Cognition*, *212*, 104695. https://doi.org/10.1016/j.cognition.2021.104695

Litman, J. (2005). Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition & Emotion*, *19*, 793–814. https://doi.org/10.1080/02699930541000101

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin, 116*, 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Lutchmaya, S., & Baron-Cohen, S. (2002). Human sex differences in social and non-social looking preferences, at 12 months of age. *Infant Behavior and Development*, *25*, 319–325. https://doi.org/10.1016/S0163-6383(02)00095-4

Lutchmaya, S., Baron-Cohen, S., & Raggatt, P. (2002). Foetal testosterone and eye contact in 12-month-old human infants. *Infant Behavior and Development*, *25*, 327–335. https://doi.org/10.1016/S0163-6383(02)00094-2

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*, 264–279. https://doi.org/10.1016/j.jecp.2014.05.004

Manohar, S. G., & Husain, M. (2013). Attention as foraging for information and value. *Frontiers in Human Neuroscience*, *7,* 711. https://doi.org/10.3389/fnhum.2013.00711

Matheson, H., Moore, C., & Akhtar, N. (2013). The development of social learning in interactive and observational contexts. *Journal of Experimental Child Psychology*, *114*, 161–172. https://doi.org/10.1016/j.jecp.2012.09.003

Matsuda, Y.-T., Okanoya, K., & Myowa-Yamakoshi, M. (2013). Shyness in Early Infancy: Approach-Avoidance Conflicts in Temperament and Hypersensitivity to Eyes during Initial Gazes to Faces. *PLoS ONE*, *8*, e65476. https://doi.org/10.1371/journal.pone.0065476

McDermott, K. B., & Roediger, H. L. (2018). Memory (encoding, storage, retrieval). In Butler, A. *General Psychology FA2018* (pp. 117-153). DEF Publishers. Retrieved June 9, 2021, from https://core.ac.uk/download/pdf/303864230.pdf#page=120

Meltzoff, A. N. (2007). "Like me": A foundation for social cognition. *Developmental Science*, *10*, 126–134. https://doi.org/10.1111/j.1467-7687.2007.00574.x

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. N. (2010). "Social" robots are psychological agents for infants: A test of gaze following. *Neural Networks*, *23*, 966–972. https://doi.org/10.1016/j.neunet.2010.09.005

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a New Science of Learning. *Science*, *325*, 284–288. https://doi.org/10.1126/science.1175626

Meng, X., Uto, Y., & Hashiya, K. (2017). Observing Third-Party Attentional Relationships Affects Infants' Gaze Following: An Eye-Tracking Study. *Frontiers in Psychology*, *7*, 2065. https://doi.org/10.3389/fpsyg.2016.02065

Mesman, J., Minter, T., Angnged, A., Cissé, I. A. H., Salali, G. D., & Migliano, A. B. (2018). Universality Without Uniformity: A Culturally Inclusive Approach to Sensitive Responsiveness in Infant Caregiving. *Child Development*, *89*, 837–850. https://doi.org/10.1111/cdev.12795

Michel, C., Kayhan, E., Pauen, S., & Hoehl, S. (2021). Effects of Reinforcement Learning on Gaze Following of Gaze and Head Direction in Early Infancy: An Interactive Eye-Tracking Study. *Child Development*, 13497. https://doi.org/10.1111/cdev.13497

Michel, C., Pauen, S., & Hoehl, S. (2017). Schematic eye-gaze cues influence infants' object encoding dependent on their contrast polarity. *Scientific Reports*, *7*, 7247. https://doi.org/10.1038/s41598-017-07445-9

Michel, C., Wronski, C., Pauen, S., Daum, M. M., & Hoehl, S. (2019). Infants' object processing is guided specifically by social cues. *Neuropsychologia*, *126*, 54–61. https://doi.org/10.1016/j.neuropsychologia.2017.05.022

Misch, A., Over, H., & Carpenter, M. (2016). I won't tell: Young children show loyalty to their group by keeping group secrets. *Journal of Experimental Child Psychology*, *142*, 96–106. https://doi.org/10.1016/j.jecp.2015.09.016

Molina, M., Van de Walle, G. A., Condry, K., & Spelke, E. S. (2004). The animate-inanimate distinction in infancy: Developing sensitivity to constraints on human actions. *Journal of Cognition and Development*, *5*, 399–426. https://doi.org/10.1207/s15327647jcd0504_1

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

Moore, C. (Ed.). (2006). *The development of commonsense psychology.* Lawrence Erlbaum Associates.

Moore, C. (2007). Understanding self and others in the second year. In Brownell & Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 43-65). Guilford Press.

Moore, C., & Corkum, V. (1998). Infant gaze following based on eye direction. *British Journal of Developmental Psychology*, *16*, 495–503. https://doi.org/10.1111/j.2044-835X.1998.tb00767.x

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Neal, R., & Schwartz, H. K. (2000). Gaze following, temperament, and language development in 6-month-olds: A replication and extension. *Infant Behavior and Development*, *23*, 231–236. https://doi.org/10.1016/S0163-6383(01)00038-8

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*, 938–954. https://doi.org/10.1111/j.1467-8624.2007.01042.x

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*, 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*, 115–128. https://doi.org/10.1007/BF02206861

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: The contribution of non-verbal communication measures. *Journal of child psychology and psychiatry, 27*, 657-669. https://doi.org/10.1111/j.1469-7610.1986.tb00190.x

Nadel, J., Carchon, I., Kervella, C., Marcelli, D., & Reserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental Science*, *2*, 164–173. https://doi.org/10.1111/1467-7687.00065

Nguyen, T., Banki, A., Markova, G., & Hoehl, S. (2020). Studying parent-child interaction with hyperscanning. *Progress in brain research, 254,* 1-24. https://doi.org/10.1016/bs.pbr.2020.05.003

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, *42*, 555–565. https://doi.org/10.1037/0012-1649.42.3.555

Nielsen, M., & Haun, D. (2016). Why developmental psychology is incomplete without comparative and cross-cultural perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*, 20150071. https://doi.org/10.1098/rstb.2015.0071

Nyström, P., Falck-Ytter, T., & Gredebäck, G. (2016). The TimeStudio Project: An open source scientific workflow system for the behavioral and brain sciences. *Behavior Research Methods*, *48*, 542–552. https://doi.org/10.3758/s13428-015-0616-x

Ockleford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of neonates to parents' and others' voices. *Early Human Development*, *18*, 27–36. https://doi.org/10.1016/0378-3782(88)90040-0

Odden, H., & Rochat, P. (2004). Observational learning and enculturation. *Educational and Child Psychology, 21*, 39–50.

O'Doherty, K., Troseth, G. L., Shimpi, P. M., Goldenberg, E., Akhtar, N., & Saylor, M. M. (2011). Third-Party Social Interaction and Word Learning From Video: Social Interaction and Video. *Child Development*, *82*, 902–915. https://doi.org/10.1111/j.1467-8624.2011.01579.x

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013). The power of human gaze on infant learning. *Cognition*, *128*, 127–133. https://doi.org/10.1016/j.cognition.2013.03.011

Okumura, Y., Kanakogi, Y., Kobayashi, T., & Itakura, S. (2017). Individual differences in object-processing explain the relationship between early gaze-following and later language development. *Cognition*, *166*, 418–424. https://doi.org/10.1016/j.cognition.2017.06.005

Okumura, Y., Kanakogi, Y., Kobayashi, T., & Itakura, S. (2020). Ostension affects infant learning more than attention. *Cognition*, *195*, 104082. https://doi.org/10.1016/j.cognition.2019.104082

Okumura, Y., Kobayashi, T., & Itakura, S. (2016). Eye Contact Affects Object Representation in 9-Month-Old Infants. *PLOS ONE, 11*, e0165145. https://doi.org/10.1371/journal.pone.0165145

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders, 24*, 247–257. https://doi.org/10.1007/BF02172225

Oudeyer, P.-Y., & Smith, L. B. (2016). How Evolution May Work Through Curiosity-Driven Developmental Process. *Topics in Cognitive Science*, *8*, 492–502. https://doi.org/10.1111/tops.12196

Over, H. (2016). The origins of belonging: Social motivation in infants and young children. *Phil. Trans. R. Soc. B*, *371*, 20150072. https://doi.org/10.1098/rstb.2015.0072

Over, H., & Carpenter, M. (2009). Eighteen-month-old infants show increased helping following priming with affiliation. *Psychological Science*, *20*, 1189–1193. https://doi.org/10.1111/j.1467-9280.2009.02419.x

Papeo, L. (2020). Twos in human visual perception. *Cortex*, *132*, 473–478. https://doi.org/10.1016/j.cortex.2020.06.005

Papeo, L., Goupil, & Hochmann. (2020). *Visual perception grounding of social cognition in preverbal infants.* The Open Science Framework. https://osf.io/preprints/v7y9x/

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological science, 28*, 369-379. https://doi.org/10.1177/0956797616685769

Paradise, R., & Rogoff, B. (2009). Side by Side: Learning by Observing and Pitching In. *Ethos, 37*, 102–138. https://doi.org/10.1111/j.1548-1352.2009.01033.x

Parise, E., Reid, V. M., Stets, M., & Striano, T. (2008). Direct eye contact influences the neural processing of objects in 5-month-old infants. *Social Neuroscience*, *3*, 141–150. https://doi.org/10.1080/17470910701865458

Pascalis, O., De Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life?. *Science, 296*, 1321-1323. https://doi.org/10.1126/science.1070223

Pascalis, O., & de Haan, M. (2003). Recognition memory and novelty preference: What model. In Hayne, H. & Fagen, J. (Eds.), *Progress in infancy research* (pp. 95–119). Psychology Press.

Pascalis, O., de Martin de Viviés, X., Anzures, G., Quinn, P. C., Slater, A. M., Tanaka, J. W., & Lee, K. (2011). Development of face processing: Development of face processing. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 666–675. https://doi.org/10.1002/wcs.146

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology, 32*, 3-25. https://doi.org/10.1080/00335558008248231

Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, *110*, E3965–E3972. https://doi.org/10.1073/pnas.1304326110

Puce, A., & Bertenthal, B. I. (Eds.). (2015). *The many faces of social attention: Behavioral and neural measures.* Springer. Retrieved June 9, 2021, from https://link.springer.com/chapter/10.1007/978-3-319-21368-2_1

Puce, A., Latinus, M., Rossi, A., Parada, F., Love, S., Ashourvan, A., & Jayaraman, S. (2015). Neural bases for social attention in healthy humans. In *The Many Faces of Social Attention* (pp. 93–127). Springer. Retrieved June 9, 2021, from https://link.springer.com/chapter/10.1007/978-3-319-21368-2_1

Putnam, S. P., & Stifter, C. A. (2002). Development of approach and inhibition in the first year: Parallel findings from motor behavior, temperament ratings and directional cardiac response. *Developmental Science*, *5*, 441–451. https://doi.org/10.1111/1467-7687.00239

Quadrelli, E., Monacò, S., Turati, C., & Bulf, H. (2020). Social context influences infants' ability to extract statistical information from a sequence of gestures. *Infant Behavior and Development*, *61*, 101506. https://doi.org/10.1016/j.infbeh.2020.101506

Raz, G., & Saxe, R. (2020). Learning in Infancy Is Active, Endogenously Motivated, and Depends on the Prefrontal Cortices. *Annual Review of Developmental Psychology*, *2*, 247–268. https://doi.org/10.1146/annurev-devpsych-121318-084841

Reddy, V., 2003. On being the object of attention: implications for self–other consciousness. *Trends Cognitive Science. 7*, 397–402. https://doi.org/10.1016/S1364-6613(03)00191-8

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science*, *14*, 302–308. https://doi.org/10.1111/1467-9280.14431

Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., & Reissland, N. (2017). The Human Fetus Preferentially Engages with Face-like Visual Stimuli. *Current Biology*, *27*, 1825-1828.e3. https://doi.org/10.1016/j.cub.2017.05.044

Reid, V. M., & Striano, T. (2005). Adult gaze influences infant attention and object processing: Implications for cognitive neuroscience. *European Journal of Neuroscience*, *21*, 1763–1766. https://doi.org/10/bp4dcf

Reid, V. M., & Striano, T. (2007). The directed attention model of infant social cognition. *European Journal of Developmental Psychology*, *4*, 100–110. https://doi.org/10.1080/17405620601005648

Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *Neuroreport*, *15*, 2553–2555.

Reuter, T., Emberson, L., Romberg, A., & Lew-Williams, C. (2018). Individual differences in nonverbal prediction and vocabulary size in infancy. *Cognition*, *176*, 215–219. https://doi.org/10.1016/j.cognition.2018.03.006

Reynolds, G. D. (2015). Infant visual attention and object recognition. *Behavioural Brain Research*, *285*, 34–43. https://doi.org/10.1016/j.bbr.2015.01.015

Reznick, J. S., Chawarska, K., & Betts, S. (2000). The Development of Visual Expectations in the First Year. *Child Development*, *71*, 1191–1204. https://doi.org/10.1111/1467-8624.00223

Rochat, P., Querido, J. G., & Striano, T. (1999). *Emerging Sensitivity to the Timing and Structure of Protoconversation in Early Infancy, 35,* 950.

Rochat, P., & Striano, T. (1999). Social-Cognitive Development in the First Year. In Rochat, P. (Ed.), *Early social cognition: Understanding of others in the first months of life* (pp. 3–34). New York, London: Psychology Press.

Rochat, P., Striano, T., & Blatt, L. (2002). Differential effects of happy, neutral, and sad still-faces on 2-, 4- and 6-month-old infants. *Infant and Child Development*, *11*, 289–303. https://doi.org/10.1002/icd.259

Rogoff, B., Dahl, A., & Callanan, M. (2018). The importance of understanding children's lived experience. *Developmental Review*, *50*, 5–15. https://doi.org/10.1016/j.dr.2018.05.006

Rogoff, B., Moore, L., Najafi, B., Dexter, A., Correa-Chávez, M., & Solís, J. (2007). Children's Development of Cultural Repertoires through Participation in Everyday Routines and Practices. In Goslin, D. A. (Ed.), *Handbook of socialization: Theory and research* (pp. 490–515). New York, NY, US: The Guilford Press.

Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand Learning Through Intent Participation. *Annual Review of Psychology*, *54*, 175–203. https://doi.org/10.1146/annurev.psych.54.101601.145118

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*, 704–713. https://doi.org/10.1037/0012-1649.18.5.704

Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. *Developmental Psychology*, *39*, 877–890. https://doi.org/10.1037/0012-1649.39.5.877

Sage, K. D., & Baldwin, D. (2010). Social gating and pedagogy: Mechanisms for learning and implications for robotics. *Neural Networks*, *23*, 1091–1098. https://doi.org/10.1016/j.neunet.2010.09.004

Sage, K. D., & Baldwin, D. (2011). Disentangling the Social and the Pedagogical in Infants' Learning about Tool-Use: Effects of Social Cues on Infant Tool-use. *Social Development*, *20*, 825–844. https://doi.org/10.1111/j.1467-9507.2011.00624.x

Salley, B., & Colombo, J. (2016). Conceptualizing Social Attention in Developmental Research: Social Attention in Developmental Research. *Social Development, 25*, 687–703. https://doi.org/10.1111/sode.12174

Schaffer, H. R. (1966). The onset of fear of strangers and the incongruity hypothesis. *Journal of Child Psychology and Psychiatry, 7*, 95–106. https://doi.org/10.1111/j.1469-7610.1966.tb02167.x

Scheel, A. M., Ritchie, S. J., Brown, N. J. L., & Jacques, S. L. (2018). Methodological problems in a study of fetal visual perception. *Current Biology, 28*, R594–R596. https://doi.org/10.1016/j.cub.2018.03.047

Schmitow, C., Kochukhova, O., & Nyström, P. (2016). Social perception: How do 6-month-old infants look at pointing gestures? *Infant Behavior and Development, 42*, 152–156. https://doi.org/10.1016/j.infbeh.2015.12.001

Scofield, J., & Behrend, D. A. (2011). Clarifying the role of joint attention in early word learning. *First Language, 3*, 326–341. https://doi.org/10.1177/0142723710395423

Seehagen, S., & Herbert, J. S. (2011). Infant Imitation From Televised Peer and Adult Models. *Infancy, 16*, 113–136. https://doi.org/10.1111/j.1532-7078.2010.00045.x

Seehagen, S., & Herbert, J. S. (2012). Selective imitation in 6-month-olds: The role of the social and physical context. *Infant Behavior and Development, 35*, 509-512. https://doi.org/10.1016/j.infbeh.2012.05.002

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology, 18*, 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Shimpi, P. M., Akhtar, N., & Moore, C. (2013). Toddlers' imitative learning in interactive and observational contexts: The role of age and familiarity of the model. *Journal of Experimental Child Psychology, 116*, 309–323. https://doi.org/10.1016/j.jecp.2013.06.008

Shneidman, L., Buresh, J. S., Shimpi, P. M., Knight-Schwarz, J., & Woodward, A. L. (2009). Social Experience, Social Attention and Word Learning in an Overhearing Paradigm. *Language Learning and Development, 5*, 266–281. https://doi.org/10.1080/15475440903001115

Shneidman, L., Gaskins, S., & Woodward, A. (2016). Child-directed teaching and social learning at 18 months of age: Evidence from Yucatec Mayan and US infants. *Developmental Science, 19*, 372–381. https://doi.org/10.1111/desc.12318

Shneidman, L., Todd, R., & Woodward, A. (2014). Why Do Child-Directed Interactions Support Imitative Learning in Young Children? *PLoS ONE, 9*, e110891. https://doi.org/10.1371/journal.pone.0110891

Shneidman, L., & Woodward, A. L. (2016). Are child-directed interactions the cradle of social learning? *Psychological Bulletin, 142*, 1–17. https://doi.org/10.1037/bul0000023

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 and Development, 55*, 77–87. https://doi.org/10.1016/j.infbeh.2019.03.005

Simion, F., Di Giorgio, E., Leo, I., & Bardi, L. (2011). The processing of social stimuli in early infancy. *Progress in Brain Research, 189*, 173-193. https://doi.org/10.1016/B978-0-444-53884-0.00024-5

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences, 105*, 809–813. https://doi.org/10.1073/pnas.0707021105

Simpson, E. A., Varga, K., Frick, J. E., & Fragaszy, D. (2011). Infants Experience Perceptual Narrowing for Nonprimate Faces: Perceptual narrowing for nonprimate faces. *Infancy, 16*, 318–328. https://doi.org/10.1111/j.1532-7078.2010.00052.x

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

Skinner, F. B. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century-Crofts.

Slaughter, V., & McConnell, D. (2003). Emergence of Joint Attention: Relationships Between Gaze Following, Social Referencing, Imitation, and Naming in Infancy. *The Journal of Genetic Psychology, 164*, 54–71. https://doi.org/10.1080/00221320309597503

Slaughter, V., Nielsen, M., & Enchelmaier, P. (2008). Interacting socially with human hands at 24 months of age. *Infancy, 13*, 185-195. https://doi.org/10.1080/15250000701795721

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*, 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition, 96*, B1-B11. https://doi.org/10.1016/j.cognition.2004.07.004

Spelke, E. S. (1985). Preferential-looking methods as tools for the study of cognition in infancy. In Gottlieb, G. & Krasnegor, N. A. (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (p. 323–363). Ablex Publishing.

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology, 109*, 132–140. https://doi.org/10.1016/j.jecp.2010.10.013

Steckler, C. M., Liberman, Z., Van de Vondervoort, J. W., Slevinsky, J., Le, D. T., & Hamlin, J. K. (2018). Feeling out a link between feeling and infant sociomoral evaluation. *British Journal of Developmental Psychology, 36*, 482–500. https://doi.org/10.1111/bjdp.12232

Stenberg, G., & Hagekull, B. (2007). Infant Looking Behavior in Ambiguous Situations: Social Referencing or Attachment Behavior? *Infancy, 11*, 111–129. https://doi.org/10.1111/j.1532-7078.2007.tb00218.x

Striano, T. (2001). From social expectations to social cognition in early infancy. *Bulletin of the Menninger Clinic, 65*, 361–370. https://doi.org/10.1521/bumc.65.3.361.19854

Striano, T., & Bertin, E. (2005). Coordinated affect with mothers and strangers: A longitudinal analysis of joint engagement between 5 and 9 months of age. *Cognition & Emotion, 19*, 781–790. https://doi.org/10.1080/02699930541000002

Striano, T., Chen, X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology, 3*, 289–299. https://doi.org/10.1080/17405620600879779

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in Cognitive Sciences, 10*, 471–476. https://doi.org/10.1016/j.tics.2006.08.006

Striano, T., & Rochat, P. (1999). Developmental link between dyadic and triadic social competence in infancy. *British Journal of Developmental Psychology, 17*, 551–562.

Su, P. L., Rogers, S. J., Estes, A., & Yoder, P. (2021). The role of early social motivation in explaining variability in functional language in toddlers with autism spectrum disorder. *Autism, 25*, 244–257. https://doi.org/10.1177/1362361320953260

Swann, W. B., Gómez, Á., Seyle, D. C., Morales, J. F., & Huici, C. (2009). Identity fusion: The interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology, 96*, 995–1011. https://doi.org/10.1037/a0013668

Swettenham, J., Baron-Cohen, S., Charman, T., Cox, A., Baird, G., Drew, A., ... Wheelwright, S. (1998). The Frequency and Distribution of Spontaneous Attention Shifts between Social and Nonsocial Stimuli in Autistic, Typically Developing, and Nonautistic Developmentally

Delayed Infants. *The Journal of Child Psychology and Psychiatry, 39,* 747–1753. https://doi.org/10.1111/1469-7610.00373

Symons, L. A., Hains, S. M., & Muir, D. W. (1998). Look at me: Five-month-old infants' sensitivity to very small deviations in eye-gaze during social interactions. I*nfant Behavior and Development, 21,* 531-536. https://doi.org/10.1016/S0163-6383(98)90026-1

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2015). Is ostension any more than attention? *Scientific Reports, 4,* 5304. https://doi.org/10.1038/srep05304

Tamir, D. I., & Hughes, B. L. (2018). Social Rewards: From Basic Social Building Blocks to Complex Social Behavior. *Perspectives on Psychological Science, 13,* 700–717. https://doi.org/10.1177/1745691618776263

Tamis-LeMonda, C. S., & Song, L. (2012). Parent–infant communicative interactions in cultural context. *Handbook of Psychology, 6,* 143-170. https://doi.org/10.1002/9781118133880.hop206006

Teller, D. Y. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development, 2,* 135–153. https://doi.org/10.1016/S0163-6383(79)80016-8

Teti, D. M., Bond, L. A., & Gibbs, E. D. (1988). Mothers, fathers, and siblings: A comparison of play styles and their influence upon infant cognitive level. International Journal of Behavioral *Development, 11,* 415-432. https://doi.org/10.1177/016502548801100402

Theuring, C., Gredebäck, G., & Hauf, P. (2007). Object processing during a joint gaze following task. *European Journal of Developmental Psychology, 4,* 65–79. https://doi.org/10.1080/17405620601051246

Thiele, M., Hepach, R., Michel, C., Gredebäck, G., & Haun, D. (2019). *Third-party social interaction targets enhance infants' visual learning.* The Open Science Framework. https://osf.io/4a9b6/

Thiele, M., Hepach, R., Michel, C., Gredebäck, G., & Haun, D. (2021). Social interaction targets enhance 13-month-old infants' associative learning. *Infancy, 26,* 409-441. https://doi.org/10.1111/infa.12393

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2017a). *Young Infants' Attentional Preference for Social Interactions.* The Open Science Framework. https://osf.io/42nyv/

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2017b). *Follow-up study: Young Infants' Attentional Preference for Social Interactions.* The Open Science Framework. https://osf.io/s4uy7/

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2021). Infants' preference for social interactions increases from 7 to 13 months of age. *Child Development.* https://doi.org/10.1111/cdev.13636

Thiele, M., Hepach, R., Michel, C., and Haun, D. (2021). Observing others' joint attention increases 9-month-old infants' object encoding. *Developmental Psychology, 57,* 837-850. https://doi.org/10.1037/dev0001189

Thiele, M., Michel, C., Hepach, R., & Haun, D. (2020a). *Infants' object processing in the context of third-party social interactions.* The Open Science Framework. https://osf.io/yfegm/

Thiele, M., Michel, C., Hepach, R., & Haun, D. (2020b). *The role of ostensive context and gaze cues on infants' object processing.* The Open Science Framework. https://osf.io/dp5cg/

Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M., & Carey, S. (2011). Big and Mighty: Preverbal Infants Mentally Represent Social Dominance. *Science, 331,* 477–480. https://doi.org/10.1126/science.1199198

Thorgrimsson, G. B., Fawcett, C., & Liszkowski, U. (2014). Infants' expectations about gestures and actions in third-party interactions. *Frontiers in Psychology, 5,* 1061-1084. https://doi.org/10.3389/fpsyg.2014.00321

Thorgrimsson, G. B., Fawcett, C., & Liszkowski, U. (2015). 1- and 2-year-olds' expectations about third-party communicative actions. *Infant Behavior and Development*, *39*, 53–66. https://doi.org/10.1016/j.infbeh.2015.02.002

Todd, J. T., & Dixon, W. E. (2010). Temperament moderates responsiveness to joint attention in 11-month-old infants. *Infant Behavior and Development*, *33*, 297–308. https://doi.org/10.1016/j.infbeh.2010.03.007

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

Tomasello, M. (2001). Cultural transmission: A view from chimpanzees and human infants. *Journal of cross-cultural psychology, 32*, 135-146. https://doi.org/10.1177/0022022101032002002

Tomasello, M. (2009). *The cultural origins of human cognition.* Harvard university press.

Tomasello, M. (2016). Cultural Learning Redux. *Child Development*, *87*, 643–653. https://doi.org/10.1111/cdev.12499

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*, 121–125. https://doi.org/10.1111/j.1467-7687.2007.00573.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–691.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child development, 57*, 1454-1463. https://doi.org/10.2307/1130423

Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of Child Language*, *23*, 157–176. https://doi.org/10.1017/S0305000900010138

Tomasello, M., & Vaish, A. (2013). Origins of Human Cooperation and Morality. *Annual Review of Psychology*, *64*, 231–255. https://doi.org/10.1146/annurev-psych-113011-143812

Trevarthen, C. (1993). Predispositions to cultural learning in young infants. *Behavioral and Brain Sciences*, *16*, 534–535. https://doi.org/10.1017/S0140525X00031502

Trevarthen, C. (2011). What is it like to be a person who knows nothing? Defining the active intersubjective mind of a newborn human being. *Infant and Child Development*, *20*, 119–135. https://doi.org/10.1002/icd.689

Trevarthen, C., & Aitken, K. J. (2001). Infant Intersubjectivity: Research, Theory, and Clinical Applications. *Journal of Child Psychology and Psychiatry*, *42*, 3–48. https://doi.org/10.1017/S0021963001006552

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The Infant's Response to Entrapment between Contradictory Messages in Face-to-Face Interaction. *Journal of the American Academy of Child Psychiatry*, *17*, 1–13. https://doi.org/10.1016/S0002-7138(09)62273-1

Tummeltshammer, K., Feldman, E. C. H., & Amso, D. (2019). Using pupil dilation, eye-blink rate, and the value of mother to investigate reward learning mechanisms in infancy. *Developmental Cognitive Neuroscience*, *36*, 100608. https://doi.org/10.1016/j.dcn.2018.12.006

Tummeltshammer, K., Mareschal, D., & Kirkham, N. Z. (2014). Infants' Selective Attention to Reliable Visual Cues in the Presence of Salient Distractors. *Child Development*, 1981-1994. https://doi.org/10.1111/cdev.12239

Tummeltshammer, K. S., Wu, R., & Kirkham, N. Z. (2013). 8-month-olds Know Which Face is Reliable. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 35,* 1420-1425. Retrieved June 9, 2021, from https://escholarship.org/uc/item/4pc7x7p2

Valdesolo, P., & DeSteno, D. (2011). Synchrony and the social tuning of compassion. *Emotion*, *11*, 262–266. https://doi.org/10.1037/a0021302

Venezia, M., Messinger, D. S., Thorp, D., & Mundy, P. (2004). The Development of Anticipatory Smiling. *Infancy*, *6*, 397–406. https://doi.org/10.1207/s15327078in0603_5

Venezia Parlade, M. V., Messinger, D. S., Delgado, C. E. F., Kaiser, M. Y., Van Hecke, A. V., & Mundy, P. C. (2009). Anticipatory smiling: Linking early affective communication and social outcome. *Infant Behavior and Development*, *32*, 33–43. https://doi.org/10.1016/j.infbeh.2008.09.007

Vernetti, Angelina, Senju, A., Charman, T., Johnson, M. H., & Gliga, T. (2018). Simulating interaction: Using gaze-contingent eye-tracking to measure the reward value of social signals in toddlers with and without autism. *Developmental Cognitive Neuroscience*, *29*, 21–29. https://doi.org/10.1016/j.dcn.2017.08.004

Vernetti, Angélina, Smith, T. J., & Senju, A. (2017). Gaze-contingent reinforcement learning reveals incentive value of social signals in young children and adults. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20162747. https://doi.org/10.1098/rspb.2016.2747

Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition*, *200*, 104270. https://doi.org/10.1016/j.cognition.2020.104270

Vestner, T., Gray, K. L. H., & Cook, R. (2021). Visual search for facing and non-facing people: The effect of actor inversion. *Cognition*, *208*, 104550. https://doi.org/10.1016/j.cognition.2020.104550

Vivanti, G., Fanning, P. A., Hocking, D. R., Sievers, S., & Dissanayake, C. (2017). Social attention, joint attention and sustained attention in autism Spectrum disorder and Williams syndrome: convergences and divergences. *Journal of autism and developmental disorders*, *47*, 1866-1877. https://doi.org/10.1007/s10803-017-3106-4

von Hofsten, C., Uhlig, H., Adell, M., & Kochukhova, O. (2009). How children with autism look at events. *Research in Autism Spectrum Disorders*, *3*, 556–569. https://doi.org/10.1016/j.rasd.2008.12.003

Wagner, Kelley, W. M., Haxby, J. V., & Heatherton, T. F. (2016). The Dorsal Medial Prefrontal Cortex Responds Preferentially to Social Interactions during Natural Viewing. *Journal of Neuroscience*, *36*, 6917–6925. https://doi.org/10.1523/JNEUROSCI.4220-15.2016

Wahl, S., Marinović, V., & Träuble, B. (2019). Gaze cues of isolated eyes facilitate the encoding and further processing of objects in 4-month-old infants. *Developmental Cognitive Neuroscience*, *36*, 100621. https://doi.org/10.1016/j.dcn.2019.100621

Wahl, S., Michel, C., Pauen, S., & Hoehl, S. (2013). Head and eye movements affect object processing in 4-month-old infants more than an artificial orientation cue. *British Journal of Developmental Psychology*, *31*, 212–230. https://doi.org/10.1111/bjdp.12001

Wang, Q., Bolhuis, J., Rothkopf, C. A., Kolling, T., Knopf, M., & Triesch, J. (2012). Infants in Control: Rapid Anticipation of Action Outcomes in a Gaze-Contingent Paradigm. *PLoS ONE*, *7*, e30884. https://doi.org/10.1371/journal.pone.0030884

Warneken, F., & Tomasello, M. (2007). Helping and Cooperation at 14 Months of Age. *Infancy*, *11*, 271–294. https://doi.org/10.1111/j.1532-7078.2007.tb00227.x

Wass, S. V., Clackson, K., & Leong, V. (2018). Increases in Arousal are More Long-Lasting than Decreases in Arousal: On Homeostatic Failures During Emotion Regulation in Infancy. *Infancy*, *23*, 628–649. https://doi.org/10.1111/infa.12243

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental psychology*, *34*, 1289. https://doi.org/10.1037/0012-1649.34.6.1289

Wimpory, D. C., Hobson, R. P., Williams, J. M. G., & Nash, S. (2000). Are infants with autism socially engaged? A study of recent retrospective parental reports. *Journal of autism and developmental disorders*, *30*, 525-536. https://doi.org/10.1023/A:1005683209438

Woodward, A., & Hoyne, K. (1999). Infants' Learning about Words and Sounds in Relation to Objects. *Child Development, 70*, 65–77. https://doi.org/10.1111/1467-8624.00006

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology, 107*, 118–136. https://doi.org/10.1016/j.jecp.2010.04.014

Wu, R., Kirkham, N. Z., Swan, K. A., & Gliga, T. (2011). Social signals scaffold learning from novel cues during infancy. In Kokinov, B., Karmiloff-Smith, A., Nersessian, N. J. (Eds.), *European Perspectives on Cognitive Science.* New Bulgarian University Press.

Wu, R., Tummeltshammer, K., Gliga, T., & Kirkham, N. Z. (2014). Ostensive signals support learning from novel attention cues during infancy. *Frontiers in Psychology, 5.* https://doi.org/10.3389/fpsyg.2014.00251

Yamashiro, A., & Vouloumanos, A. (2018). How do infants and adults process communicative events in real time? *Journal of Experimental Child Psychology, 173*, 268–283. https://doi.org/10.1016/j.jecp.2018.04.011

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*, 13690–13695. https://doi.org/10.1073/pnas.0804388105

Yu, C., & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLoS ONE, 8*, e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2017). Hand-Eye Coordination Predicts Joint Attention. *Child Development, 88*, 2060–2078. https://doi.org/10.1111/cdev.12730

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology, 106*, 208–220. https://doi.org/10.1016/j.jecp.2010.03.002

Zmyj, N., Daum, M. M., Prinz, W., Nielsen, M., & Aschersleben, G. (2012). Fourteen-month-olds' imitation of differently aged models. *Infant and Child Development, 21,* 250-266. https://doi.org/10.1002/icd.750

Appendix A – Supplementary Materials Study I

A1 - General Supplementary Information

Additional Information Eye Tracking Task

Content of the video stimuli. We manipulated the relation between the two adults as interacting or non-interacting by using the relative positioning of their bodies (facing versus standing back-to-back), gaze direction (eye contact versus looking away), the execution of an action (co-regulated versus individually). These criteria were a result of the following considerations in the planning phase:

Even though prior research has shown that (even static) body orientation and gaze direction is sufficient to guide 9-month-old infants to facing dyads (Handl et al., 2013), we decided to include the execution of an action, as we aimed to increase the co-regulated turn-taking aspect between the two actors to highlight their interactive relationship. In similar previous studies this has been done, for example, by showing infants a turn-taking conversation between two people (Augusti et al., 2010; Beier & Spelke, 2012). Since in our study, in contrast to these studies, we presented social interaction and control video simultaneously, we needed to create stimuli without auditory stimulation.

In addition to co-regulated activity between the two actors, we further decided to only include (a) neutral interactions (i.e., without pro- nor anti-social meaning), (b) interactions that could be cut in the center to horizontally mirror the videos of the separate actors in the control condition (which would not have been possible if the actors would pass an object, e.g., a ball), (c) interactions that were not physically impossible in the mirrored back-to-back videos (to avoid saliency due to impossible actions), and (d) interactions that were "too interactive" even in isolation to remove the interactive aspect in the control condition as much as possible (e.g., waving hands while standing back-to-back could have been interpreted as interacting with another individual outside the visible region of the video). For the same reason, the agents in the control videos were positioned in such a way that they never "touched" the border of the video. Criterion (b) was particularly important as we presented the stimuli in a forced-choice preferential looking procedure. Differences in motion or synchronicity might have had an effect on infants' preferential orienting (Valdesolo & DeSteno, 2011).
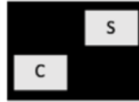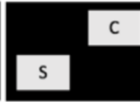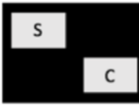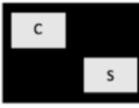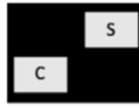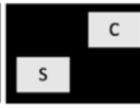
Based on the previously described criteria, we created stimuli showing the three interactions clapping game, leaning heads and touching hands. The actors first looked forward for two seconds before turning toward or away from one another (as in Augusti et al., 2010). All three interactions contained mutual touch between the actors, which was not planned initially, but consistent with stimuli used by Galazka and collegues (2014).

Creating the video stimuli. We filmed the actors individually to ensure flexible and accurate positioning of the dyad partners, consistent timing of actions between actors and trials, and identical levels of motion between social and control stimuli within trials. The cont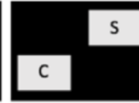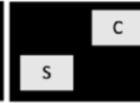rol stimuli were created by horizontally mirroring the actions of the individual actors. We filmed the individuals in front of a green screen. This way, we could control for color and luminance differences between and within videos. We used Adobe Premiere Pro for cutting and editing the videos. Adobe Premiere's Ultra Key tool was used to isolate the actors from the background and replace it with an even colored, grey background layer which was identical over all videos.

Actors. Four female actors acted in the stimuli. They all wore white t-shirts, had their hair tied back, and did not wear any glasses or jewelry. We combined the four actors in four dyadic arrangements. This means that every actor was seen together with two different actors, one time on the right side and one time on the left side. Each of the four dyads of actors recurred in all three interactions but never occurred in the same position twice. Each of the three interactions (and corresponding control video) were shown in four possible diagonal arrangements on the screen (see Figure A1).

Figure A1

Positioning of the videos on screen for the twelve experimental trials. In the left column, control and social stimulus are exemplarily illustrated for each of the three interactions with one of the four dyads

Notes. S = social interaction, C = non-interactive control.

Video duration. Each trial lasted 12 seconds. Our decision to use a video duration of 12 seconds was based on our observation during piloting that this time was long enough to give the

youngest participants (7-month-olds) enough time to look at both videos, while being short enough to keep the oldest infants' attention throughout 12 trials.

Additional Information Free Play

Free play coding categories. Our pre-registered coding scheme included the three infant behaviors "general look at parent", "eye contact", and "joint attention look". We decided to focus on these three behaviors to be able to examine changes and variability in social engagement between 7 and 11 months of age. For this purpose, we created a coding scheme that was sensitive to different levels of social interactions, including both "earlier" developing interaction patterns (i.e., face-to-face interactions), as well as "richer" interaction behaviors presumably shown with increasing age (i.e., joint attention). We added the fourth category "look at parent's face" to our coding scheme after seeing parts of the video recordings for the first time and prior to coding. The reason for adding this category was to get a more precise picture of infants' social engagement, as it would reflect infants' motivation to engage in eye contact without necessarily requiring the parent to look back to them. Accordingly, looks at the parent's face did not reflect higher social engagement skills as compared to eye contact. Figure A3 provides an illustration of the total occurrences of the four infant behaviors.

Free play coding procedure. The coding of the free play sessions was conducted in Microsoft Excel and proceeded as follows: The coder watched each video recording (5 min) in sixty 5-second intervals in reduced speed (0.35x). For each interval, the coder decided if the infant showed one of the four relevant behaviors. Behaviors were only coded if the face of the child was visible. If a behavior was not shown, the infant received a "0" in the respective category. If none of the behaviors were shown, the infant received a "0" in all categories. If an infant showed one of the four behaviors at least once during the 5-second interval, they received a "1" in the respective category, even if the behavior occurred more than one time during the interval. Based on the hierarchical structure of the four coding behaviors, infants automatically received a "1" in all "lower" behavior categories, when showing behaviors from category 2 ("looks at parent's face"), category 3 ("eye contact"), and category 4 ("joint attention looks", see Figure A2).

To give an example: If an infant and their parent looked at each other's eyes during one coding interval, not only category 3 ("eye contact") was scored with a "1", but also the lower categories, since eye contact necessarily involves a general interest in the parent (category 1), as well as looking at the parent's face (category 2). Note that this hierarchical coding was only used during the coding procedure. For calculating the social engagement proportion score, we used the "highest" looking behavior displayed in each interval (i.e., "Sum" row in Figure A2). If a behavior was shown longer than 5 seconds, succeeding intervals were coded with the respective behavior. Table A1 provides an overview about the detailed coding instructions for all coded behaviors.

Table A1

Detailed coding procedure for the four infant looking behaviors coded during parent-child free play

Looking behavior	Coding Instructions
1. General look at the parent	Infants receive a "1" if they look at their parent. This includes looks at objects if the parent holds it in their hand. **Coding Note.** If the mother "accidentally" crosses the infant's vision field, this episode should not be coded as general look (e.g., the infant looks at an object while the parent reaches for another object next to this object). The parent does not necessarily have to initiate the action (e.g., parent takes a rattle, then the child looks at rattle in their hand). It would also count if the child looks at a rattle before the parent touches it (if the child keeps looking at the rattle in their parent's hand).
2. Looking at the parent's face	Infants receive a "1" if they look at the face of their parent but the parent does not look back.
3. Eye contact between parent and infant	Infants receive a "1" if they looked at their parent's eyes with the parent looking back at their eyes. **Coding Note.** To distinguish "pure" eye contact from a joint attention look, follow the instructions given in the row below.
4. Joint attention looks between parent, infant and an object	Infants receive a "1" if they mutually look at the same object with their parent, before or after they have engaged in mutual eye contact. **Coding Note.** In most cases a complete joint attention period (i.e., including eye contact and look at the object) lasts longer than one 5-sec interval. Since it is difficult to determine the timepoint when a joint attention look ends and "pure" eye contact or looking at the object begins, only those intervals during which the eye contact between parent and child is established should be coded as joint attention interval. If the child alternates their gaze between object and parent, multiple intervals in a row can be coded as joint attention. To give an example: An infant looks at an object in interval 1 (no joint attention interval), then looks at the parent's eyes in interval 2 (joint attention interval), then the infant looks back at the object within the same interval, before looking back to the parent's eyes in interval 3 (joint attention interval). To differentiate between eye contact and joint attention episode, the coder should first identify the intervals in which eye contact between parent and infant takes place. Then, the coder should check the sequence immediately before the eye contact had started and after it ends. If child and parent both look at an object together in one of these periods, the coder should go back to the interval in which the eye contact is first established and code it as joint attention look. If not, the corresponding interval should be coded as eye contact.

Figure A2

Illustration of the hierarchical structure of the free play coding procedure

|	**1** **0** 00-05'	**2** **0** 06-10'	**3** **0** 11-15'	**4** **0** 16-20'	**5** **0** 21-25'	**6** **0** 26-30'	...	**60** **4** 56-60'
1) General look at parent	0	1	1	1	1	1		1
2) Look at parent's face	0	0	1	1	1	0	...	1
3) Eye contact	0	0	1	1	1	0		1
4) Joint attention look	0	0	0	1	0	0		0
Sum	0	1	3	4	3	1	...	3

Notes. Screenshot from Microsoft Excel for one simulated child. Visible is the coding for the first six 5-second intervals as well as the last interval (Total number of intervals = 60). The "Sum" row reveals the "highest" looking behavior that the child had shown in each interval. This value was included for the calculation of the social engagement score (e.g., eye contact with their parent during interval 3, 5, and 60).

Social engagement proportion score. The proportion score that we used for the main analyses of the free play data was created based on our observation during coding, that most "general looks at the parent" (category 1) were actually looks at toys that parents held in their hands. An infant who constantly looked at a toy in their parent's hand would thus frequently receive a "1" in category 1, even without being necessarily interested in their parent (but potentially rather the toy). Using total frequencies of category 1 behaviors as a measure of infants' active social engagement would thus carry the risk of overestimating their actual social interest—especially in parent-child dyads in which the parents were particularly active. To control for such potential confounding, we relativized the sum of the higher-order social behaviors (i.e., the number of intervals during which behaviors from category 2, 3, or 4 were shown) at the total amount of all coded behaviors (i.e., the number of intervals during which behaviors from categories 1, 2, 3, and 4 were shown) for each individual child. By using this proportion score instead of the raw sum of the four behaviors, we aimed to (a) isolate infants' actual social interest from their overall looking behaviors related to their parent, (b) indirectly control for individual differences in the parental activity level, and (c) improve differentiation between different levels of social engagement at the group level. To give an example: Consider a mother constantly moving a rattle within their child's field of vision. Even though this child might have low social interest, they might reveal high scores in category 1 when being interested in the rattle. However, compared to a child with higher social interest and the same amount of "general looks at their parent", this child would likely receive a lower social engagement score, as the more socially motivated child would presumably show more eye contact and joint attention looks in addition to (or in combination with) their looks at the parent-toy interaction.

To calculate the number of intervals during which the behaviors were shown, we scored the "highest" looking behavior that a child had shown in each interval (see "Sum" row in Figure A2). For example, the exemplary child in Figure A2 would have shown behavior 1 in two of the seven visible intervals, behavior 3 in three intervals, behavior 4 in one interval, whereas behavior 2 never occurred.

Figure A3

Scatterplots for the total number of occurrences of the four coded infant behaviors during free play
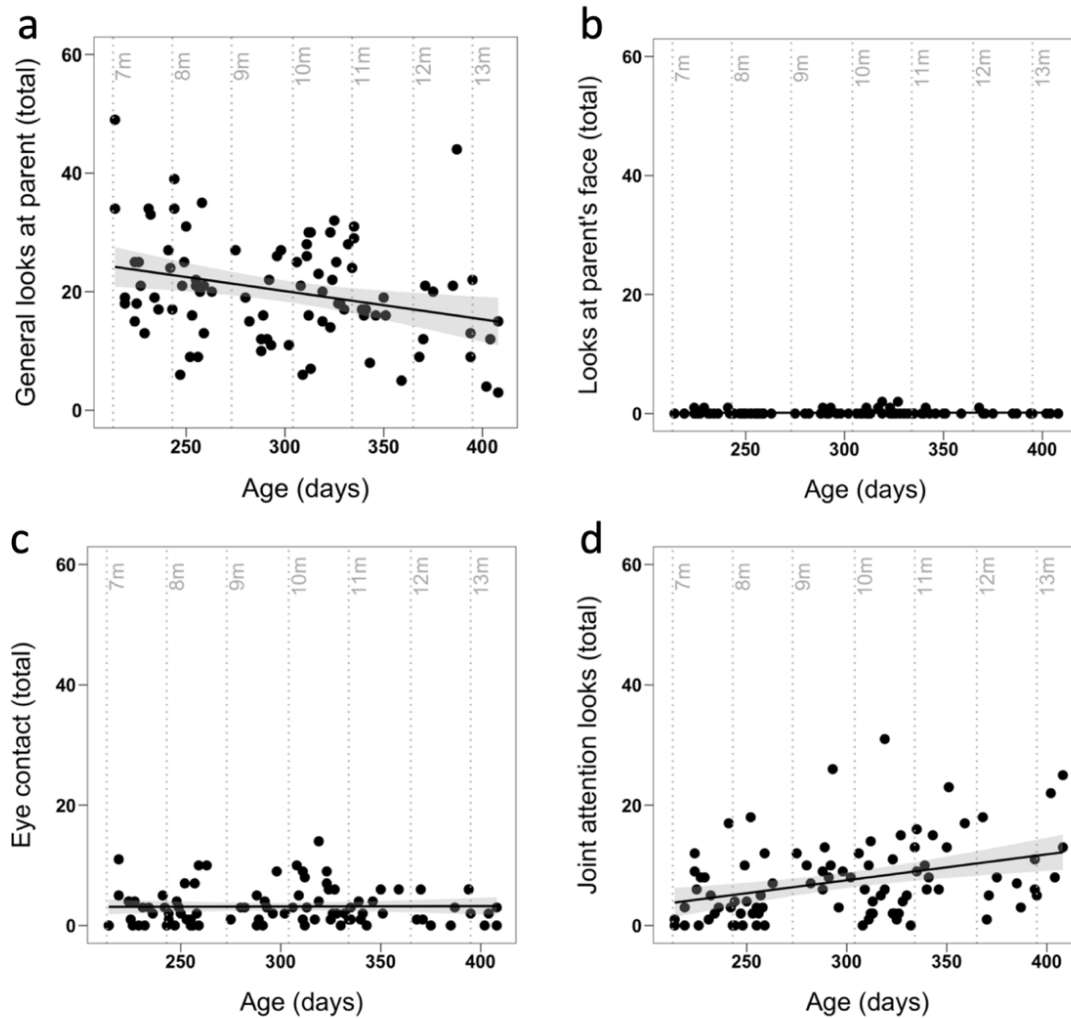
Notes. The dots represent individual data points for a merged sample including participants from Experiment 1 and 2. The vertical dashed lines indicate age in months. The linear regression lines with confidence ribbons fit to the data of the plot. Statistically significant was only (a) the decrease in general looks at the parent (p = .01), and (d) the increase in joint attention looks (p < .001).

Additional Analyses and Results across Both Experiments

Parental gender. The primary caregiver participated in the free play phase of the study (over both experiments: n = 10 fathers and n = 101 mothers). To account for the possible influence of parental gender, we repeated the overall analysis of age on infants' social engagement score after excluding all father-child interactions (Beta = 0.08 ± *SE* = .02, $t(1,99)$ = 4.14, p <.001, η^2 = .15).

Age group differences. To validate our findings from Experiment 1 regarding age group differences between 7- to 8.5-month-old infants and 9.5- to 11-month-olds, we repeated our analyses including infants from both Experiments who provided valid data in the required age ranges. For the eye tracking analysis this resulted in an extended sample of n = 30 per age group. For the free play analysis an extended sample of n = 42 in the younger age group and n = 34 in the

older age group. The results from these additional analyses were consistent with our findings in Experiment 1: Only infants in the older age group preferentially looked at the social stimuli ($M = .54$, $SD = .07$; $t(29) = 3.47$, $p = .002$, $d = 0.63$). Infants in the younger age group did not show any preference ($M = .48$, $SD = .08$; $t(29) = -1.37$, $p = .18$, $d = 0.25$). Moreover, social behavior scores were significantly higher in the older age group ($M = .36$, $SD = .19$) compared to the younger age group ($M = .25$, $SD = .22$; $F(1,74) = 5.65$, $p = .02$, $\eta^2 = .07$).

Eye tracking model. Even though we used different eye tracking models in Experiment 1 (SMI RED250mobile) and Experiment 2 (SMI RED-m), we used the same SMI software to record and export the gaze data, the same event detection filters to define gaze events, and identical R scripts for processing the data. To statistically test whether eye tracking model had an influence on the overall analysis of the merged eye tracking data, we repeated our main model for infants' proportional looking time to the social interaction stimuli, including age as continuous predictor and eye tracking model as control variable. The effect of age on infants' proportional looking time to the social interaction stimuli remained stable (Beta = $.04 \pm SE = .01$, $t(2,88) = 4.62$, $p < .001$, $\eta^2 = .20$). We did not find any effect of eye tracking model (Beta = $.004 \pm SE = .02$, $t(2,88) = 0.24$, $p = .81$, $\eta^2 = .0004$).

Effect of age on correlational analysis. In addition to the linear model reported in the main manuscript in section "Overall analysis of individual differences across both experiments", we ran a partial correlation analysis to explore the impact of age on the relation between infants' visual preference and their active social attention behavior further. When controlling for age, the partial correlation between infants' visual preference for others' interactions and their active social attention behavior was not statistically relevant ($N = 90$; $r(88) = .05$, $p = .63$, $R^2 = .0025$). Table A2 provides an overview of the correlational and post-hoc power analyses for both experiments as well as for the merged sample.

Table A2

Results from correlation analyses between infants' proportional looking time at others' social interactions (eye tracking) and their active social engagement scores (free play)

	N	r	R^2	p	Power
Experiment 1	40	.15	.02	.36	.15
Experiment 2	50	.23	.05	.11	.37
Experiment 1 & 2 (merged)	90	.24	.06	.03*	.63

Notes. N = Number of participants included in the correlation analysis.

Table A3

Results from post-hoc power analyses for the two main analyses over the merged sample (Experiment 1 and Experiment 2)

Dependent variable	N	Beta	SE	t	η^2	p	Power
Prop. looking time at others' social interactions	91	.04	.01	5.09	.23	<.001	.99
Active social engagement score	111	.09	.02	4.49	.16	<.001	.99

Notes. Results from linear models for the dependent variables, including age (in days) as continuous predictor.

A2 - Supplementary Information Experiment 1

Additional Analyses and Results Eye Tracking

We ran some additional analyses to explore the eye tracking data further. First, to allow for a more direct comparison with the main analysis of Experiment 2, we conducted a model for the mean proportional looking time to the social interaction stimuli, including age as continuous rather than categorical predictor. The mean proportional looking time increased with age (Beta = .04 ± *SE* = .01, $t(1,38)$ = 2.87, $p < .01$, η^2 = .18).

Second, we compared infants' looking preferences before and after the actors had started to turn towards or away from one another. We did not find any effects of condition during the first two seconds—neither in the older group (M = .52, SD = .11; $t(19)$ = 0.65, p = .52, d = 0.15), nor in the younger group (M = .51, SD = .09; $t(19)$ = 0.76, p = .45, d = 0.17), with no difference between age groups ($F(1,38)$ = 0.004, p = .95, η^2 = .0001). During the last ten seconds, in contrast, the mean proportional looking time at the social stimuli was significantly greater in the older as compared to the younger sample ($F(1,38)$ = 9.16, p = .004, η^2 = .19), with a looking preference for the social stimuli in only the older (M = .56, SD = .09; $t(19)$ = 3.10, p = .006, d = 0.69), not the younger sample (M = .47, SD = .10; $t(19)$ = 0.77, p = .45, d = 0.29).

Third, we explored whether the infants' looking preference for the social interaction videos varied over trials or between the kinds of interactions that we used in our stimuli. For this purpose, we conducted a general linear mixed model (GLMM, Gaussian error distribution) for the mean proportional looking time to the social stimuli, including the interaction between age group (between-subject factor) and type of interaction (within-subject factor: clapping, leaning, touching), and the interaction between age group (between-subject factor) and trial (within-subject factor: 12 trials) as fixed effects. In line with our main analysis, we additionally included gender as fixed effect in the model. As random effects, we included subject and gender as intercept, as well as the random slopes on subject for trial, type of interaction, and the interactions between age group and trial, and age group and type of interaction. Infants' looking preference did not differ between the different kinds of interactions, neither did it change over trials. We did not find

any effect of interaction type or trial, neither in interaction with age group (age × trial: $\chi^2(1) = 1.73$, $p = .19$, estimate = −0.01, $SE = 0.009$; age × type of interaction: $\chi^2 (1) = 0.63$, p = .43, estimate = −0.007, $SE = 0.009$), nor as overall main effects (trial: $\chi^2 (1) = 0.75$, $p = .39$, estimate = 0.004, $SE = 0.005$; type of interaction: $\chi^2 (1) = 0.61$, p = .43, estimate = −0.004, SE = 0.005). In line with our main analysis reported in the main manuscript, the model revealed a significant effect of age group on infants' looking preference ($\chi^2 (1) = 5.97$, $p = .01$, estimate = 0.07, $SE = 0.03$). Further in line with our main analysis, gender had no significant effect on infants' looking preference ($\chi^2(1) = 0.03$, $p = .86$, estimate = −0.005, $SE = 0.43$). We thus conclude that the effect of age on infants' looking preference was independent from the kind of interaction and did not change over time.

Table A4

Results from post-hoc power analyses for all main analyses in Experiment 1

	ANOVAs				
Dependent variable	*N*	*F*	*η²*	*p*	*Power*
Prop. looking time at others' social interactions	40	7.50	.16	.009	.69
Active social engagement score	47	5.06	.10	.03	.56
	One-sample tests against chance level				
	N	*t*	*d*	*p*	*Power*
Prop. looking time at others' social interactions *(Group 1: 7- to 8.5-month-olds)*	20	−1.56	.35	.13	.35
Prop. looking time to social interaction *(Group 2: 9.5- to 11-month-olds)*	20	2.38	.53	.03	.66

Notes. In the two analyses of variance (ANOVAs) age group was included as categorical between-subject factor.

Additional Analyses and Results Free Play

As an alternative approach to the proportion score of social engagement, we repeated our main analysis for the effect of age on social behavior, using the sum of the raw frequencies of occurrence of behaviors from categories 2 ("looks at parent's face"), category 3 ("eye contact"), and category 4 ("joint attention looks") as dependent variable (i.e., the numerator of the proportion score). Analogous to our findings based on the proportion score, this sum score was significantly higher in the older age group ($M = 11.05$, $SD = 6.78$) compared to the younger age group ($M = 6.40$, $SD = .6.28$; $F(1,45) = 5.87$, $p = .02$, $\eta^2 = .12$, see Figure A4).

We initially added the 90-seconds sequence at the beginning of the free play based on findings from prior studies revealing differences in infants' joint attention behavior due to differences in parental engagement (e.g., Bigelow, MacLean, & Proctor, 2004). By restricting the parents' active behavior at the beginning of the free play, we aimed to standardize the first 90 seconds of the free play as much as possible without entirely diminishing the naturalness of the

situation. We did not find any difference in infants' social engagement score before and after 90 seconds. Neither overall infants (before: $M = .32$, $SD = .26$; after: $M = .26$, $SD = .21$; $t(46) = 1.78$, $p = .08$), nor for the separate younger sample (before: $M = .28$, $SD = .28$; after: $M = .21$, $SD = .20$; $t(26) = 1.70$, $p = .10$) or older sample (before: $M = .38$, $SD = .23$; after: $M = .34$, $SD = .19$; $t(19) = 0.76$, $p = .45$). This suggests that the social engagement score was not significantly affected by the parental activity level.

A3 - Supplementary Information Experiment 2

Additional Analyses and Results

As described above, we repeated our main analysis for the effect of age on social behavior, using the sum score described in the section above (sum of behaviors from categories 2,3, and 4). Analogous to our findings based on the proportion score, the frequency of social behaviors increased with age both in the separate sample (Beta = $2.80 \pm SE = .95$, $t(1,62) = 2.95$, $p < .001$, $\eta^2 = .12$) and in the merged sample (Beta = $3.03 \pm SE = .67$, $t(1,109) = 4.51$, $p < .001$, $\eta^2 = .16$, see Figure A4). The social behavior sum score did not correlate with infants' proportional looking time at social interactions in the merged sample ($N = 90$; $r(88) = .13$, $p = .21$).

Since the analyses of the separate infant behaviors revealed a significant age effect for the occurrence of joint attention looks during free play, we ran exploratory analyses for the correlation between infants' joint attention looks and their orienting to third-party interactions. The proportional looking time at social interactions did not correlate with infants' joint attention score, neither in separate analyses of Experiment 2 ($N = 50$; $r(48) = .12$, $p = .41$), nor in the merged sample ($N = 90$; $r(88) = .15$, $p = .15$).

Table A5

Results from post-hoc power analyses for all main analyses over a merged sample in Experiment 2

Dependent variable	N	Beta	SE	t	η^2	p	Power
Prop. looking time at others' social interactions	51	.04	.01	3.73	.22	<.001	.91
Active social engagement score	64	.09	.03	3.35	.15	<.001	.86

Notes. Results from linear models for the dependent variables, including age (in days) as continuous predictor.

Figure A4

Scatterplot with individual data points illustrating the effect of age on infants' social engagement sum score (p < .001)
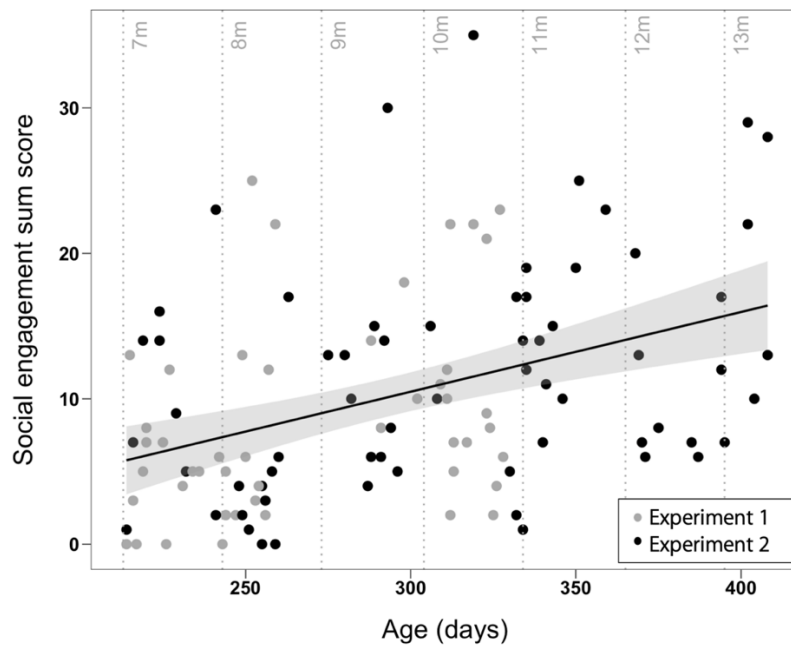
Notes. The social engagement sum score represents the sum of the frequencies of behaviors from category 2 ("looks at parent's face"), category 3 ("eye contact"), and category 4 ("joint attention looks"). The vertically dashed lines indicate age in months. The linear regression lines with confidence ribbons fit to the overall data of the plots.

Appendix B – Supplementary Materials Study II

B1 - Supplementary Information Visual Learning Task

Timing of the Task

The timing of the visual learning task was based on prior studies and our observations during piloting. The final timing is illustrated in Figure 1 in the main document and explained in detail in Table B1.

Table B1

Detailed information regarding the timing of the visual learning task

Phase	Duration	Explanation
Gap	200 ms	We added this gap between trials to clearly separate the trials from one another. The delay created the impression that the next cue "popped up" on the screen after the target video from the previous trial had stopped playing.
Cue	Gaze dependent (max 4000 ms) + 300 ms delay	The cue was gaze dependent to support children's learning by (a) making sure that children had seen the cue and (b) adjusting the task to individual differences as infants could control the trial-to-trial speed with their gaze.
		After the child had fixated the cue for 150 ms (Deligianni, Senju, Gergely, & Csibra, 2011), it remained on screen for another 300 ms before it disappeared. This delay was to ensure that children had processed the cue shape before it disappeared. Initially, we piloted a delay of 1500 ms (Reuter et al., 2018), but reduced it to 300 ms as children started to explore the screen during the longer delay. We aimed to (a) clearly point out that cues were associated with following targets and (b) support the sensation of eliciting the target by looking at the cue.
Gap	600	The delay between cue and target was due to give infants' the opportunity to perform predictive gaze shifts to the cued target region (timing as in Wang et al., 2012).
Target	4000	The duration of the target video was based on our aim to keep it as short as possible to keep the "flow" of the task while keeping it long enough that it contained the relevant phases: (a) facing forward, (b) turning towards/away from one another, (c) establishing the interaction/non-interactive action.

Additional Analysis and Results

Saccadic latency and looking time. Following visual inspection of the data, we ran further exploratory analyses in addition to the analyses described in the main document. First, we observed that the latencies in the social interaction condition seemed to rapidly decline during the first half of trials (trials 1-6), before inclining during the second half again (trials 7-12). In order to explore possible habituation effects to the target videos as one possible explanation of this pattern,

we conducted the same GLMMs for looking time at the target videos as we ran for our main analysis of saccadic latency. Looking times were assessed by calculating the total duration of fixations within the social interaction and control AOI for each trial, including fixation data from target video onset until video offset. We found a continuous decrease in looking time at the target videos, indicating a general decrease in sustained attention throughout the experiment (main effect of trial, $\chi^2(1)$ = 11.12, p < .001, estimate = −187.09, SE = 52.06). Following the assumption that this general decrease in attention had caused the continuous increase in saccadic latencies throughout the second half of the experiment, we ran an exploratory GLMM for saccadic latency over the first six trials only. The results of this analysis are reported in the main manuscript.

First look. In complementing the first look analysis in the main document, Table B2 shows the total number of first looks at the two target AOIs following social interaction and control cue as well as the proportional number of first looks used for the first look analysis.

Table B2

Total number of first looks in the visual learning task

	Social Interaction AOI	Control AOI	Prop. Number of Looks to Target AOI *(SD)*
Social Interaction cue	261	103	*.72 (.32)*
Control cue	101	266	*.72 (.31)*

Notes. Total numbers over all participants and trials and proportional number of first looks to the correct target region (i.e., number of first looks to the cued target AOI divided by the total number of first looks in both AOIs). Total number of trials over all conditions and participants = 768.

Stimuli

Although seemingly acting in dyads the actors were filmed individually. This ensured an accurate positioning of the dyad partners, consistent timing of actions between actors and trials, and identical levels of motion between social and control stimuli within trials. The control stimuli were created by horizontally mirroring the actions of the individual actors. To clarify that the agents in the control videos were not interacting with another individual outside the visible region of the video, the actors were positioned in such a way that they never "touched" the border of the video. All actors were filmed in front of a green screen to control for color and luminance differences between and within videos. Adobe Premiere Pro was used for cutting and editing the videos. Adobe Premiere's Ultra Key tool was used to isolate the actors from the background and replace it with an even colored, grey background layer which was identical over all videos. All social interactions contained mutual touch between the actors. The actors were visible from the waist up. They were all female, wore white t-shirts, had their hair tied back, and did not wear any glasses or jewelry.

B2 - Supplementary Information Preferential-Looking Task

Additional Analyses and Results

As described in the pre-registration, we explored sub-group differences between enhanced and less-enhanced learners further by using a more sophisticated group assignment procedure (see also Mani & Huettig, 2014). For this purpose, we divided the sample based on a median split of a beta-coefficient difference score, calculated for each individual by subtracting the beta-coefficient of their learning function during social interaction trials from the beta-coefficient of their learning function during control trials. This procedure was more sophisticated than using mean latencies, since it focused on latency changes over time. Children with enhanced performance in the learning task did not look significantly longer at the social interaction shape ($M = .48$; $SD = .13$) compared to less enhanced learners ($M = .43$; SD = .18, $t(30) = -1.38$; $p = .18$, $d = -.49$).

B3 - Supplementary Information Manual Forced-Choice Task

Procedure

The procedure was adapted from previous studies (e.g., Hamlin & Wynn, 2011). The experimenter, sitting opposite of the child, held the choice board upside down with the stimuli facing her body. At the beginning of the task, the experimenter said "Hi" while looking at the child. As soon as the child looked at the experimenter, she flipped around the board out of the child's reach, saying "Look!". Once the child had looked at both shapes, the experimenter pushed the board forward and asked "Which one do you want?". We ended the task if the child had made a choice or if the child did not make any choice for two minutes.

B4 - Additional Gaze Following Task (Exploratory)

In addition to our main research question, we included a gaze-following task to explore the possible relation between children's performance in the visual learning task and their gaze following abilities. Infants' seeking of learnable content might not only be reflected in their tendency to learn associations between arbitrary shapes and situations with observational learning opportunities, but also relate their tendency to use others' gaze as social cue guiding them to relevant information. Since we added the gaze-following task as an exploratory measure, it was conducted during an additional eye tracking phase (5 min) at the very end of each testing session. The same eye tracking hardware was used as for the main tasks of the study. We run the task by using Tobii Studio (version 3.4.8.1348).

Stimuli and Design

Each child was presented with six videos during which an actress shifted her gaze to one of two target objects. During half of the trials the actress looked to an object being located to her right side, while she looked to an object located to her left side during the other half of the trials. The order of trials was randomized for each participant. As dependent variable we measured the participant's first gaze shift to either of the two objects. We used video stimuli designed for a previous study by Astor & Gredebäck (2019).

Data Analysis and Coding

We defined three areas of interest: one elliptical AOI for the head and two rectangular areas, one for each target. A gaze following difference score was calculated by subtracting the number of incongruent trials in the gaze following task (child first looked at the not-attended object after actor's head turn) from the number of congruent trials (child first looked at the attended object after the actor's head turn). Trials were only counted as valid when children had looked at the head of the actor before looking at one of the two target objects. On average, each participant contributed 4.8 valid gaze-following trials to the analysis (SD = .71, range = 3-6 trials). To check whether the participants followed gaze at all, we tested the gaze following score against chance level by running a one sample test against zero. To investigate possible relations between visual learning and gaze following abilities, we correlated the latency difference score from the visual learning task with a gaze following difference score by using Pearson's r correlation. We used the MATLAB based open source software TimeStudio (Nyström, Falck-Ytter, & Gredebäck, 2016) for defining AOIs and pre-processing the data (TobiiStudio version 3.19; MATLAB version R2018b).

Results

Children followed gaze in the additional gaze-following task (M = 3; SD = 1.66; $t(29)$ = 9.89, p = .00, d = 1.8). Gaze following abilities were not correlated with visual learning abilities, N = 30, $r(28)$ = .20, p = .30.

　　　Complementary analyses. In addition to our pre-registered plan, we compared the proportional number of congruent trials between enhanced and less enhanced learners in the visual learning task (see main document section 'Data Analyses and Coding' for the procedure of sub-sample creation). The proportional number of trials in which children followed gaze was higher for the sub-sample of enhanced learners (M = .87; SD = .13) compared to less enhanced learners (M = .74; SD = .19, $t(25)$ = 2.05; p = .05).

Appendix C – Supplementary Materials Study III

C1 - Supplementary Information Video Stimuli

Additional Information Objects

As objects, we used pictures of abstract toys from a stimulus collection initially used in a study by Wahl, Michel, Pauen, & Hoehl (2013). From the overall pool of 160 pictures, we selected 32 objects that could be fitted into a square shape. All infants saw the same pairs of objects, whereby each individual object served equally often as novel and as familiar object in all conditions. For the purpose of our study, we edited the original pictures as follows: First, we removed the background-layer and replaced it with a transparent background. Then, we fitted each object into a 260×260 pixels square shape format. The side lengths of the square shape were determined by the height of the smallest image. Objects that were not square-shaped initially, were stretched or compressed into the required format to ensure that all objects covered a similar area on the screen. As a consequence, some objects had a slightly different shape as in the study by Wahl and colleagues. We selectively adjusted the color of the objects to ensure that the luminance and saturation was similar between objects. For one object we changed the color entirely from black to blue, since this was the only black object in the selection of object, presumably making it less salient compared to the other, more colorful objects. All editing was done by using Adobe Photoshop.

Additional Information Video Content

This section provides detailed information regarding our considerations during stimulus development. We describe all deviations from the stimuli used previous studies, explain our decision to deviate from these stimuli, and elaborate on why we did not expect the deviations to have an impact on infants' object encoding.

The first deviation from previously used stimuli was that we only showed one object in the encoding phase (the familiar object). This means, that infants did not see the novel object until it appeared in the preferential-looking phase. In other screen-based studies, the novel object had been visible in the encoding phase already, but the actor did not pay attention to it (e.g., Okumura et al., 2013, 2020; Theuring et al., 2007). Since most previous studies have looked at the relation between infants' gaze following and subsequent object encoding, the presence of two objects in the encoding phase was particularly required to allow calculating a gaze-following difference score. With regard to our specific research question, our primary goal was to create an arrangement between actors and object that allowed both actors to look away from the object without looking at any other object. Moreover, we wanted to create a scenario in which both actors

played the same role in the interaction and were provided with the same amount of visual information. This excluded the possibility of providing one of the actors with two toys (one on both sides), as opposed to the other actor only having visual "access" to only one toy (as, e.g., in Meng et al., 2017). During the stimulus planning phase, we considered that the presence of only one object in the encoding phase may raise the concern that infants' novelty preference could be at ceiling when seeing the novel object in the subsequent preferential-looking phase. However, since previous object-processing studies with real interactive settings had been done with only one object in the encoding phase as well, we assumed that the paradigm should principally work with the novel object first introduced in the preferential-looking phase (Begus, Southgate, & Gliga, 2015; Cleveland et al., 2007; Cleveland & Striano, 2007; Ishikawa, Yoshimura, Sato, & Itakura, 2019).

Another difference from previous studies was the actor-object positioning during the encoding phase. Due to the presence of two (instead of one) actors, the actors in our videos were not positioned in the center, but instead within the left and right third of the screen. Correspondingly, the object was not positioned in the left or the right area of the screen, but at the bottom center instead. In contrast to previous studies, where the objects were positioned in left-right arrangement in both phases (encoding and preferential looking), the left-right positioning of the objects in the preferential-looking phase represented a new visual arrangement compared to the central placement in the encoding phase. However, since previous studies had shown that object encoding depends on object identity rather than object location (Okumura et al., 2016), we did not consider the novel arrangement in the preferential-looking phase to have an effect on infants' processing.

When planning the manipulation of third-party ostensive cues, we aimed to create a context that was (according to previous findings) rich enough to elicit object encoding, but at the same time lean enough to trace infants' object encoding back to purely *third-party* ostensive cues. A recent study by Okumura and colleagues (2020) suggests that 9-month-old infants may need infant directed speech in addition to eye contact to allow processing of an object (but see, e.g., Cleveland & Striano, 2007). Despite this finding, we did not include a corresponding initial greeting phase between the two actors, since we were concerned that the pure sound of the word "hello" could diminish the purely third-party context by giving infants the feeling of being addressed themselves, and thereby increasing their responsiveness (see also Senju & Csibra, 2008). To provide another ostensive cue in addition to third-party eye contact, we included the turning of the actors' entire body (toward or away from one another) as attention-grabbing social motion prior to third-party eye contact.

Related to the previous point, another deviation from previous videos was that the actors were shown in back view in the initial non-social sequence, rather than showing them facing forward with lowered gaze (Okumura et al., 2013, 2017, 2020; Theuring et al., 2007). We decided on this initial position as a clear demonstration of third-party context. We would argue that the

back-view in our study has an even stronger non-social meaning compared to the previously used lowered-gaze-sequence, which is why we did not consider this a relevant deviation with regard to infants' processing performance.

Even though the overall duration of eye contact was consistent with previously used videos (2 seconds), we split this sequence in two one-second phases: one face-to-face (or back-to-back) phase before the actors look toward (or away from) the object, and one corresponding phase afterwards. Other screen-based studies have only used one eye contact sequence in the beginning. We decided for this twofold eye contact sequence for two reasons: First, we aimed to generally increase the interactive dynamic between the two actors, and second, we wanted to highlight the relation between the actors in the end of the video. We did not expect the presence of two eye contact sequences to have an impact on infants' object encoding compared to other studies, because the minimum requirement for communicative cueing remained fulfilled (namely eye contact *before* the gaze shift toward the object). To ensure that infants had payed attention to this minimum requirement, we only included trials during which infants had looked at the *first* eye contact sequence and the gazing sequence.

Additional Information Video Creation

For the stimuli of both experiments, the actors were filmed individually in front of a green screen. To ensure consistent timing of actions between actors and trials, metronome clicks were played at 120 bpm while filming. If necessary, we corrected the action timing of each actor post hoc and frame-by-frame in Adobe Premiere. Filming the actors in front of a green screen ensured flexible and accurate positioning of the dyad partners. Moreover, it allowed us to control for color and luminance differences between and within videos. We used Adobe Premiere Pro for cutting and editing the videos. Adobe Premiere's Ultra Key tool was used to isolate the actors from the background and replace it with an even colored, grey background layer which was identical over all videos. We positioned the actors in such a way that their overall motions were centered around the same vertical axis. In Figure C1 (Experiment 1) and Figure C2 (Experiment 2) we illustrate the areas of interest and maximum areas that the actors' movements covered over all conditions and trials.

Figure C1

Areas of interest (AOIs) during the encoding phase of Experiment 1

Notes. All videos were presented in full-screen view (1920×1080 pixels). Green area = Object AOI (340×340 pixels), defined 1° visual angle larger than the maximum dimensions of the object. Blue areas = Face AOIs covering all possible head movements (570×430 pixels), defined 1° visual angle larger than the areas covering all possible head movements from all actors in Experiment 1 and 2.

Figure C2

Areas of interest (AOIs) during the encoding phase of Experiment 2 (a) during trials showing the actor the right side of the object and (b) during trials showing the actor on the left side
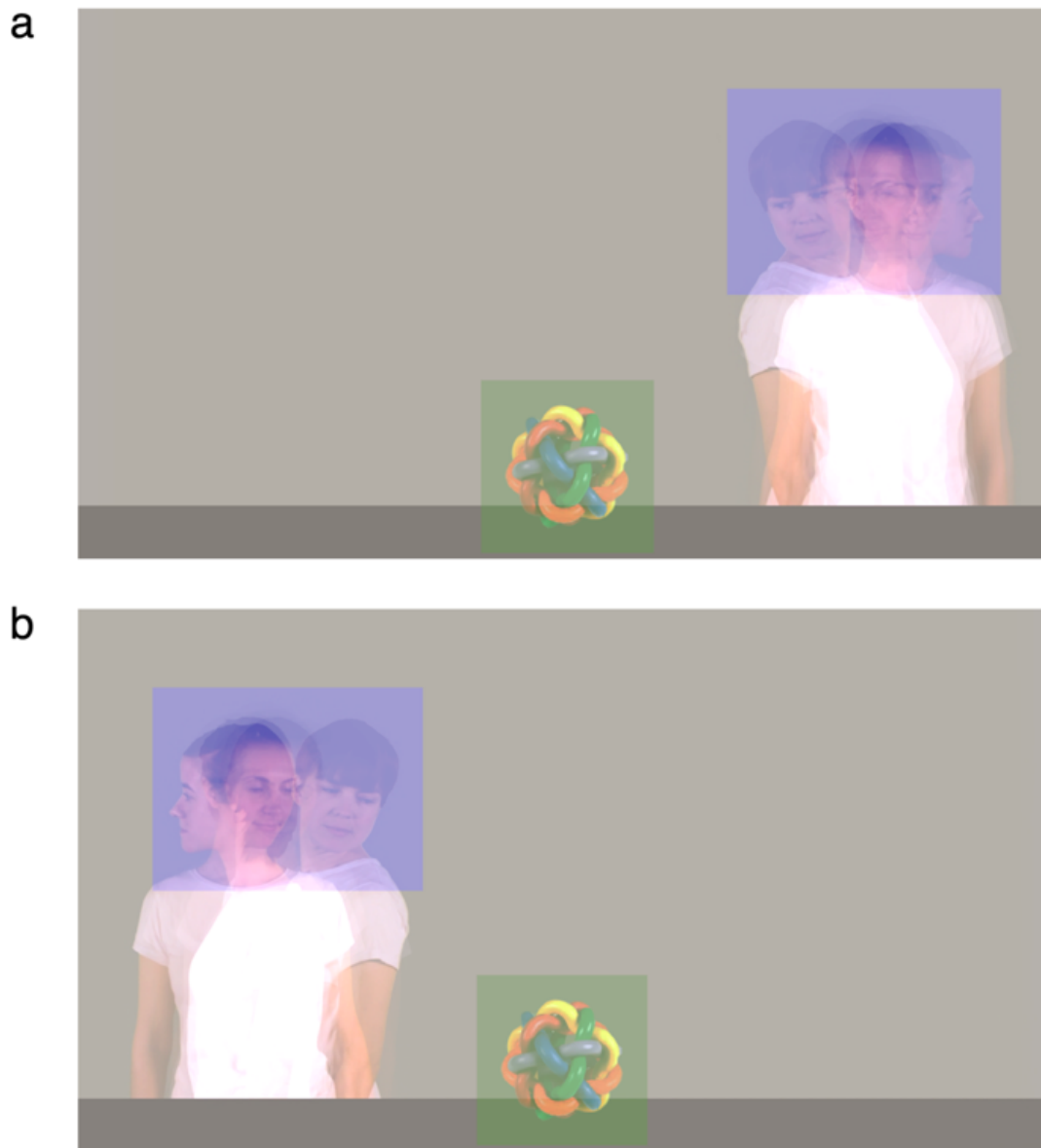
Notes. All videos were presented in full-screen view (1920×1080 pixels). Green area = Object AOI (340×340 pixels), defined 1° visual angle larger than the maximum dimensions of the object. Blue area = Face AOIs covering all possible head movements (570×430 pixels), defined 1° visual angle larger than the areas covering all possible head movements from all actors in Experiment 1 and 2.
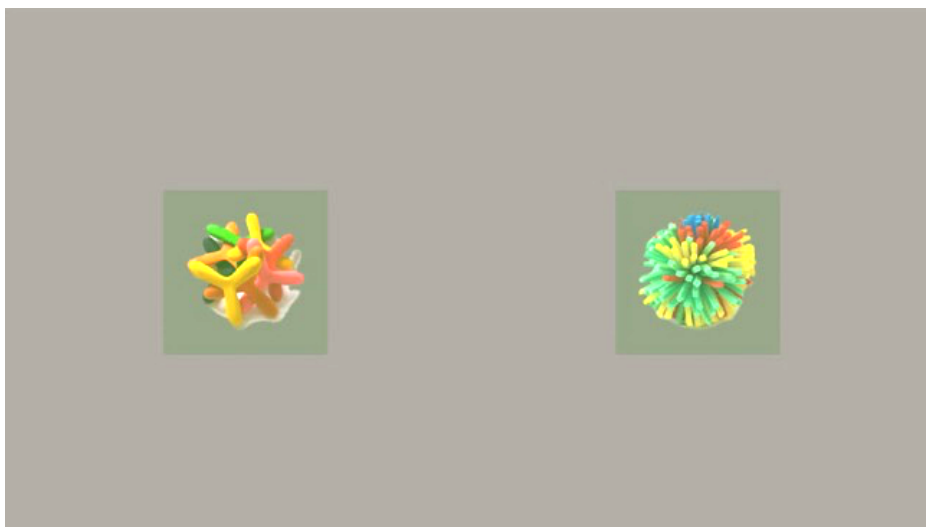
Figure C3

Areas of interest (AOIs) during the preferential-looking phase in Experiment 1 and Experiment 2

Notes. All videos were presented in full-screen view (1920×1080 pixels). Object AOIs (340×340 pixels) were defined 1° visual angle larger than the maximum dimensions of the objects.

C2 – Supplementary Analyses

We ran some analyses in addition to the analyses described in the main document to better understand the impact of infants' overt attention on their encoding performance. All following analyses served exploratory purposes and were not planned in the pre-registration. All face AOIs were defined 1° visual angle larger than the maximum areas covering all possible head movements from all actors in Experiment 1 and 2, to ensure comparability between the two experiments. The resulting AOIs covered an area of 14.4° × 10.9° (see Figure C1 and C2).

Experiment 1

Valid trial statistics. As described in more detail in the main manuscript, infants were only included in the analysis if they provided valid data in at least one trial per condition after being filtered according to our pre-registered criteria. Table C1 provides an overview of the corresponding valid trial statistics for each of the four conditions.

Table C1

Valid trial statistics for the four conditions in Experiment 1

Condition	Number of valid trials				Infant looking at object (fam.)	Infant not looking at object (fam.)
	Total	Min	Max	M *(SD)*		
Third-Party Eye Contact Looking at object	111	2	4	3.47 *(.67)*	73	38
No Third-Party Eye Contact Looking at Object	108	1	4	3.38 *(.71)*	63	45
Third-Party Eye Contact No Looking at Object	110	2	4	3.44 *(.76)*	82	28
No Third-Party Eye Contact No Looking at Object	114	2	4	3.56 *(.62)*	82	32
Total	443				300	143

Notes. Each individual infant could provide between 1 (min.) and 4 (max.) trials per condition. The maximum number of total trials over all infants was 128 per condition. The "infant looking at object (fam.)" column represents the number of valid trials during which infants had looked at the object at all over the total duration of the encoding phase (i.e., fixation duration > 0 ms within the object AOI). The "infant not looking at object (fam.)" column represents the number of valid trials during which infants had not looked at the object at all over the total duration of the encoding phase (i.e., fixation duration = 0 within the object AOI).

Overall attention during encoding. First, we investigated condition differences in infants' overall attention to the encoding videos. For this purpose, we conducted a GLMM for infants' total looking time to the screen during the encoding phase (including fixation data over the entire trial sequence). We included the same fixed and random effects in the model as in our main model for infants' novelty preference score. We found that overall attention to the stimuli did not vary statistically across condition. Neither the interaction between third-party eye contact (eye contact, no eye contact) and others' looking at the object (looking at object, not looking at object) revealed a significant effect ($\chi^2(1) = 0.06$, $p = .80$, estimate = 96.8, $SE = 399.9$), nor the main effects of the two factors (eye contact: $\chi^2(1) = 0.29$, $p = .59$, estimate = −104.1, $SE = 192.0$; looking at the object: $\chi^2(1) = 2.08$, $p = .15$, estimate = −284.5, $SE = 194.7$). However, infants' looking time to the familiarization videos decreased over trials (main effect of trial: $\chi^2(1) = 32.25$, $p < .001$, estimate = −1449.8, $SE = 194.3$). We found the same effect when repeating our model for infants' total looking times in the preferential looking phase (main effect of trial: $\chi^2(1) = 41.24$, $p < .001$, estimate = −1291.6, $SE = 140.8$), indicating a general decrease in visual attention throughout the experiment.

Looking times to the object and faces during encoding. To complement our analyses regarding the necessity of direct attention for object encoding, we assessed condition differences in infants' attention to the object in the encoding phase. For this purpose, we repeated the model

of our main analysis for infants' fixation duration within the object AOI. We included the same fixed and random effects as in our main model. We found a main effect of others' looking at the object in that infants looked longer to the object when the actors did *not* look to the object ($\chi^2(1)$ = 10.05, *p* = .002, estimate = 315.10, *SE* = 93.46). In addition, infants' attention to the object decreased over trials (main effect of trial: $\chi^2(1)$ = 4.22, *p* = .04, estimate = −215.03, *SE* = 101.31). One possible explanation for the main effect of others' looking at the object is that the faces of the actors were systematically less visible in these conditions, meaning that they carried less information and social salience. As a possible consequence, the object may have received relatively more attention compared to the two conditions in which the actor's faces were visible. To explore this possibility further we repeated our analysis over the 5-second gazing phase only (i.e., the "still" sequence during which the actors looked away from or toward the object), revealing the same main effect of others' looking at the object ($\chi^2(1)$ = 8.84, *p* = .003, estimate = 244.55, *SE* = 81.36). In addition, we compared infants' attention to the faces during the gazing phase. For this purpose, we calculated the sum of fixation durations within the two face AOIs for the corresponding sequence. We found a reversed main effect of others' looking at the object, in that infants looked longer to the faces in conditions during which the actors looked to the object ($\chi^2(1)$ = 17.68, *p* < .001, estimate = −705.52, *SE* = 149.53). The pattern was the same when including fixations over the entire duration of the encoding phase ($\chi^2(1)$ = 10.39, *p* = .001, estimate = −605.70, *SE* = 176.90). Table C3a provides a summary of the descriptive statistics for looking times during the encoding phase in all four conditions.

Taken together, our findings regarding infants' looking times suggest the following pattern: When the faces of the actors were visible, infants looked longer at the socially salient faces and shorter to the object. When the actors turned away from the object, their faces were less visible, causing longer looking times to the object and shorter looking times to the faces.

Gaze shifts to the object and between the faces during encoding. In addition to looking times, we explored infants' scanning pattern while they watched the videos in the encoding phase. First, we examined potential condition differences in the number of gaze shifts between the two faces of the actors. For this purpose, we determined the number of looks within each of the face AOIs. A "look" was defined as the interval between the first fixation on the active AOI and the end of the last fixation within the same active AOI when there were no fixations outside the AOI (Tobii Studio User Manual, Version 3.2, see also Meng et al., 2017). According to this definition one look could entail a group of multiple fixations. As a next step, we determined for each look whether the latest previous fixation (i.e., the last fixation immediately before the first fixation within the look) had hit the respective other face AOI. If this was the case, we counted this gaze event as a gaze shift from one to the other AOI. If not (i.e., if the previous fixation before a look had been somewhere else on the screen), the corresponding look was labeled as irrelevant and discarded from the analysis. To examine condition differences, we conducted a GLMM for infants' total

number of gaze shifts between the two faces, including all fixation data over the interaction phases (i.e., during the still face-to-face and back-to-back sequences). We included the same fixed and random effects as in our main model for infants' novelty preference score.

Neither the interaction between third-party eye contact and others' looking at the object revealed a significant effect ($\chi^2(1) = 0.36$, $p = .55$, estimate = 0.09, $SE = .11$), nor the main effects of the two factors (eye contact: $\chi^2(1) = 0.29$, $p = .59$, estimate = 0.05, $SE = 0.09$; looking at the object: $\chi^2(1) = 1.67$, $p = .20$, estimate = 0.10, $SE = 0.08$). The pattern remained the same when including fixation data over the entire duration of the video. This is in contrast to previous studies showing an increased number of gaze shifts between facing dyads as compared to people standing back-to-back (Augusti et al., 2010; Meng et al., 2017). One possible explanation for the equal number of gaze shifts between the faces in the back-to-back conditions of our study is that infants were seeking information that could explain why the actors had turned away from one another in the first place.

In addition to gaze shifts between the two actors, we explored infants' object looks further. Specifically, we aimed to examine whether the origin of a specific look had an impact on infants' encoding performance, such that a socially caused referential look might be more relevant and therefore increase infants' processing compared to a look without any social origin. To test this assumption, we first calculated all looks within the object AOI, proceeding as described for the face-to-face gaze shift analysis above. Then, we decided for each look whether the latest previous fixation had hit one of the two face AOIs. If this was the case, this look was labeled as socially-caused referential look. If the previous latest fixation did not hit any of the two face AOIs, the corresponding look was labeled as not socially-caused and discarded from the analysis. A trial was discarded from the analysis if no gaze shift had been performed toward the object at all. To examine condition differences, we conducted a GLMM for the total number of socially-caused object looks, including all fixation data of the gazing phase (i.e., the still sequence during which the actors looked at the object or away from the object).

Neither the interaction between third-party eye contact and others' looking at the object revealed a significant effect ($\chi^2(1) = 1.51$, $p = .22$, estimate = 0.22, $SE = 0.18$), nor the main effects of the two factors (eye contact: $\chi^2(1) = 2.33$, $p = .13$, estimate = 0.14, $SE = 0.09$; looking at the object: $\chi^2(1) = 0.05$, $p = .82$, estimate = -0.02, $SE = 0.09$). Moreover, in an additional analysis including fixation data over the entire video sequence, we did not find any significant correlation between the number of socially-caused object looks in the encoding phase and infants novelty preference score in the subsequent preferential-looking phase ($r(441) = -.01$, $p = .89$, $R^2 = .0001$). These results speak against the assumption that the observed triadic joint attention situation had increased infants' own attention to the object and thereby deepened their encoding of the object. Table C3b provides a summary of the descriptive statistics for gaze shifts during the encoding phase in all four conditions.

Taken together, we could not find any indication that infants' overt scanning pattern in the encoding phase of Experiment 1 had caused their increased processing in the third-party joint attentional condition. We did not find any evidence for increased gaze shifts between the two actors while they faced each other, nor did we find that the actors' gazing to the object had a direct impact on infants' own attention to the object.

Table C2

Means (and standard deviations) for looking times and gaze shifts during the encoding phase in all four conditions of Experiment 1

	Conditions Experiment 1 (Third-party)			
	Eye Contact/ Looking at object	No Eye Contact/ Looking at object	Eye Contact/ No Looking at object	No Eye Contact/ No Looking at Object
(a) Looking times				
LT Screen (Overall)	6458.62 *(1708.40)*	6598.85 *(1817.30)*	6215.88 *(1506.73)*	6380.40 *(1575.40)*
LT Object AOI (Overall)	1087.98 *(905.48)*	966.64 *(984.07)*	1335.97 *(1072.26)*	1399.38 *(1174.15)*
LT Object AOI (Gazing)	779.77 *(657.61)*	668.98 *(711.82)*	943.73 *(673.56)*	1022.40 *(751.51)*
LT Face AOIs (Overall)	4875.75 *(1894.56)*	5060.34 *(1791.05)*	4393.43 *(1728.33)*	4421.97 *(1661.28)*
LT Face AOIs (Gazing)	2951.35 *(1391.08)*	3171.61 *(1211.17)*	2390.10 *(1213.94)*	2415.97 *(1082.29)*
(b) Gaze shifts				
Gaze shifts between two face AOIs (Overall)	1.95 *(1.28)*	1.97 *(1.20)*	1.88 *(1.27)*	1.63 *(1.24)*
Gaze shifts between two face AOIs (Interaction)	0.91 *(0.68)*	0.91 *(0.48)*	1.05 *(0.72)*	0.96 *(0.69)*
Socially-caused looks at the object (Gazing)	0.96 *(0.54)*	1.0 *(0.65)*	1.09 *(0.56)*	0.90 *(0.55)*
Overall looks at the Object (Gazing)	1.63 *(0.67)*	1.63 *(0.63)*	1.74 *(0.63)*	1.68 *(0.58)*

Notes. In the first column in parentheses, "Overall" refers to the entire duration of the encoding phase (max. duration = 11000 ms), "Gazing" refers to the phase in which the actors looked toward or away from the object (max. duration = 5000 ms), "Interaction" refers to the phases in which the actors looked at each other's eyes or away from one another (max. duration = 2000 ms). (a) Looking times (LT) represent the sum of fixation durations within the corresponding area of interest (AOI) in milliseconds (ms). (b) Gaze shifts represent the total numbers of gaze movements between the two face AOIs, as well as gaze shifts toward the object. "Socially-caused looks at the object" include gaze shifts from the face AOI to the object AOI, "Overall looks at the object" include all gaze shifts toward the object (i.e., both socially-caused and not socially-caused gaze shifts).

Experiment 2

Valid trial statistics. As described in more detail in the main manuscript, infants were only included in the analysis if they provided valid data in at least one trial per condition after being filtered according to our pre-registered criteria. Table C2 provides an overview of the detailed valid trial statistics for each of the four conditions.

Table C3

Valid trial statistics for the four conditions in Experiment 2

Condition	Number of valid trials				Infant looking at object (fam.)	Infant not looking at object (fam.)
	Total	Min	Max	M *(SD)*		
First-Party Eye Contact Looking at Object	108	1	4	3.38 *(.91)*	89	19
No First-Party Eye Contact Looking at Object	109	2	4	3.41 *(.76)*	100	9
First-Party Eye Contact No Looking at Object	109	2	4	3.41 *(.76)*	93	16
No First-Party Eye Contact No Looking at Object	112	1	4	3.50 *(.84)*	102	10
Total	438				384	54

Notes. Each individual infant could provide between 1 (min.) and 4 (max.) trials per condition. The maximum number of total trials over all infants was 128 per condition. The "infant looking at object (fam.)" column represents the number of valid trials during which infants had looked at the object at all over the total duration of the encoding phase (i.e., fixation duration > 0 ms within the object AOI). The "infant not looking at object (fam.)" column represents the number of valid trials during which infants had not looked at the object at all over the total duration of the encoding phase (i.e., fixation duration = 0 within the object AOI).

Overall attention during encoding. In contrast to Experiment 1, we found that infants' overall attention to the stimuli varied across conditions. Trials during which the actor looked to the object captured more global attention compared to trials during which the actor looked away from the object (main effect of others' looking at the object: $\chi^2(1)$ = 7.29, *p* = .007, estimate = −545.6, *SE* = 190.3). One conceivable explanation for the difference between the experiments is that the videos in which the actors looked away from the object may have been less interesting in Experiment 2 compared to Experiment 1. Even though the visual appearance of the separate actors was the same in both Experiments, the presence of two actors looking to the side may have been more interesting compared to one actor performing the identical movement. As in Experiment 1, we found a continuous decrease in infants' looking time throughout the experiment—both in the encoding phase (main effect of trial: $\chi^2(1)$ = 21.50, *p* < .001, estimate = −1224.6, *SE* = 221.2), as well

as in the preferential-looking phase (main effect of trial: $\chi^2(1) = 29.31$, $p < .001$, estimate = −1102.1, $SE = 159.1$).

Looking times to the object and face during encoding. Neither the interaction between eye contact and others' looking at the object ($\chi^2(1) = 2.39$, $p = .12$, estimate = −253.85, $SE = 163.66$), nor the main effects of these two factors had a significant effect on infants' looking time to the object (eye contact: $\chi^2(1) = 1.01$, $p = .31$, estimate = −83.20, $SE = 82.24$; looking at the object: $\chi^2(1) = .46$, $p = .50$, estimate = 56.10, $SE = 82.21$). A main effect of trial indicated that infants' attention to the object decreased over trials ($\chi^2(1) = 19.04$, $p < .001$, estimate = −286.44, $SE = 56.15$). We did not find any systematic pattern regarding infants' looking duration to the faces either. Neither the interaction between eye contact and others' looking at the object ($\chi^2(1) = .09$, $p = .77$, estimate = 157.88, $SE = 539.75$), nor the main effects of these two factors revealed a significant effect on infants' looking time to the actor's face (eye contact: $\chi^2(1) = .04$, $p = .85$, estimate = 50.92, $SE = 269.81$; looking at the object: $\chi^2(1) = 1.24$, $p = .27$, estimate = −300.64, $SE = 269.81$). Overall, infants' attention to the faces decreased over trials ($\chi^2(1) = 6.93$, $p = .009$, estimate = −355.74, $SE = 135.07$). To ensure consistency between Experiment 1 and 2, we repeated our analyses over the 5-second gazing phase only. Infants' attention patterns remained the same for both the object as well as the actor's face. Table C4a provides a summary of the descriptive statistics for looking times during the encoding phase in all four conditions.

Gaze shifts to the object during encoding. As in Experiment 1, we examined the origin of infants' looks to the object further. We ran the same analyses as described in Experiment 1, and conducted the same pre-processing steps to extract socially caused looks at the object. Neither the interaction between third-party eye contact and others' looking at the object revealed a significant effect on the number of socially-caused looks in the gazing phase ($\chi^2(1) = 0.06$, $p = .81$, estimate = 0.04, $SE = 0.16$), nor did the main effects of the two factors (eye contact: $\chi^2(1) = 0.001$, $p = .99$, estimate = −0.001, $SE = 0.09$; looking at the object: $\chi^2(1) = 2.48$, $p = .12$, estimate = −0.12, $SE = 0.08$). Moreover, including fixation data from the entire video sequence, we did not find any significant correlation between the number of socially-caused object looks and infants' novelty preference score in the subsequent preferential-looking phase ($r(436) = −.01$, $p = .82$, $R^2 = .0001$). Given the previous literature on gaze following, one could have assumed that infants perform more socially-caused looks while the actor looked at the object. However, since we only presented one object in the encoding phase (rather than two objects as in previous gaze following studies), we could not directly calculate the difference score that has been previously used as standard measure of gaze following. One possible explanation for our finding is that there was no other visual stimulation on screen (such as a second object in previous gaze following studies), causing a continuous back-and forth looking between head and object regardless of condition. Table C4b provides a summary of the descriptive statistics for gaze shifts during the encoding phase in all four conditions.

Taken together, we did not find any indication from infants' overt looking behavior during the encoding phase of Experiment 2 (including looking time and gaze shift measures) that may account for their increased object encoding performance in the critical joint attention condition.

Table C4

Means (and standard deviations) for looking times and gaze shifts during the encoding phase in all four conditions of Experiment 2

	Conditions Experiment 2 (First-party)			
	Eye Contact/ Looking at object	No Eye Contact/ Looking at object	Eye Contact/ No Looking at object	No Eye Contact/ No Looking at Object
(a) Looking times				
LT Screen (Overall)	6854.07	7018.86	6721.62	6145.27
	(2304.36)	*(2114.37)*	*(2097.61)*	*(1968.59)*
LT Object AOI (Overall)	1232.92	1203.86	1174.88	1369.24
	(836.77)	*(751.38)*	*(749.56)*	*(890.84)*
LT Object AOI (Gazing)	929.87	861.02	792.21	1013.16
	(610.74)	*(598.35)*	*(572.78)*	*(560.83)*
LT Face AOIs (Overall)	2591.37	2635.04	2391.37	2241.02
	(1018.47)	*(1184.88)*	*(1191.77)*	*(1263.93)*
LT Face AOIs (Gazing)	1537.59	1652.01	1407.35	1180.11
	(697.44)	*(873.90)*	*(888.28)*	*(898.04)*
(b) Gaze shifts				
Socially-caused looks at the object (Gazing)	1.33	1.31	1.22	1.17
	(0.57)	*(0.49)*	*(0.48)*	*(0.42)*
Overall looks at the object (Gazing)	1.69	1.58	1.60	1.70
	(0.48)	*(0.51)*	*(0.56)*	*(0.40)*

Notes. In the first column in parentheses, "Overall" refers to gaze events over the entire duration of the encoding phase (max. duration = 11000 ms) and "Gazing" refers to the phase in which the actor looked toward or away from the object (max. duration = 5000 ms). (a) Looking times (LT) represent the sum of fixation durations within the corresponding area of interest (AOI) in milliseconds (ms). (b) Gaze shifts represent the total numbers of gaze movements toward the object. "Socially-caused Looks at the object" include gaze shifts from the face AOIs to the object AOI, "Overall looks at the object" include all gaze shifts toward the object (i.e., both socially-caused and not socially-caused gaze shifts).

Merged Analyses Experiment 1 and 2

Overall looking times during encoding. Infants' overall attention to the stimuli in the encoding phase did not differ between the two Experiments (Experiment 1: M = 6413.436, SD = 3001.84; Experiment 2: M = 6684.953; SD = 3146.908; $\chi^2(1)$ = .46, p = .50, estimate = 277.5, SE = 410.7). This indicates that videos showing one person were equally interesting compared to videos showing two persons from a third-party perspective.

Relative attention to faces over objects during encoding. In Experiment 1 (third-party), infants' proportional looking time to the faces over the object was significantly higher (M = .79, SD = .23) compared to Experiment 2 (M = .42, SD = .41; χ^2(1) = 64.05, p < .001, estimate = −0.36, SE = 0.03). To calculate this proportion score, we divided infants cumulated fixation duration in the face AOIs by the sum of their cumulated fixation duration in the face AOIs and the object AOI. We included fixations from the total encoding video duration. The difference between the two experiments is not surprising since twice as many actors (and faces) were visible in in Experiment 1 compared to Experiment 2.

Overall, we did not find any systematic variation in infants' overt attention patterns that would suggest that infants' superior processing in the first- or third-party "eye contact – looking at object" condition depended on attention differences during encoding. This provides further support for the assumption that infants' object encoding had been driven by covert attentional processes.

Supplementary Information Fixation Filter

To define fixations, we used the Tobii Velocity-Threshold Identification (I-VT) fixation filter with default parameter values, that is: a velocity and distance threshold of 30° per second, no noise reduction, a maximum time between fixations of 75 ms, a maximum angle between fixations of 0.5°, a minimum fixation duration of 60 ms, and an interpolated of missing data for data segments below 75 ms. More details on the I-VT Filter can be found here:

https://www.tobiipro.com/siteassets/tobii-pro/learn-and-support/analyze/how-do-we-classify-eye-movements/determining-the-tobii-pro-i-vt-fixation-filters-default-values.pdf

C3 - Pilot Study

We conducted a pilot study to ensure that the video stimuli, the timing of the procedure, and the overall duration of the experiment were adequate for infants in the required age range. In addition, we used the pilot data to run a simulation-based a priori power analysis to determine whether our planned sample size was sufficient to detect the expected effect size. Piloting took place in March 2020 under the same conditions as the final study took place. We tested a version with 24 trials before deciding on the 16-trial version. Piloting was finished as soon as we had determined an age at which infants remained attentive throughout the experiment, while being old enough to ensure that they were sensitive to third party-interactions based on previous findings.

Participants

Overall, $N = 21$ infants between 9 months, 8 days and 13 months, 22 days participated in the pilot study ($M = 346.7$ days, $SD = 51.2$ days). The participants were recruited from the same data base as the participants for the final sample. We started piloting with versions of Experiment 1 ($n = 18$), since this experiment represented the main focus of this study. As soon as we had addressed all procedural concerns and decided on a concrete participant age range, we finished piloting for Experiment 1 and piloted three more infants in a corresponding version of Experiment 2. This was done to rule out that infants would be less attentive when only one person was visible on screen, and to check whether our piloting decisions based on Experiment 1 could also account for Experiment 2. For the a priori Power analysis, we only included infants who had participated in Experiment 1, and who provided at least one valid trial per condition after being filtered according to our criteria described in Experiment 1 in the main manuscript. This applied to $n = 10$ infants between 9 months, 8 days and 13 months, 22 days ($M = 334.0$ days, $SD = 52.38$ days).

A Priori Power Analysis

Since we piloted different versions of Experiment 1, some pilot participants were presented with 24 trials instead of 16 trials. To make best use of all data while ensuring consistency with our finally aimed data structure, we included the first 16 trials of these children in the power analysis, if they provided the sufficient number of one valid trial per condition. Due to the counterbalancing of conditions within trial-blocks, the first 16 trials in the 24-trial-version of the experiment included 4 trials of each condition, consistent with the final Experiment version.

We ran a simulation-based power analysis with the R package "simr" (Version 1.0.5, Green & MacLeod, 2019). The power analysis script and the pilot data are available online (https://osf.io/yfegm/). We followed the following steps described by Green & MacLeod (2016). First, we fitted the main model of Experiment 1 (see analysis section of Experiment 1 in the main manuscript for details). The model estimates were calculated based on our pilot data. Second, we specified the effect size. As effect size, we calculated R-squared as the difference in the model fit between a full model including all fixed and random effects, and a null model including only the control variables and random effects without the fixed effects. R-squared was calculated with the R package "MuMIn" (Barton, 2019). The model comparisons revealed a marginal effect size of $R^2 = .077$ (conditional $R^2 = .089$). To detect an effect of this magnitude with a sample size of $N = 32$, the power analysis based on 1000 simulations indicated a power of 100% CI [99.63, 100.0]. We did not run a separate power analysis for the pilot version of Experiment 2, since the number of data points was too low to calculate a valid estimation of effect sizes. However, due to the closely matched study design, we expected a similar power in both experiments.

Curriculum Vitae

Name:	Kyra Maleen Thiele
Date of Birth:	26th May, 1989
Place of Birth:	Münster, Germany

Education

06/2020 – present	Ph.D. Student Max Planck Institute of Evolutionary Anthropology Department of Comparative Cultural Psychology Leipzig, Germany
08/2018 – 12/2018	Visiting Researcher Uppsala University Department of Psychology, Uppsala Child and Baby Lab Uppsala, Sweden
10/2016 – 05/2020	Research Associate | Ph.D. Student Leipzig University Department of Early Child Development and Culture Leipzig, Germany
10/2014 – 10/2016	Master of Science in Psychology (M.Sc.) Leipzig University Leipzig, Germany
10/2010 – 10/2013	Bachelor of Science in Psychology (B.Sc.) Leipzig University Leipzig, Germany
2008	Abitur Geschwister-Scholl-Gymnasium Unna, Germany

Scientific Publications
and Conference Contributions

Publications in the Context of this Dissertation

Thiele, M., Hepach, R., Michel, C., Gredebäck, G., & Haun, D. (2021). Social interaction targets enhance 13-month-old infants' associative learning. *Infancy, 26*, 409-441. https://doi.org/10.1111/infa.12393

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2021). Observing others' joint attention increases 9-month-old infants' object encoding. *Developmental Psychology, 57*, 837-850. https://doi.org/10.1037/dev0001189

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2021). Infants' preference for social interactions increases from 7 to 13 months of age. *Child Development.* https://doi.org/10.1111/cdev.13636

Additional Publications

Astor, K., Thiele, M., & Gredebäck, G. (2021). Gaze following emergence relies on both perceptual cues and social awareness. *Cognitive Development, 60*, 101121. https://doi.org/10.1016/j.cogdev.2021.101121

Scientific Presentations and Conference Contributions

Thiele, M., Michel, C., Hepach, R., & Haun, D. (2021). *Observed joint attention promotes 9-month-old infants' learning about novel objects.* Poster presented at Society for Research in Child Development Biennial Meeting (SRCD), Virtual Conference.

Thiele, M., Michel, C., Hepach, R., & Haun, D. (2021). *Third-party joint attention increases 9-month-old infants' object processing.* Poster presented at Budapest CEU Conference on Cognitive Development (BCCCD), Virtual Conference.

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2020). *Investigating developmental trajectories of infants' preference for social interactions.* Poster presented at The International Congress of Infant Studies (ICIS), Virtual Conference.

Thiele, M., Hepach, R., Michel, C., Gredebäck, G., & Haun, D. (2020). *13-month-old infants seek out third-party social interactions.* Poster presented at the International Congress of Infant Studies (ICIS), Virtual Conference.

Thiele, M. (2020). *Using eye tracking to investigate the foundations of infants' observational learning from third-party social interactions.* Talk given at the Tobii pro Webinar: Using Eye Tracking in Developmental Psychology Research, Virtual Webinar.

Thiele, M., Hepach, R., Michel, C., Gredebäck, G., & Haun, D. (2020). *13-month-olds' visual learning is reinforced by social interaction targets.* Poster presented at Budapest CEU Conference on Cognitive Development (BCCCD), Budapest, Ungarn.

Thiele, M., Hepach, R., Michel, C., Gredebäck, G., & Haun, D. (2019). *13-month-olds' visual learning is reinforced by social interaction targets.* Poster presented at Third Annual ESLR Workshop, Leipzig, Germany

Thiele, M. (2019). *Investigating infants' preference for social interactions.* Talk given at the Workshop für Nachwuchswissenschaftlerinnen in der Entwicklungspsychologie, Münster, Deutschland.

Thiele, M. (2019). *Social interaction targets enhance infants' visual learning performance.* Talk given at the Gemeinsame Tagung der Pädagogischen und Entwicklungspsychologie (PAEPSY), Leipzig, Deutschland.

Thiele, M. (2019). *Are social interactions rewarding for young children?* Symposium organized at the Gemeinsame Tagung der Pädagogischen und Entwicklungspsychologie (PAEPSY), Leipzig, Deutschland.

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2018). *Investigating social interaction behavior in 7- to 11-month-old infants.* Poster presented at the International Congress of Infant Studies (ICIS), Philadelphia, USA.

Thiele, M. (2018). *Investigating young infants' preference for social interactions.* Talk given at the International Congress of Infant Studies (ICIS), Philadelphia USA.

Thiele, M., Hepach, R., Michel, C., & Haun, D. (2018). *Investigating young infants' attentional preference for social interactions and its relation to their active social behavior.* Poster presented at Budapest CEU Conference on Cognitive Development (BCCCD), Budapest, Ungarn.

Thiele, M., Hepach, R., & Haun, D. (2017). *Do young infants have an attentional preference for social interactions?* Poster presented at Understanding Others Workshop, Jena, Deutschland.

Thiele, M., Peoples, S., Hepach, R., & Tomasello, M. (2017). *Young children are sensitive towards their own but not towards others' deservingness in an individual context.* Poster presented at the Gemeinsame Fachtagung der Pädagogischen und Entwicklungspsychologie (PAEPSY), Münster, Deutschland.

Contributions of Authors

Nachweis über Anteile der Co-Autor:innen, Kyra Maleen Thiele
The Social Attentional Foundations of Infants' Learning from Third-Party Social Interactions

Nachweis über Anteile der Co-Autor:innen

Titel	Infants' Preference for Social Interactions Increases From 7 to 13 Months of Age
Journal	Child Development
Autor:innen	Maleen Thiele, Robert Hepach, Christine Michel, Daniel Haun

Anteil Maleen Thiele (Erstautorin):
- Projektidee
- Konzeption
- Stimulus Erstellung
- Administration
- Datenerhebung
- Datenaufbereitung & Datenauswertung
- Visualisierung
- Schreiben der Publikation
- Redigieren der Publikation

Anteil Robert Hepach (Autor 2):
- Projektidee
- Supervision
- Konzeption
- Schreiben der Publikation
- Redigieren der Publikation

Anteil Christine Michel (Autorin 3):
- Projektidee
- Supervision
- Konzeption
- Schreiben der Publikation
- Redigieren der Publikation

Anteil Daniel Haun (Senior-Autor):
- Projektidee
- Supervision
- Konzeption
- Redigieren der Publikation
- Finanzierung

Maleen Thiele
(Erstautorin)

Daniel Haun
(Senior-Autor)

Nachweis über Anteile der Co-Autor:innen

Titel	Social Interaction Targets Enhance 13-month-old Infants' Associative Learning
Journal	Infancy
Autor:innen	Maleen Thiele, Robert Hepach, Gustaf Gredebäck, Christine Michel, Daniel Haun

Anteil Maleen Thiele (Erstautorin):
- Projektidee
- Konzeption
- Stimulus Erstellung
- Administration
- Datenerhebung
- Datenaufbereitung & Datenauswertung
- Visualisierung
- Schreiben der Publikation
- Redigieren der Publikation

Anteil Robert Hepach (Autor 2):
- Projektidee
- Supervision
- Konzeption
- Datenauswertung
- Schreiben der Publikation
- Redigieren der Publikation

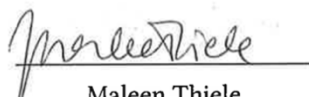Anteil Christine Michel (Autorin 3):
- Projektidee
- Supervision
- Konzeption
- Schreiben der Publikation
- Redigieren der Publikation

Anteil Gustaf Gredebäck (Autor 4):
- Projektidee
- Supervision
- Konzeption
- Redigieren der Publikation

Anteil Daniel Haun (Senior-Autor):
- Projektidee
- Supervision
- Konzeption
- Redigieren der Publikation
- Finanzierung

Maleen Thiele
(Erstautorin)

Daniel Haun
(Senior-Autor)

Nachweis über Anteile der Co-Autor:innen

Titel	Observing Others' Joint Attention Increases 9-month-old Infants' Object Encoding
Journal	Developmental Psychology
Autor:innen	Maleen Thiele, Robert Hepach, Christine Michel*, Daniel Haun* *geteilte Letztautorenschaft

Anteil Maleen Thiele (Erstautorin):
-   Projektidee
-   Konzeption
-   Stimulus Erstellung
-   Administration
-   Datenerhebung
-   Datenaufbereitung & Datenauswertung
-   Visualisierung
-   Schreiben der Publikation
-   Redigieren der Publikation

Anteil Robert Hepach (Autor 2):
-   Supervision
-   Konzeption
-   Datenauswertung
-   Redigieren der Publikation

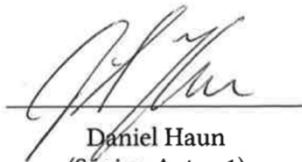Anteil Christine Michel (Senior-Autorin 1):
-   Projektidee
-   Supervision
-   Konzeption
-   Schreiben der Publikation
-   Redigieren der Publikation

Anteil Daniel Haun (Senior-Autor 2):
-   Supervision
-   Konzeption
-   Redigieren der Publikation
-   Finanzierung

Maleen Thiele
(Erstautorin)

Daniel Haun
(Senior-Autor 1)

Christine Michel
(Senior-Autorin 2)

Declaration of Authorship

Hiermit versichere ich, Krya Maleen Thiele, geboren am 26.05.1989 in Münster, dass ich die vorliegende Arbeit ohne unzulässige Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken und Zitate sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials sowie der Herstellung der Manuskripte für die drei einzelnen Studien erhielt ich Unterstützung von Prof. Dr. Daniel Haun (Max-Planck-Institut für evolutionäre Anthropologie, Leipzig), Prof. Dr. Robert Hepach (Oxford University, Oxford, UK), Dr. Christine Michel (Universität Leipzig und Max-Planck-Institut für Kognitive und Neurowissenschaften, Leipzig) und Prof. Dr. Gustaf Gredebäck (Uppsala Universitet, Uppsala, Schweden). Ich versichere, dass außer den oben genannten Personen keine weiteren an der geistigen Herstellung der Arbeit beteiligt waren. Ich habe zu keiner Zeit die Hilfe eines Promotionsberaters in Anspruch genommen. Dritte Personen haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten enthalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Ich versichere, dass die vorliegende Arbeit in gleicher oder ähnlicher Form keiner anderen wissenschaftlichen Einrichtung zum Zwecke einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt wurde. Ich habe zu keinem früheren Zeitpunkt erfolglose Promotionsversuche unternommen.

Leipzig, 21. Juli 2021

Kyra Maleen Thiele