Journal of Experimental Child Psychology 207 (2021) 105115

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Inferring hidden objects from still and communicative onlookers at 8, 14, and 36 months of age

Marianna Jartó *, Ulf Liszkowski

Department of Developmental Psychology, University of Hamburg, 20146 Hamburg, Germany

A R T I C L E I N F O

Article history:
Received 3 March 2020
Revised 23 December 2020

Keywords:
Visual perspective taking
Referential expectation
Social cognition
Eye tracking
Pointing
Visual cueing

A B S T R A C T

The current study investigated across five eye-tracking experiments children's developing skill of adopting others' referential perspective (Level 1 perspective taking) and to what extent it involves automatic processes or requires ostensive communicative cues. Three age groups (8-, 14-, and 36-month-olds) were tested on their expectation of an object appearing behind one of two peripheral occluders. A centrally presented person in profile either provided an ostensive communicative pointing cue or sat still, oriented to one of the two occluders. The 14-month-olds anticipated the hidden object when the onlooker had communicatively pointed to the location, as revealed by faster target detection in congruent trials (latency effect) and longer dwell times to the empty side in incongruent trials (violation-of-expectation effect). This was not the case when a still person was only oriented to one side. Adding emotional expressions to the still person (Experiment 2) did not help to produce the effects. However, at 36 months of age (Experiment 3), children showed both effects for the still person. The 8-month-olds did not show the violation-of-expectation effect for communicative pointing (Experiment 4) or for a matched abbreviated reach (Experiment 5b), showing it only for a complete reach behind the occluder (Experiment 5a), although they were faster to detect the congruent object in Experiment 4 and 5a. Findings reveal that automatic perspective taking develops after communicative perspective taking and that

* Corresponding author.
E-mail address: marianna.jarto@uni-hamburg.de (M. Jartó).

https://doi.org/10.1016/j.jecp.2021.105115

communicative perspective taking is a developmental outcome of the first year of life. The developmental pattern suggests a continuous social construction process of perspective-taking skills.

Introduction

Interacting with others in meaningful ways requires taking their perspectives. The scopes and limits of this ability during early development are currently not well understood. Research distinguishes different levels of perspective taking that develop at different ages in ontogeny and involve different levels of cognitive complexities (Flavell, Everett, Croft, & Flavell, 1981). On the cognitively complex end ("Level 2 perspective taking"), children represent others' subjective representations of the world, that is, *how* others see something as, for example, implicated in passing standard false belief tests. On a somewhat simpler, and earlier emerging, variant ("Level 1 perspective taking"), infants understand *that* others see something, that is, that others mentally represent something as, for example, when infants infer that someone does or does not see a hidden object. On the least complex end (here dubbed "Level 0"), young infants orient to certain aspects in the world because their attention is cued into a direction by others' movements and visual orientations based on lower-level perceptual biases and orienting responses (e.g., Heyes, 2014). The current study investigated the underlying processes of Level 1 perspective taking and how it contrasts with Level 0 during development.

Research with adults and young infants has suggested that Level 1 perspective taking is based on automatic processes (Apperly & Butterfill, 2009), but it is less clear whether these processes entail cognitive expectations that others see something (e.g., Teufel, Alexis, Clayton, & Davis, 2010) or rather pertain to perceptual cueing responses (Conway, Lee, Ojaghi, Catmur, & Bird, 2017). It is further unclear whether perspective taking is automatic in a rather mandatory sense and occurs even outside interactions or whether it requires some form of communicative cues. Communicative accounts such as relevance theory (Sperber & Wilson, 1986) and natural pedagogy theory (Csibra & Gergely, 2009) suggest that ostensive cues are required to convey the relevance of information and are privileged in the scope and depths of processing. Finally, developmentally, it is unclear whether these automatic processes are a starting point in development or rather a developmental outcome of automatization. Despite a wealth of findings on perspective taking, the evidence is not well integrated along these dimensions. Distinct paradigms reveal different levels and processes of perspective taking based on different types of cues at different ages during development.

Level 0 findings

Posner-like central cueing paradigms tap basic Level 0 processes. Infants as early as 4–6 months of age are spatially cued by averted gaze (Farroni, Mansfield, Lai, & Johnson, 2003) and pointing-shaped hands (Bertenthal, Boyer, & Harding, 2014), but typically only when the cue is ostensive communicative or at least dynamic, not when it is static (Rohlfing, Longo, & Bertenthal, 2012). As in adult research, the stimulus onset asynchrony (SOA) plays a crucial role (see Bukowski, Hietanen, & Samson, 2015). A reflexive bottom-up cueing effect with a very short SOA (100 ms) is present from early on (4–6 months), whereas a reflective top-down cueing effect (measured with an SOA longer than 300 ms) seems to develop sometime from 10 to 12 months (Daum, Ulber, & Gredebäck, 2013). The latter reflective cueing effect is present first when the cue is accompanied by communicative speech. Reflective attentional orienting is further revealed with simple overt gaze-following paradigms. Here, too, findings are that infants are directed by others' gaze more often when it is accompanied by ostensive cues such as communicative pointing and verbalizing (at 9 months; Flom, Deák, Phill, & Pick, 2004).

Converging findings thus reveal an early form of reflexive attentional bias to orient to indicated directions that becomes reflective toward the end of the first year of life. These effects are obtained when the cues are ostensive communicative, or dynamic, which suggests that even Level 0 attention following, as a first step of perspective taking, first emerges in ostensive settings. However, the findings do not necessarily tap Level 1 perspective taking in the sense of understanding that others see, and mentally represent, something because the findings are amenable to leaner perceptual orienting interpretations, especially those based on a short SOA.

Level 1 findings

In contrast, so-called occlusion paradigms go beyond perception and reflexive orienting and clearly tap Level 1 perspective-taking skills. The main finding is that around their first birthday, infants understand seeing as a mental act and expect a referent when someone is communicatively looking or pointing in a direction. For example, 12-month-olds follow a gaze less when the view of the model is blocked by closed eyes, a blindfold (14 and 18 months; Brooks & Meltzoff, 2002), or barriers (18 months; Butler, Caron, & Brooks, 2000). They also follow expressive communicative gaze around barriers instead of just extrapolating along the line of sight to the barrier (Moll & Tomasello, 2004). Furthermore, by 12 months, infants search for a hidden object when its hiding location is disambiguated through a communicative pointing gesture (Behne, Liszkowski, Carpenter, & Tomasello, 2012), but not when the gesture is non-ostensive (Behne, Carpenter, & Tomasello, 2005). In addition, 12-month-olds look longer (Csibra & Volein, 2008) and have larger pupils (Pätzold & Liszkowski, 2019) when a hidden object is not revealed at an ostensively indicated location, demonstrating a violation of their cognitive referential expectations. Moll and Tomasello (2007) found that 14-month-olds represented the perspective of an interaction partner when determining which of three objects she had previously not seen, but only when in communicative joint engagement with the partner, not when the partner was a mere onlooker. The developmental question, however, is whether infants younger than 12 months have cognitive referential expectations.

Emergence of Level 1 perspective taking

Little research to date has employed occlusion paradigms with younger infants. Csibra and Volein (2008) reported that their finding of referential object expectations pertained to 12- and 8-month-olds because their omnibus analyses did not yield significant age interactions. However, in Pätzold and Liszkowski's (2019) pupillometry study, only 12-month-olds, but not 8-month-olds, had relatively larger pupils when an actress pointed communicatively to an occluded site that was then revealed to be empty compared with when there had not been a point. In support of developmental differences, Rüther and Liszkowski (2020) recently found that only 12-month-olds, but not 10-month-olds, searched reliably above chance for a hidden object when its location was disambiguated by a communicative pointing cue. On the other hand, a different violation-of-expectation looking time study suggested that even 6-month-olds represent another person's perspective when they expect the person to reach for one of two objects, and only one of them is visible to the person (but both objects are visible to the infants; Luo & Johnson, 2009). Furthermore, Krehm, Onishi, and Vouloumanos (2014) found that 9- to 11-month-olds looked longer when a recipient of a pointing gesture subsequently chose a non-referred target object instead of a referred target object (albeit the paradigm did not involve visual occlusion). To provide more data on the developmental emergence, therefore, the current study tested for skills of Level 1 perspective taking in younger infants before 12 months.

Perspective taking without ostension

Another question is whether ostensive cues are necessary for Level 1 perspective taking. In the Level 1 occlusion paradigms reviewed before, the cues to inferring the object are typically ostensive communicative, and so it is currently less clear whether non-ostensive social cues would yield similar effects. In their pupillometry study, Pätzold and Liszkowski (2019) reported negative findings with exogenous light cues (in contrast to the positive findings with communicative pointing cues). Furthermore, as mentioned before, Daum et al. (2013) found that a reflective cueing effect first emerged when the cue was accompanied by communicative speech. In contrast, in a different paradigm (von Hofsten, Dahlström, & Fredriksson, 2005), 12-month-olds watched stills of a static person looking at one object from an array of several objects, without communicative ostension, and infants appropriately oriented to the target object.

Yet other paradigms have investigated automatic intrusions of others' perspectives in non-communicative settings. Findings with adults and infants as young as 7 months suggest that a neutral passive onlooker shown in profile yields an altercentric perspective interference, that is, an imprecision in one's own perspective calculation due to an intruding different perspective of the bystander (Kampis, Parise, Csibra, & Kovács, 2015; Kovács, Téglás, & Endress, 2010; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010; Southgate & Vernetti, 2014). This effect appears to hold even when the bystander is no longer present (Kovács et al., 2010). The effect is obtained without ostensive communication, which contrasts with the theoretical accounts and evidence from cueing and occlusion paradigms reviewed before. However, it is debated to what extent these paradigms indeed tap Level 1 perspective taking or rather recruit a Level 0 form of perceptual cueing and orienting (Heyes, 2014).

Summary

Different paradigms have yielded different cognitive levels and ages of perspective taking. Cueing paradigms with a short SOA suggest automatic orienting following ostensive dynamic cues early in the first year of life, likely based on perceptual orienting biases. Occlusion paradigms reveal Level 1 processes robustly around 12 to 14 months of age but have yielded inconclusive evidence at younger ages. Occlusion paradigms have mostly employed ostensive communicative cues. In contrast, perspective intrusion paradigms suggest an earlier age of emergence even without ostensive communicative cues, but the underlying process level has remained less clear.

The current study

In the current study, across five experiments, we employed an eye-tracking video-based paradigm that affords standardized use across age groups. To test for Level 1 cognitive referential expectations, we developed a cueing paradigm with a long SOA and implemented the logic of occlusion paradigms based on Csibra and Volein (2008) and Pätzold and Liszkowski (2019). We used ostensive and non-ostensive social cues across different age groups to establish under which conditions, and at what ages, children form cognitive referential expectations. Participants watched a video of a person sitting in profile between two occluding screens. The person disappeared at test, and after a brief central fixation cross the occluders fell flat to reveal an object either at the side to which the person had been oriented or at the other side. We measured the latency of looking to the object and the looking time (dwell time) to the empty location. The logic of these measures in common cueing and violation-of-expectation paradigms is that latencies to expected locations are shorter than those to unexpected locations and looking times to unexpected outcomes are longer than those to expected outcomes. Thus, in the current study, in accordance with the reviewed literature, if the cue indeed led to referential expectations, we predicted a shorter latency to fixate the object in congruent trials than in incongruent trials, and we predicted a longer dwell time for the empty side in incongruent trials than in congruent trials. These predictions were thus directed, and findings in the opposite direction would not be interpretable, typically qualifying for one-tailed probability tests (e.g., Salkind, 2006). The paradigm rests on the logic that participants form a cognitive representation of the occluded object in expectation of seeing it once occlusion ceases (Behne et al., 2012; Csibra & Volein, 2008). Because the cue has ceased before the object is revealed (long SOA), and the attention is reoriented to the center before the object is revealed, latency effects are not purely perceptual. Whereas the latency measure could theoretically also entail an expectation only about the location, dwell time is a common measure of cognitive expectation in infancy research such that longer dwell time reveals a violation of expectation, here the expectation of a referent object (Csibra & Volein, 2008). If participants had solid cognitive expectations of an object, they should show the predicted directed effects on both measures.

Across experiments, we manipulated the communicativeness of the cue that the person in profile provided. The experiments followed up on each other, with Experiment 1 being the main experiment that addressed the main question of whether infants represent another's perspective (i.e., Level 1) following a still directional cue (as in perspective intrusion tasks) or only following a communicative cue (as expected from theoretical accounts and previous findings). Accordingly, in Experiment 1, a still person simply looked ahead in the direction of one of the two occluders (still condition) or the person pointed ostensively behind one of the two occluders (pointing condition). We tested 14-month-olds because the literature reviewed above shows that they have referential expectations about occluded objects in response to communicative cues. Accordingly, the prediction for the main Experiment 1 was that the ostensive communicative condition should yield positive latency and dwell time effects. As we will show, we did obtain the significant effects in the communicative condition, but not in the still condition. To follow up on the negative finding of the still cue, in Experiment 2 we tried to enrich the still cue by adding facial still expressions of disgust or excitement to the onlooker. These emotions are salient and referential and thus could help to make the still cue more relevant and instigate referential expectations. As we will show, these modifications of the still cue did not produce significant effects, thereby corroborating the absence of evidence from Experiment 1. In Experiment 3, therefore, we tested older children (36-month-olds) in the still condition to assess whether it had an effect at all (as adult studies ultimately would suggest). Because we obtained positive findings, we reasoned that the findings of Experiment 1 were valid and that non-ostensive perspective taking emerges only with development. In Experiment 4, we followed up on the positive finding of the ostensive communicative cue in Experiment 1 and examined whether we would obtain the effects in younger infants at 8 months, as suggested by Csibra and Volein (2008). Because we did not obtain conclusive positive effects, in Experiment 5 we tested whether reaching actions would lead to the expected effects because at that age infants are familiar with reaching but do not yet point. Although the understanding of individual reaching is different from perspective taking and perhaps subserved by different cognitive systems (Csibra, 2003), we reasoned that positive evidence could internally validate our paradigm and demonstrate its general suitability with infants at 8 months of age. All experiments used the same general paradigm, but each experiment was based on hypotheses and experimental manipulations in their own right, with mild variations in timing or sampling rates, thereby mandating separate analyses. Stimuli and data are available at the Open Science Framework (https://osf.io/4r2xf).

Experiment 1

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

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

Method

Participants

Parents were contacted via birth register in the metropolis Hamburg, Germany. The final sample consisted of 32 infants (15 girls) with a mean age of 14 months 15 days and an age range of 14;1 to 14;27 (months; days). An additional 16 infants were invited but excluded due to fussiness

($n = 5$), bad calibration (more than 2° deviance; $n = 10$) or recording problems ($n = 1$). We note that the calibration procedure apparently ran less smoothly in the current experiment compared with subsequent experiments, but we cannot pinpoint this to any obvious factor(s). One possibility is that 14-month-olds are motorically more advanced than 8-month-olds yet less self-regulatory than 3-year-olds. At any rate, these dropouts occurred before, and thus independent of, the experiment itself.

Setup and procedure

Families were individually invited to the laboratory at a time during the day that they anticipated to be optimal for their infants (e.g., awake, not hungry, in a good mood). During a 10-min warm-up in a welcome area, parents were informed about the procedure and data privacy. Eye-tracking experiments were approved by the local ethical committee and were in accordance with the Declaration of Helsinki for the protection of human participants. Then, parents followed with their infant into the eye-tracking room. A Tobii X120 eye tracker (Tobii Technology, Stockholm, Sweden) was used to record infants' gaze with a 60-Hz sampling rate. Two screens were connected to the computer via DVI-D (18 + 1) and set to extended mode. One screen was not visible to the infant and was used to monitor calibration and the infant's behavior during the session. The other screen was the presentation screen with 1920 × 1200-pixel resolution (52 × 34 cm, refresh rate = 60 Hz, response time = 13.5 ms) and was positioned above the eye tracker. The presentation screen and eye tracker were protruded through a black canvas (2.5 m height × 1.5 m width) that had triptych-like wings to the left and right side to minimize distraction for the infant. Above the screen was a web camera to monitor the behavior of the infant during the session. Parents sat on a swivel chair in front of the screen with their infant on their lap at a distance of approximately 62–65 cm. They were told not to interact with the infant or to direct the infant's attention to the monitor during the session and were instructed to hold their infant in the same position. Parents wore opaque glasses during calibration and stimuli presentation so as not to record parents' gaze and to keep them ignorant about the presented events to prevent any interferences. A 5-point calibration with rotating balls and infant-friendly music was used.

Stimulus and design

Participants watched self-made video-based stimuli. Videos were extracted with 25 frames per second (fps) and, as recommended for Tobii Studio software, were converted to AVI format with 15 fps with video codec WMV1. One video (a congruent trial) was unintentionally not converted to 15 fps and was presented with 25 fps. Timing for this trial was adjusted (every event happened 120 ms earlier). Fig. 1 illustrates the unfolding and the timing of the test stimuli events. A person

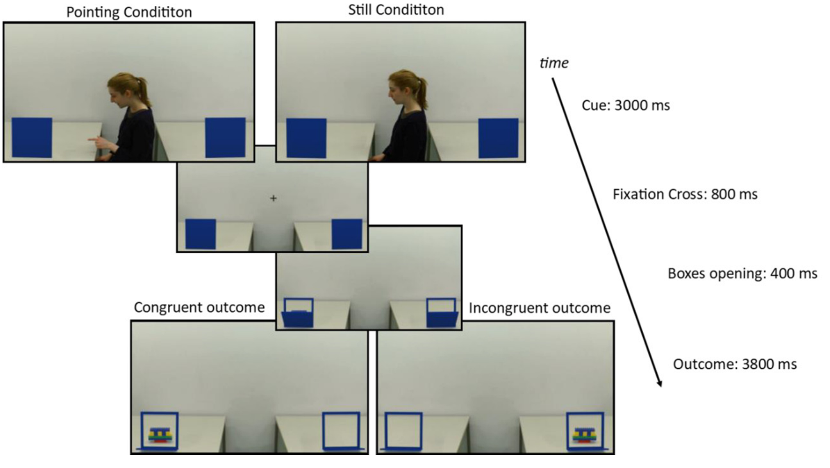

Fig. 1. Sequence of the video stimuli used in Experiment 1.

was displayed in the middle of the screen in profile, sitting between two equidistant tables. During the recording of the stimuli, the person had been instructed to fixate on a predefined spot on the table. During the process of editing the videos, two screens were superimposed onto the tables to cover the predefined spot. The screens were animated to lower forward like a drawbridge with a door-like frame remaining. The video display measured 21.7° vertical × 37.7° horizontal. The occluders measured 6.7 × 6.7°, and the distance between the occluders was 23.5°. On the screen, the person in profile measured 16.6° and her head about 4.8 × 4.8°. The distance between the middle of the head and the middle of the occluder was 17.1°. After the person disappeared, a fixation cross (1.4 × 1.4°) appeared instead at the position where the person's head had previously been. A LEGO DUPLO object was superimposed onto the video at the location of the predefined fixation spot, behind one of the two occluders, so that it was revealed inside the door-like frame when the occluder opened. There were eight different LEGO DUPLO objects, a different one for each of the eight trials.

There were two types of cues. In the still condition, the person in profile remained still while staring behind the occluder. In the pointing condition, the person pointed behind the two occluders, with the point starting from her lap, and said "Ah." In the pointing condition, the person was 10.5° wide at the peak of the pointing hand. The distance between the end of the pointing finger and the middle of the pointed-to occluder was 8.6°. An object appeared either on the side of the cued occluder in front of the person (congruent trial) or on the side of the uncued occluder behind the person (incongruent trial). Each session started with a 30-s animation video with music (*Barbapapa*) to increase the infant's attention and motivation. Then two familiarization videos without a person familiarized the infant with the opening of the occluders and the appearance of a toy. Object sides alternated. Then, the two different conditions were presented blockwise, with the presentation order being counterbalanced across participants. Each condition included eight trials: four revealing a congruent outcome and four revealing an incongruent outcome. Cueing side, object side, and congruency were randomized across trials but never the same in more than two consecutive trials. Between the two condition blocks, another 30-s *Barbapapa* video was presented. Every trial started with a small moving attention getter with a short sound in the middle of the screen for 3 s to center the infant's attention before the beginning of a trial.

Data reduction and analysis

Our main dependent measures were latency to the object and dwell time on the empty side. As additional information, we measured the looking behavior during the cue phase. We calculated the proportion of trials in which infants looked longer to the cued occluder than to the uncued occluder and vice versa. In addition, we calculated the proportion of trials in which infants did not look away from the person. We defined four square areas of interest (AOIs). Two AOIs covered each occluder with a horizontal and vertical dimension of 7.7°, another AOI was positioned around the fixation cross with a dimension of 2.4 × 2.4°, and a further AOI comprised the entire video 21.7 × 37.7°.

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

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

To calculate mean values, infants who yielded less than 50% of valid trials per congruency trial type were excluded. We employed conservative exclusion criteria to reduce noise. First, infants needed to watch the cue for at least one third of its duration (i.e., 1000 ms) to make sure that infants attended to the communicative features of the cue and reflectively processed it and that the processing time was comparable between conditions. Second, infants needed to fixate on the cross for a minimum of 60 ms to exclude that infants' gaze only swayed past the cross. Latency and dwell times were measured only after this requirement to exclude fixations on the boxes from the cueing phase. Results indeed revealed no latency shorter than 200 ms. The following numbers of trials were excluded due to insufficient attention during the cue phase (n = 26 trials of 10 infants in the pointing condition, n = 26 trials of 13 infants in the still condition) and inattention of the fixation cross (n = 49 trials of 12 infants in the pointing condition, n = 40 trials of 20 infants in the still condition). The maximum dropout rate of trials was about 21%, which is comparable to other studies (e.g., Bertenthal et al., 2014).

To assess overall effects, we submitted our dependent measures, latency and dwell time, to 2 × 2 repeated-measures analysis of variance (ANOVA) with congruency (congruent or incongruent) and condition (pointing or still) as within-participant factors. To control for order (pointing condition first or still condition first) as a between-participant factor, an additional 2 × 2 × 2 mixed-design ANOVA was run. Our hypothesis, based on existing findings, was that 14-month-olds do have referential object expectations and thus show directed latency and dwell time effects. Of question was whether latency and dwell time effects are also present in the still condition. Accordingly, we subsequently ran planned t tests for each condition. For the planned comparisons within each condition, sample sizes differed according to exclusion criteria. The maximum increase for all dependent variables was n = 4 compared with the omnibus test. For confirmatory analyses, we also ran the planned directed comparisons only on the sample from the overall omnibus ANOVA. The two main dependent measures are based on directed hypotheses such that shorter latencies (not longer times) are predicted against the null model when comparing congruent and incongruent trials and that longer dwell times (not shorter times) are predicted against the null model when comparing incongruent and congruent trials. To assess the effects on an individual level, we used binomial tests to compare the number of infants who had a faster mean latency in congruent trials than in incongruent trials with the number of infants who showed the opposite effect and to compare the number of infants who had longer dwell times to the empty side in incongruent trials than in congruent trials with the number of infants who showed the opposite effect. We winsorized variables with extreme outliers (3 standard deviations above the mean) by changing values above the 95% percentile to a value within the 95% percentile (Reifman & Keyton, 2010) using an online winsorizing calculator (Hemmerich, 2018).

Results

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

Cueing phase

A 2 × 2 Congruency (cued or uncued box) × Condition (still or pointing) repeated-measures ANOVA showed a significant main effect of condition, $F(1, 28) = 25.99$, $p < .001$, $\eta_p^2 = .481$, and congruency, $F(1, 28) = 7.07$, $p = .013$, $\eta_p^2 = .201$. The Congruency × Condition interaction was not significant, $F(1, 28) < 0.01$, $p > .990$, $\eta_p^2 < .001$. Infants disengaged more often from the person and looked to the boxes in the non-ostensive condition (M = 18.2%, SE = 2.6) than in the communicative condition (M = 4.6%, SE = 1.1). In both conditions, infants looked more to the cued box (M = 14.7%, SE = 2.3) than to the uncued box (M = 8.2%, SE = 1.5). Infants did not disengage from the person in most of the trials (M_{pointing} = 90.7%, SD = 11.5; M_{still} = 64.6%, SD = 28).

Latency

Due to an outlier, latency was winsorized in incongruent trials in both conditions. There were no significant effects for order, so this factor was not further included. A 2 × 2 (Congruency [congruent or incongruent] × Condition [still or pointing]) repeated-measures ANOVA showed no significant main effect of congruency, $F(1, 18) = 1.06$, $p = .316$, $\eta_p^2 = .056$, or condition, $F(1, 18) = 1.37$, $p = .257$, $\eta_p^2 = .071$, but the Congruency × Condition interaction was significant, $F(1, 18) = 5.12$, $p = .036$, $\eta_p^2 = .221$.

The first column of Fig. 2 displays the latencies for Experiment 1. Planned paired t tests for each condition according to the directed hypothesis confirmed that in the pointing condition, infants were

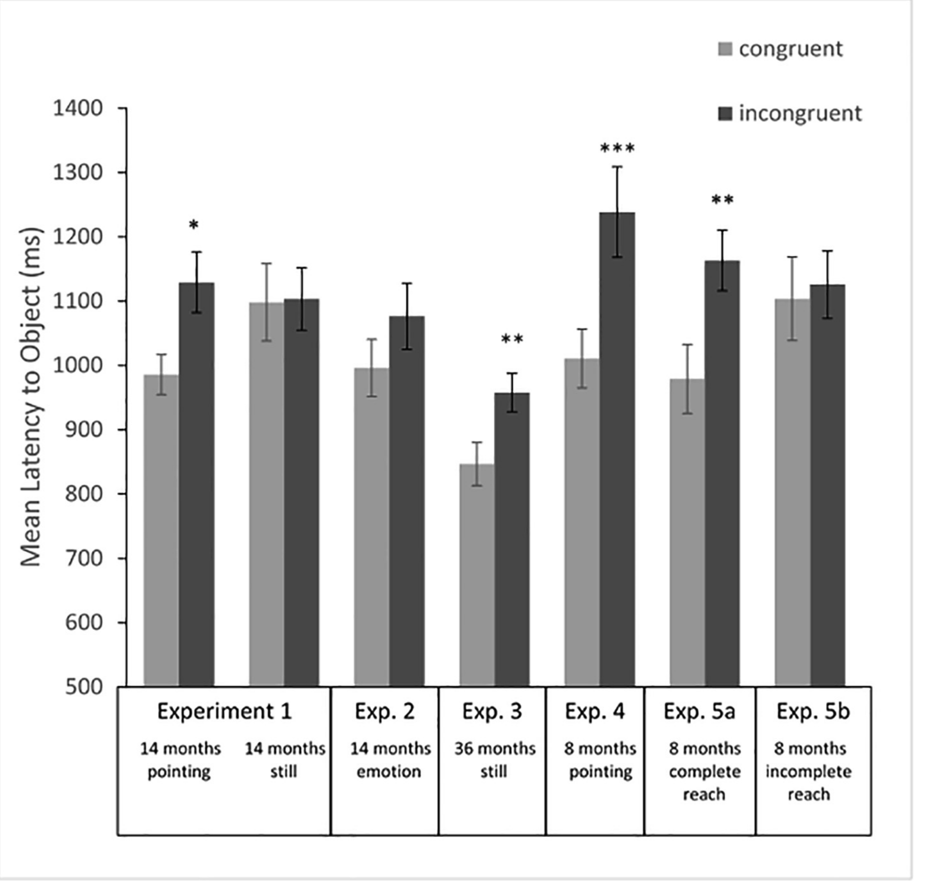

Fig. 2. Mean latency to first fixation of object. Error bars represent ± 1 standard error. *p < .05; **p < .01; ***p < .001.

significantly faster to detect the object in congruent trials than in incongruent trials, $t(22) = 2.37$, $p = .027$ (two-tailed), $d_z = 0.65$. No such effect was found in the still condition, $t(19) = 0.08$, $p = .940$ (two-tailed), $d_z = 0.02$. The pattern of statistical findings held when including the same number of infants from the omnibus ANOVA who had sufficient trial numbers in both conditions.

In the pointing condition, 16 of 23 infants looked faster to the congruent object than to the incongruent object (binomial test, $p = .094$, two-tailed) according to the directed hypothesis ($p = .047$, one-tailed). In the still condition, there was no difference in the number of infants (10 of 20, $p = 1.00$, two-tailed).

Dwell time

The 2 × 2 × 2 (Congruency [congruent or incongruent] × Condition [pointing or still] × Order [pointing first or still first]) mixed-design ANOVA with order as a between-participant factor and all others as within-participant factors showed no main effects [congruency: $F(1, 21) = 0.79$, $p = .383$, $\eta_p^2 = .036$; condition: $F(1, 21) = 0.20$, $p = .658$, $\eta_p^2 = .010$]. The interaction between congruency and condition was not significant, $F(1, 21) = 1.89$, $p = .184$, $\eta_p^2 = .082$. Of all other interactions, only Condition × Order was significant, $F(1, 21) = 6.13$, $p = .022$, $\eta_p^2 = .226$ (pointing first: $M_{\text{pointing}} = 395.64$, $SE = 68.29$, $M_{\text{still}} = 264.70$, $SE = 62.54$; still first: $M_{\text{still}} = 339.34$, $SE = 50.15$, $M_{\text{pointing}} = 248.61$, $SE = 54.76$), suggesting that infants looked less in the second block. The first column of Fig. 3 displays the mean dwell time to the empty location for Experiment 1. To test directly for the predicted effect, a planned

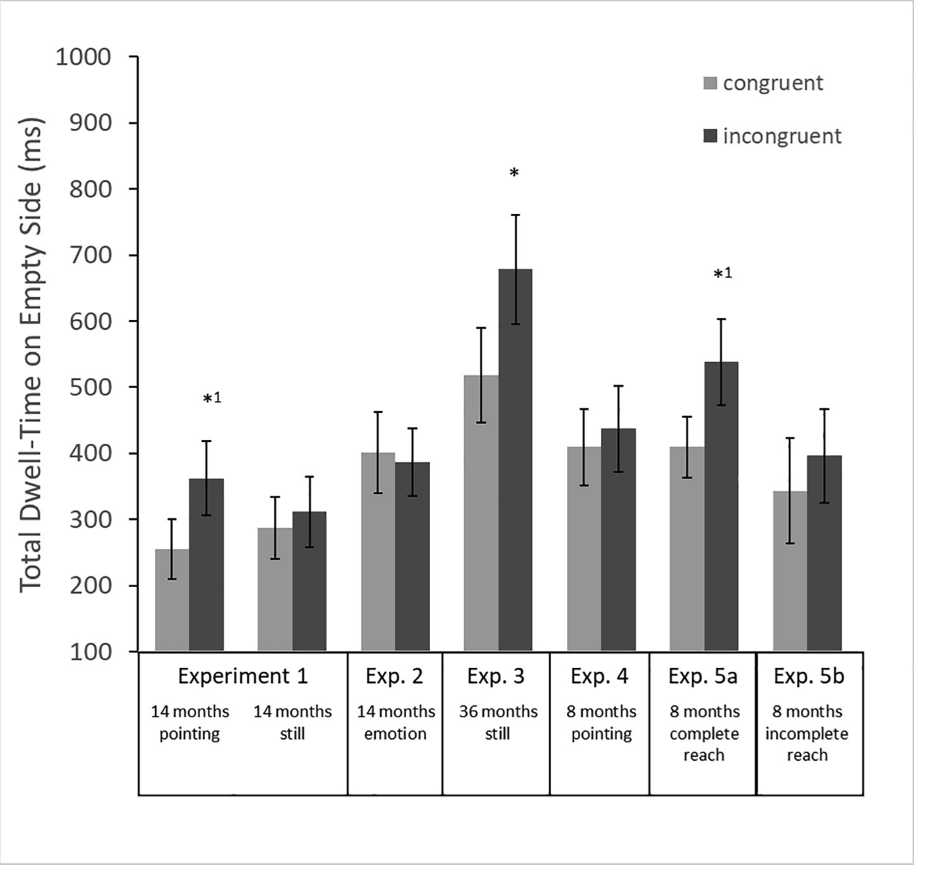

Fig. 3. Mean dwell time on the empty side. Error bars represent ± 1 standard error. $^{*}p < .05$; [1]One-sided significance according to directed hypothesis. Experiments 4, 5a, and 5b display means for the same time frame as in Experiments 1 to 3.

paired t test revealed that infants in the pointing condition looked significantly longer to the empty side of the occluder in incongruent trials than in congruent trials, $t(23) = 1.91$, $p = .070$ (two-tailed; according to the directed hypothesis, $p = .035$, one-tailed), $d_z = 0.42$. No such effect was found in the still condition, $t(23) = 0.38$, $p = .706$ (two-tailed), $d_z = 0.10$. The pattern of statistical findings held when including the same number of infants from the omnibus ANOVA who had sufficient trial numbers in both conditions.

In the pointing condition, 18 of 24 infants looked longer to the empty side in the incongruent videos than in the congruent videos (binomial test. $p = .024$, two-tailed). In the still condition, the difference was not significant (10 of 24 infants, binomial test, $p = .542$, two-tailed).

Discussion

Findings from the communicative condition of Experiment 1 revealed cognitive referential expectations following a communicative cue, which is consistent with the previous literature. In extension, these findings show that direct gaze toward the participant is not necessary, although we note that the condition remained clearly ostensive and communicative through the actress's pointing and voice display. The group-level analysis of the dwell time effect was significant on a one-sided test according to the directed hypothesis. However, the analysis on the individual level clearly confirmed the significance of the dwell time effect. Furthermore, a positive interpretation is in line with previous findings on perspective taking in communicative settings at the current age (see Introduction). It would have been desirable to administer even more trials to increase test power; however, in the current design with two within-participant factors, this would have meant doubling the test time, which did not seem appropriate in the current age range given that infants seemed to become more restless over time. A comparable study with 12-month-olds (Csibra & Volein, 2008) employed only half the number of trials compared with the current study. In our subsequent experiments, accordingly, we administered only one within-participant factor to decrease trial and participant exclusion. In light of 14-month-olds' competence in the communicative condition, the developmental question is then: When do these cognitive referential expectations emerge? We addressed this question in Experiment 4, where we tested 8-month-olds in the communicative pointing condition.

In contrast, in the non-ostensive still condition, when the actress was simply oriented to one side as in onlooker visual perspective-taking tasks, results did not reveal referential expectations. This negative finding adds to the literature, which has suggested that infants first adopt others' visual perspective in communicative settings and initially do not infer the perspective of an onlooker (Moll & Tomasello, 2007). Thus, seeing a person look in a direction appears not to be sufficient for 14-month-olds to infer a hidden referent. Although static cues (von Hofsten et al., 2005) or non-ostensive gaze cues (Collicott, Collins, & Moore, 2009) may elicit simple gaze following, as is also evident from our analysis of the cueing phase before the test outcome, they do not seem to instigate referential expectations at 14 months of age. While automatic visual perspective-taking findings with infants may appear not to be easily compatible with this view, an alternative interpretation of those findings is that an onlooker only enhances infants' perspective during the cue but does not instigate the expectation after the cue that the onlooker has seen something.

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

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

Experiment 2

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

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

Given that infants link an adult's emotion to a referent before 14 months of age, we reasoned that in our paradigm emotional expressions may help 14-month-olds to expect a referent when someone is

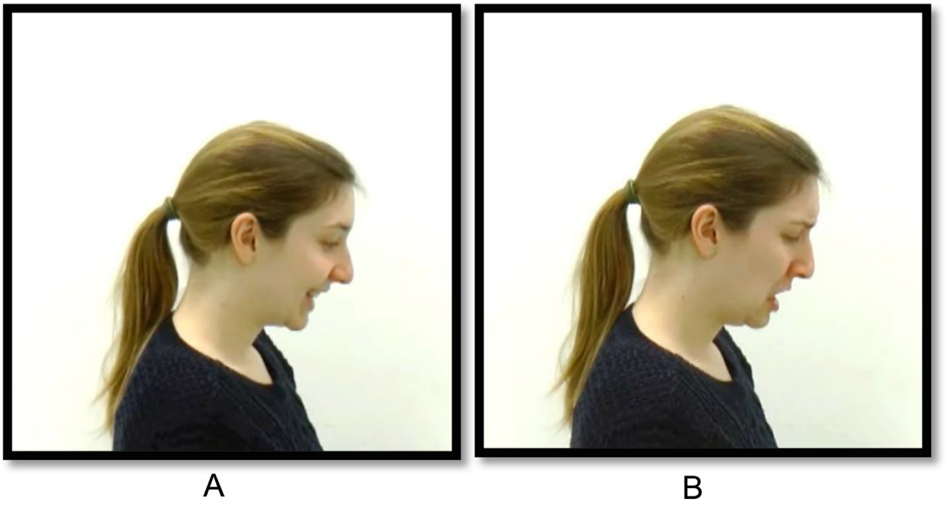

Fig. 4. Facial expressions used in Experiment 2. Shown are the emotion joy (A) and the emotion disgust (B). Setup of the video stimuli remained the same as in Experiment 1.

looking behind a barrier. Note, however, that our primary question was still to test whether still non-ostensive communicative expressions would yield the effects. We used two different emotions for stimulus variety (joy and disgust), but our focus was not to test which emotion works better. Thus, we had no specific expectation that one emotion or the other would yield a stronger object expectation.

Method

Participants

Parents were contacted as described in Experiment 1. The final sample consisted of 26 infants (13 girls). Infants had a mean age of 14 months 15 days and an age range of 14;2 to 14;29. An additional 4 infants were invited but excluded due to fussiness.

Setup and procedure

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

Stimuli and design

The setting and size of the video stimuli were the same as in the still condition of Experiment 1. The person in profile showed a still facial expression while staring at the spot behind the occluder (see Fig. 4). We presented two different emotions, joy and disgust, mainly for variation in the stimulus material. The person presented the facial expression from the first frame onward and held the expression. Instead of a still frame, we presented a video sequence to maintain validity such that the person's lips and body were moving slightly in a natural way. The video format and codec were the same as in Experiment 1. As in Experiment 1, videos were converted to AVI format, and the video codec was WMV1. Due to recommendations from Tobii eye-tracking systems support, we did not convert videos to 15 fps. The fixation cross was extended to 1000 ms to give infants enough time to fixate it. To keep the timing the same, we had the occluder come down slightly faster (200 ms rather than 400 ms). Presentation started with two familiarization trials, and then both emotions were presented blockwise in a within-participant design. Number of congruent and incongruent trials, presentation order, and attention getter were the same as in Experiment 1.

Data reduction and analysis

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

Exclusion criteria were the same as in Experiment 1. For looking to the cue less than 1000 ms, 17 trials of 8 infants were excluded in the disgust emotion and 26 trials of 10 infants were excluded in the joy emotion. For looking at the fixation cross less than 60 ms, 21 trials of 7 infants were excluded in the disgust emotion and 21 trials of 8 infants were excluded in the joy emotion. Based on these criteria, 12 infants watched enough videos of both emotions, making a within-participant analysis of the type of emotion less feasible. Regarding order, 11 infants watched the joy condition first and 7 infants watched the disgust condition first.

Results

Cueing phase

Across both emotions, infants followed to the cued occluder in 18.1% (SD = 19.0) of the trials and to the uncued occluder in 11.2% (SD = 12.0) of the trials, $t(25) = 1.67$, $p = .108$, $d_z = 0.43$. Infants did not disengage from the person in 70.7% (SD = 23.7) of the trials.

Latency

A repeated-measures ANOVA with congruency (congruent or incongruent) as a within-participant factor and emotion (joy or disgust) as a between-participant factor showed no significant main effect of congruency, $F(1, 16) < 0.01$, $p = .985$, $\eta_p^2 < .001$, and no significant interaction for congruency and condition, $F(1, 16) = 0.20$, $p = .661$, $\eta_p^2 = .012$.

The second column of Fig. 3 displays mean latencies for Experiment 2. A planned t test following the directed predictions revealed no significant difference between congruent and incongruent trials, $t(19) = 1.31$, $p = .206$ (two-tailed), $d = 0.37$, with 11 of 20 infants looking faster to the object side in the congruent videos than in the incongruent videos ($p = .824$, two-tailed).

Dwell time

A repeated-measures ANOVA with congruency (congruent or incongruent] as a within-participant factor and condition (joy or disgust) as a between-participant factor showed no significant main effect of congruency, $F(1, 17) = 1.97$, $p = .179$, $\eta_p^2 = .104$, no significant main effect of condition, $F(1, 17) = 0.15$, $p = .702$, $\eta_p^2 = .009$, and no interaction effect of congruency and condition, $F(1, 17) = 1.03$, $p = .324$, $\eta_p^2 = .057$.

The second column of Fig. 3 displays mean dwell time for Experiment 2. A planned t test following the directed predictions revealed no significant difference between congruent and incongruent trials, $t(21) = 0.26$, $p = .794$ (two-tailed), $d_z = 0.06$, with 8 of 22 infants looking longer to the empty side in the incongruent videos than in the congruent videos ($p = .286$, two-tailed).

Discussion

Adding emotional expressions to the still cue did not help to instigate object expectations in 14-month-olds. The emotions were successful in attracting attention to the cue. Overt gaze shifts to the cued versus uncued side were only marginally significant on a one-sided test. Unlike following the explicit communicative pointing cue of Experiment 1, the current emotional still cue was not effective in instigating a referential expectation of an object. Thus, these findings are in support of our interpretation of the negative findings from the still condition of Experiment 1 that at 14 months of age infants do not automatically take another's perspective but likely require communicative cues to do so.

The findings may appear at odds with studies showing that facial expressions enhance gaze following already at 9 months of age. However, in contrast to the current paradigm, gaze following and social referencing paradigms present the face frontal and usually include a variety of social communicative cues such as eye contact and vocal expression, which were absent in our paradigm by intended design. Furthermore, none of the previous studies on emotion cueing tested whether these effects would go beyond perceptual directing and indeed instigate cognitive expectations. Nevertheless, it is important to stress that although the negative findings support our interpretation of Experiment 1, they do not question infants' ability to process and understand facial emotions in general.

Experiment 3

Although Experiment 1 showed that communicative cues instigate object expectations, it could be that the absence of such expectations following the directional still cue were an artifact of the experimental design. Experiment 2 did not provide support for that interpretation, but it could be that the still cue does not work at all in the current paradigm. Therefore, in Experiment 3 we tested whether the still cue would produce an effect in children older than 14 months of age. We tested 36-month-old preschoolers because 3-year-olds have been shown to draw inferences from observations of non-ostensive demonstrations (Schmidt, Rakoczy, & Tomasello, 2011) and are already at the brink of understanding Level 2 perspective taking (Moll & Meltzoff, 2011). If the experiment would produce negative findings, it would rather speak against the validity of our paradigm than reveal a lack of competence. If our paradigm was indeed sensitive to revealing spontaneous, non-ostensive visual perspective taking, we expected that 36-month-olds should show the predicted cueing and dwell time effects.

Method

Participants

Parents were contacted as described in Experiment 1. The final sample consisted of 30 toddlers (14 girls) with a mean age of 36 months 15 days and an age range of 36;2 to 36;27. An addition 7 toddlers were invited but excluded due to fussiness (n = 2), bad calibration (n = 3), or failed calibration (n = 2). According to exclusion criteria (see "Data reduction and analysis" section below), sample size differs across analyses.

Setup and procedure

The setup and procedure remained the same as in Experiment 1. A Tobii eye tracker X120 was sampling with 120 Hz.

Stimulus and design

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

Data reduction and analysis

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

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

Results

Cueing phase

Toddlers followed the cue in 42.2% (SD = 27.8) of the trials and looked to the uncued occluder in 13.6% (SD = 14.8) of the trials, $t(29) = 4.64$, $p < .001$, $d_z = -1.30$. Toddlers did not disengage from the person in 44.2% (SD = 29.1) of the trials.

Latency

The third column of Fig. 2 displays the mean latencies of Experiment 3. A planned t test following the directed hypothesis revealed that toddlers were significantly faster to detect the object in congruent trials than in incongruent trials, $t(25) = 3.01$, $p = .006$ (two-tailed), $d_z = 0.69$, with 19 of 26 toddlers looking faster to the object in the congruent videos than in the incongruent videos (binomial test, $p = .030$, two-tailed).

Dwell time

The third column of Fig. 3 displays the mean dwell time of Experiment 3. A planned *t* test following the directed hypothesis revealed that toddlers looked significantly longer to the empty side in incongruent trials than in congruent trials, $t(26) = 2.72$, $p = .011$ (two-tailed), $d_z = 0.39$, with 20 of 27 toddlers looking longer to the empty side in the incongruent videos than in the congruent videos (binomial test, $p = .020$, two-tailed).

Discussion

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

Experiment 4

Experiment 4 investigated how early in ontogeny communicative cues may instigate visual perspective taking in the current paradigm. Classic views have suggested that before 9 months of age, infants' joint engagement skills are not fully triadic referential and likely do not entail following into others' perspectives (Bakeman & Adamson, 1984; Tomasello, 2019). Electroencephalography (EEG) studies on covert attention, however, have found that 8-month-olds show sensitivity to incongruences between a target location and a subsequent point direction (Gredebäck & Melinder, 2010). But further eye-tracking studies have suggested that cueing a target requires communicative cues and first emerges after 8 months of age (i.e., by 12 months; Daum et al., 2013). Those cueing paradigms, however, measure rather reflexive than reflective following with short SOAs. Two violation-of-expectation studies tested more directly cognitive object expectations following referential gaze and point cues but yielded both positive findings (Csibra & Volein, 2008) and negative findings (Pätzold & Liszkowski, 2019) for communication-induced object expectations at 8 months of age. In the current experiment, therefore, we tested 8-month-olds with the communicative pointing cue condition from Experiment 1. Because it was less clear whether processing time might take longer overall for infants, we extended the time window for the dwell time analysis by 1 s.

Method

Participants

Parents were contacted as described in Experiment 1. The final sample consisted of 22 infants (12 girls) with a mean age of 8 months 14 days and an age range of 8;2 to 8;29. An addition 6 infants were invited but excluded because calibration was not possible ($n = 1$) or not accurate enough ($n = 5$). According to exclusion criteria (see "Data reduction and analysis" section below), sample size differs in each analysis.

Setup and procedure

The setup and procedure remained the same as in Experiment 1. Tobii eye tracker X120 was sampling with 60 Hz.

Stimuli and design

The setting of the video was the same as in the pointing condition of Experiment 1. Following Tobii Studio advice, rendering was set to 25 fps (instead of 15 fps; see Experiment 2) so that every time the sequence happened 120 ms earlier than in Experiment 1. To account for possibly longer processing time overall, the outcome phase was extended by 1 s to 4800 ms.

Data reduction and analysis

The same dependent variables and AOIs were used as in Experiment 1. According to the sampling rate of 60 Hz, noise reduction was disabled as in Experiment 1.

Exclusion criteria were the same as in Experiment 1. For watching the cue phase less than 1000 ms, 1 trial of 1 infant was excluded. For fixating the fixation cross for less than 60 ms, 22 trials of 6 infants were excluded.

Results

Cueing phase

Infants followed the cue in 11.1% (SD = 21.1) of the trials and looked to the uncued occluder in 3.3% (SD = 9.4) of the trials, $t(22) = 1.72$, $p = .100$, $d_z = 0.47$. Infants did not disengage from the person in 85.6% (SD = 24.2) of the trials.

Latency

The fourth column of Fig. 2 displays the mean latency of Experiment 4. We winsorized 5% of the variable latency in congruent trials. Infants were significantly faster to detect the object in congruent trials than in incongruent trials, $t(17) = 4.00$, $p = .001$ (two-tailed), $d_z = 0.85$, with 14 of 18 infants looking faster to the congruent object than to the incongruent object (binomial test, $p = .012$, two-tailed).

Dwell time

The fourth column of Fig. 3 displays the mean dwell time of Experiment 4. We winsorized 5% of the variable dwell time in congruent and incongruent trials. There was no difference in the dwell time on the empty location between the congruent and incongruent trials, $t(19) = 0.58$, $p = .571$ (two-tailed), $d_z = 0.10$, with 10 of 20 infants looking longer to the empty side in the incongruent videos than in the congruent videos, which is not significantly above chance ($p = 1$, two-tailed). The pattern of results remained the same when analyzing the time window with the extended 1000 ms during the test phase, $t(19) = 0.31$, $p = .758$ (two-tailed), $d_z = 0.06$ ($M_{\text{congruent}} = 447.71$, $SD = 291.70$; $M_{\text{incongruent}} = 465.00$, $SD = 323.48$).

Discussion

The 8-month-olds followed the communicative pointing cue to the indicated location, and they displayed a latency effect for the location of the target object, as revealed by significantly faster latencies for that location. However, unlike for the 14-month-olds of Experiment 1, in Experiment 4 the dwell time analysis did not reveal reliable violations of object expectations, neither for the extended 5-s time window nor for the shorter 4-s time window and neither on a one-tailed test nor in an individual-level analysis. The lack of a violation-of-expectation finding is in line with the interpretation that communicative cues do not yet instigate referential expectations in 8-month-olds (Pätzold & Liszkowski, 2019; Rüther & Liszkowski, 2020). However, they contrast with a previous finding by Csibra and Volein (2008). The latter study did not involve a re-fixation cross before the target object was revealed, although the authors did present a control analysis that excluded looks that had remained on the cued side before the object was revealed. It may be that in their study the older 12-month age group mainly drove the significance of the study's effect given that the analyses were not run on the 8-month-old sample independently.

The latency result of Experiment 4 contrasts with an earlier finding that only 12-month-olds show a latency effect with a pointing hand including communitive speech (Daum et al., 2013). However, in our paradigm, a person was fully visible and moved, providing a more naturalistic cue than a static hand. Naturalistic cues and movement (Rohlfing et al., 2012) seem to be important when directing infants' attention at 8 months of age.

It is possible that 8-month-olds are still less familiar with the pointing cue, especially because they do not themselves point at that age. Infants do, however, reach for objects at that age, perhaps even to get others to help them obtain objects (Ramenzoni & Liszkowski, 2016), and they begin to understand reaching as object directed (Brandone, Horwitz, Aslin, & Wellman, 2014; Hamlin, Hallinan, & Woodward, 2008). In a final experiment, therefore, we substituted the pointing cue with a reaching cue to test whether infants would expect it to be related to occluded objects, perhaps as an indication of an early understanding of the reacher's perspective.

Experiment 5

In Experiment 5, we again used the same general paradigm and hypotheses to test whether 8-month-olds expect an occluded object when someone is reaching behind a barrier. In Experiment 5a, infants watched a dynamic complete grasp with the extended arm ending behind the occluder. In Experiment 5b, infants watched a dynamic reach matched in kinematics and distance to the pointing cue, with the reaching hand stopping where the pointing hand of Experiment 4 had stopped, to equate for the surface features of the cue.

Infants reach, and understand others' reaching actions, as goal directed from around the second half of the first year of life (Applin & Kibbe, 2019; Luo & Johnson, 2009). By 10 months of age, infants visually anticipate the goal of an unfulfilled reaching action (Brandone et al., 2014). Sensorimotor alpha attenuation in EEG suggests that 9-month-olds understand reaching actions behind an occluder as object directed (Southgate, Johnson, El Karoui, & Csibra, 2010). Given the findings, in Experiment 5a we expected that if our paradigm and measures work in general with 8-month-olds, it should yield latency and dwell time effects indicative of infants' understanding of reaching actions as object directed. It was less clear whether this would also be the case for the abbreviated reach of Experiment 5b, which was matched to the pointing cue. Findings from Brandone et al. (2014) would suggest that object-directed anticipations of dynamic reaches (i.e., before the reach is fully executed) would not occur before 10 months of age. Perceptual cueing paradigms employing EEG signatures have found latency effects with shorter SOA to static cues at earlier ages (Bakker, Sommerville, & Gredebäck, 2016; Daum & Gredebäck, 2011).

Method

Participants

Parents were contacted as described in Experiment 1. The final sample in Experiment 5a consisted of 25 infants (14 girls) with a mean age of 8 months 14 days and an age range of 8;2 to 8;29. An addition 8 infants were invited but excluded because calibration was not possible ($n = 1$), calibration was not accurate enough ($n = 5$), or the eye tracker lost eye gaze due to high fussiness and movement ($n = 2$).

In Experiment 5b, the final sample consisted of 27 infants (12 girls) with a mean age of 8 months 14 days and an age range of 8;2 to 8;29. An additional 7 infants were invited but excluded because calibration was not accurate enough ($n = 5$) or the eye tracker lost eye gaze due to high fussiness and movement ($n = 2$).

According to exclusion criteria (see "Data reduction and analysis" section below), sample size differs in each analysis.

Setup and procedure

The setup and procedure remained the same as in Experiment 1. The sampling rate was adjusted to 120 Hz.

Stimuli and design

The setting and size of the video stimuli were the same as in the pointing condition of Experiment 1 except that during the cueing phase the person was reaching instead of pointing. The person started reaching from her lap. In Experiment 5a, she grasped behind one of the two occluders until the hand was not visible anymore (see Fig. 5). In Experiment 5b, the reach stopped at the position of the pointing cue from Experiment 1 (see Fig. 5). The videos had a frame rate of 25 fps. Timing was the same as in Experiment 4. Again, two familiarization trials and eight test trials, with four congruent and four incongruent trials, were presented with small attention getters between each trial. The order of cueing side and congruency was counterbalanced as in Experiment 1.

Data reduction and analysis

The dependent variables and AOIs were the same as in Experiment 1. Again, we enabled a Moving Median noise reduction (three-sample window size in Tobii IVT filter) to adjust the higher sampling rate (120 Hz) to the smaller one from Experiments 1 and 4 (60 Hz).

Exclusion criteria were the same as in Experiment 1. In Experiment 5a, for watching the cue phase less than 1000 ms, 21 trials of 8 infants were excluded. For looking at the fixation cross less than 60 ms, 28 trials of 14 infants were excluded. In Experiment 5b, for watching the cue phase less than 1000 ms, 15 trials of 11 infants were excluded. For looking at the fixation cross less than 60 ms, 40 trials of 15 infants were excluded.

Results: Experiment 5a

Cueing phase

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

Latency

The fifth column of Fig. 2 displays the mean latency of Experiment 5a. Infants were significantly faster to detect the object in congruent trials than in incongruent trials, $t(18) = 3.60$, $p = .002$ (two-tailed), $d_z = 0.83$, with 16 of 19 infants looking faster to the object side in the congruent videos than in the incongruent videos (binomial test, $p = .004$, two-tailed).

Dwell time

The fifth column of Fig. 3 displays the mean dwell time of Experiment 5a. Infants looked slightly longer to the empty side in incongruent trials than in congruent trials, $t(19) = 2.08$, $p = .051$ (two-tailed; according to the directed hypothesis, $p = .026$, one-tailed, $d_z = 0.50$). When including the last 1000 ms in the extended test window, the comparison was not significant, $t(19) = 1.18$, $p = .254$ (two-tailed), $d_z = 0.31$ ($M_{\text{congruent}} = 471.01$, $SD = 266.05$; $M_{\text{incongruent}} = 560.90$, $SD = 314.02$). The number of infants who looked longer to the empty side in the incongruent videos than in the congruent videos was 13 of 20 infants ($p = .262$, two-tailed) for both time windows.

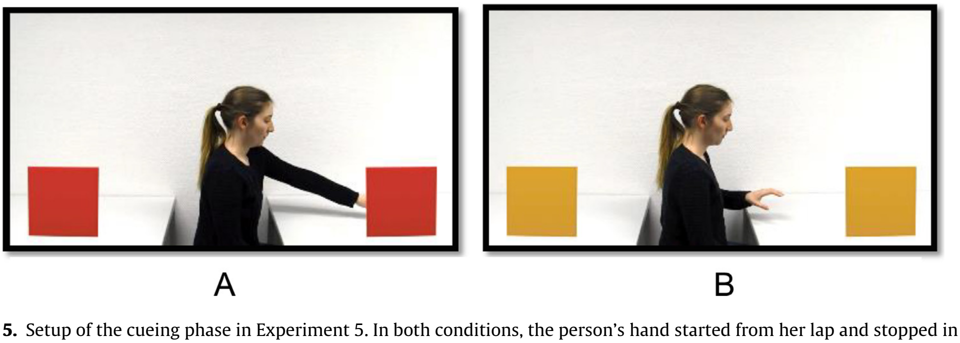

Fig. 5. Setup of the cueing phase in Experiment 5. In both conditions, the person's hand started from her lap and stopped in the position presented in the figure. Shown are Experiment 5a (A) and Experiment 5b (B).

Results: Experiment 5b

Cueing phase

Infants followed the cue to the cued occluder in 12.7% (SD = 17.3) of the trials and looked to the uncued occluder in 9.8% (SD = 21.8) of the trials, revealing no significant difference, $t(26)$ = 0.50, p = .620, d_z = 0.15. Infants did not disengage from the person in 77.5% (SD = 25.1) of the trials.

Latency

The last column of Fig. 2 displays mean latency of Experiment 5b. A dependent t test revealed that infants were not significantly faster to detect the object in congruent trials than in incongruent trials, $t(17)$ = 0.28, p = .784 (two-tailed), d_z = 0.09, with 10 of 18 infants looking faster to the object side in the congruent videos than in the incongruent videos (binomial test, p = .816, two-tailed).

Dwell time

The last column of Fig. 3 displays mean latency of Experiment 5b. A dependent t test revealed that infants did not look longer to the empty side in incongruent trials than in congruent trials, $t(21)$ = 0.72, p = .480 (two-tailed), d_z = 0.15. The number of infants who looked longer to the empty side in incongruent trials than in congruent trials did not differ (11 of 22 infants; binomial test, p = 1, two-tailed). The nonsignificant pattern remained the same when analyzing dwell time with the extended 1000 ms, $t(21)$ = 0.73, p = .474 (two-tailed), d_z = 0.17 ($M_{\text{congruent}}$ = 367.42, SD = 421.44; $M_{\text{incongruent}}$ = 437.28, SD = 407.32).

Discussion

Individual reaching for an occluded object instigated object expectations at 8 months of age, as revealed by latency and dwell time measures, but only if infants watched a full grasp (Experiment 5a), not when it was an abbreviated grasp matched to the pointing cue (Experiment 5b). The dwell time effect of Experiment 5a was not strong, but it appears to be meaningful considering the directed hypothesis of the effect and previous findings, which have revealed infants' object-directed understanding of actions (e.g., Applin & Kibbe, 2019; Luo & Johnson, 2009).

Attention to the cue in Experiment 5a was conflated with attention to the location because the cue did not stop before the location. As such, it may have exogenously cued the location. Indeed, when the reaching cue in Experiment 5b was matched to the distance and kinematics of the pointing cue, it did not cue the congruent side during the cueing phase and it did not enhance the latency to detect the target at that side after the cueing phase. Thus, although Experiment 5a may suggest that infants do seem to understand something about others' directedness of actions, Experiment 5b suggests that this understanding does not lend itself to understanding others' actions as communicative. The lack of a cueing effect to a dynamic reach at 8 months of age is supported by previous findings (Brandone et al., 2014). It may appear to contrast with reported cueing effects by Bakker et al. (2016); however, that latter study tested a short SOA in an EEG setting and employed a static cue. In the current dynamic setting, the interruption of the grasping act halfway through may have appeared to be more artificial than the pointing act (which typically does not result in touching). On the other hand, this difference in distance is a crucial distinction between ostensive communicative and individual goal-directed actions.

General discussion

The current study investigated the age and conditions under which infants and toddlers begin to engage in visual perspective taking of the kind that enables representing that others see something in their line of sight. This basic form of Level 1 visual perspective taking is conceptually simpler than more complex forms of representing what exactly someone else represents (Level 2; Flavell et al., 1981) or (falsely) believes (Tomasello, 2018). But it is conceptually more complex than simply following others' line of sight perceptually or orienting covertly to a cued location (Level 0).

Paradigmatically, one can distinguish Level 0 from Level 1 with occlusion paradigms when a participant cannot see an object by just following another's line of sight but can infer the presence of the object if the participant envisions the other's perspective. Previous research has shown that this ability is in place around 12 months of age when the cue is ostensive communicative (Behne et al., 2012; Csibra & Volein, 2008; Moll & Tomasello, 2004; Pätzold & Liszkowski, 2019). Other paradigms, primarily from adult visual perspective-taking tasks, have revealed that non-ostensive directed cues, such as a person or face in profile, are sufficient to cue a perspective seemingly automatically even during infancy (Kampis et al., 2015; Kovács et al., 2010; Samson et al., 2010; Teufel et al., 2010). The current study investigated whether and when during development infants and toddlers take another's perspective following non-ostensive still displays of a directed perspective or following ostensive communicative cues.

The main Experiment 1 established that, as expected from the literature, communicative pointing indeed induced cognitive representations of a hidden object. In addition, it revealed the absence of this effect for a still directed perspective of a person in profile. The two subsequent experiments were conducted to follow up on the absence of this effect. They established, on the one hand, that adding still emotional expressions did not help 14-month-olds to take the perspective and, on the other hand, that 3-year-olds did successfully take the perspective of a non-ostensive still onlooker. Together, the pattern of findings of the first three experiments thus provides a firm basis to suggest that seemingly automatic visual perspective taking following non-ostensive still directive perspectives is a developing skill rather than a starting point of some form of System 1 processes (e.g., Apperly & Butterfill, 2009). This interpretation is supported by previous findings showing that visual perspective taking first emerges in communicative situations through joint engagement (Moll & Tomasello, 2007).

The second set of Experiments 4, 5a, and 5b followed up on the positive effect of communication-induced perspective taking in Experiment 1 and asked how early in development such an effect may be present. Findings revealed that by the same standards as in Experiment 1, which had revealed positive findings at 14 months of age, there was no sound evidence for object expectations following communicative cues at 8 months of age. Experiments 5a and 5b again followed up on the absence of the effect and revealed that grasping, a more familiar action than pointing at that age, instigated cognitive expectations of a hidden object (Experiment 5a), but when the action was matched in surface structure to the pointing cue (i.e., an abbreviated grasp), the effects were again absent. The latter two experiments confirmed the general applicability of our current paradigm. Although not of immediate centrality for current arguments about perspective taking, they are relevant in their own right for future investigations of infants' understanding of non-ostensive individual actions.

The pattern of findings from the first and second sets of experiments allows for a developmental hypothesis by which social interactional experiences continuously shape social cognitive development from early on (see Liszkowski, 2018). The social cognitive skill of Level 1 perspective taking emerges around 12–14 months of age in ostensive communicative settings. Across subsequent development, ostensive communicative exchanges are then a basis to develop later automatized forms of visual perspective taking outside of communicative exchange, as our findings with 3-year-olds suggest. However, Level 1 perspective taking around 12–14 months of age is also a developmental outcome of the first year of life given that these skills are not fully in place around 8 months of age. We suggest that simpler forms of social interaction, which invoke simpler cognitive processes (e.g., Level 0), scale up to cognitive expectations of Level 1 perspective taking through social interactional experiences.

The developmental pattern of the current findings that (a) communication-induced Level 1 perspective taking is not fully developed at 8 months of age but is present at 14 months and that (b) spontaneous non-ostensive perspective taking is absent at 14 months of age but is present at 36 months suggests a continuous construction of social understanding through social interaction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by a grant from the German Federal Ministry of Education and Research (BMBF) (01DL14007). We thank all families who participated in these experiments. We thank Nicola Ballhausen and Maike Fuchs for their help in preparing the stimuli and data collection. We thank Wiebke Pätzold for valuable feedback during the early stages of the project.

References

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

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

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother–infant and peer–infant interaction. *Child Development, 55*, 1278–1289.

Bakker, M., Sommerville, J. A., & Gredebäck, G. (2016). Enhanced neural processing of goal-directed actions after active training in 4-month-old infants. *Journal of Cognitive Neuroscience, 28*, 472–482.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science, 8*, 492–499.

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology, 30*(Pt 3), 359–375.

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

Brandone, A. C., Horwitz, S. R., Aslin, R. N., & Wellman, H. M. (2014). Infants' goal anticipation during failed and successful reaching actions. *Developmental Science, 17*, 23–34.

Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology, 38*, 958–966.

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

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

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

Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance, 43*, 454–465.

Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 358*, 447–458.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148–153.

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

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.

Daum, M. M., Ulber, J., & Gredebäck, G. (2013). The development of pointing perception in infancy: Effects of communicative signals on covert shifts of attention. *Developmental Psychology, 49*, 1898–1908.

De Groote, I., Roeyers, H., & Striano, T. (2007). Gaze following as a function of affective expression in 3-, 6- and 9-month-old infants. *Infant Behavior and Development, 30*, 492–498.

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology, 85*, 199–212.

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

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

Flom, R., & Johnson, S. (2011). The effects of adults' affective expression and direction of visual gaze on 12-month-olds' visual preferences for an object following a 5-minute, 1-day, or 1-month delay. *British Journal of Developmental Psychology, 29*, 64–85.

Flom, R., & Pick, A. D. (2005). Experimenter affective expression and gaze following in 7-month-olds. *Infancy, 7*, 207–218.

Gredebäck, G., & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition, 114*, 197–206.

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

Hemmerich, W. (2018). StatistikGuru: Winsorizing-Rechner. Retrieved from https://statistikguru.de/rechner/winsorizing-rechner.html.

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

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science, 9*, 131–143.

Kampis, D., Parise, E., Csibra, G., & Kovács, Á. M. (2015). Neural signatures for sustaining object representations attributed to others in preverbal human infants. *Proceedings of the Royal Society B: Biological Sciences, 282*, 20151683.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830–1834.

Krehm, M., Onishi, K. H., & Vouloumanos, A. (2014). I see your point: Infants under 12 months understand that pointing is communicative. *Journal of Cognition and Development, 15*, 527–538.
Liszkowski, U. (2018). Emergence of shared reference and shared minds in infancy. *Current Opinion in Psychology, 23*, 26–29.
Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science, 12*, 142–149.
Moll, H., & Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Development, 82*, 661–673.
Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science, 7*, F1–F9.
Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology, 43*, 309–317.
Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development, 72*, 718–735.
Pätzold, W., & Liszkowski, U. (2019). Pupillometry reveals communication-induced object expectations in 12- but not 8-month-old infants. *Developmental Science, 22* e12832.
Ramenzoni, V. C., & Liszkowski, U. (2016). The social reach: 8-Month-olds reach for unobtainable objects in the presence of another person. *Psychological Science, 27*, 1278–1285.
Reifman, A., & Keyton, K. (2010). Winsorize. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 1636–1638). Thousand Oaks, CA: Sage.
Rohlfing, K. J., Longo, M. R., & Bertenthal, B. I. (2012). Dynamic pointing triggers shifts of visual attention in young infants. *Developmental Science, 15*, 426–435.
Rüther, J., & Liszkowski, U. (2020). Ontogenetic emergence of cognitive reference comprehension. *Cognitive Science, 44* e12869.
Salkind, N. J. (2006). *Encyclopedia of measurement and statistics*. Thousand Oaks, CA: Sage.
Samson, D., Apperly, I., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1255–1266.
Schmidt, M. F., Rakoczy, H., & Tomasello, M. (2011). Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science, 14*, 530–539.
Southgate, V., Johnson, M. H., El Karoui, I., & Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychological Science, 21*, 355–359.
Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition, 130*, 1–10.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.
Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception, & Psychophysics, 72*, 695–705.
Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences of the United States of America, 115*, 8491–8498.
Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Cambridge, MA: Belknap Press.
von Hofsten, C., Dahlström, E., & Fredriksson, Y. (2005). 12-Month-old infants' perception of attention direction in static video images. *Infancy, 8*, 217–231.