The temporal dynamics of infants' joint attention: Effects of others' gaze cues and manual actions

Ty W. Boyer[a], Samuel M. Harding[b], Bennett I. Bertenthal[b],*

[a] *Georgia Southern University, Department of Psychology, United States of America*
[b] *Indiana University, Department of Psychological and Brain Sciences and Program in Cognitive Science, United States of America*

A B S T R A C T

Infants' development of joint attention shows significant advances between 9 and 12 months of age, but we still need to learn much more about how infants coordinate their attention with others during this process. The objective of this study was to use eye tracking to systematically investigate how 8- and 12-month-old infants as well as adults dynamically select their focus of attention while observing a social partner demonstrate infant-directed actions. Participants were presented with 16 videos of actors performing simple infant-directed actions from a first-person perspective. Looking times to faces as well as hands-and-objects were calculated for participants at each age, and developmental differences were observed, although all three groups looked more at hands-and-objects than at faces. In order to assess whether visual attention was coordinated with the actors' behaviors, we compared participants looking at faces and objects in response to gaze direction as well as infant-directed actions vs. object-directed actions. By presenting video stimuli that involved continuously changing actions, we were able to document that the likelihood of joint attention changes in both real and developmental time. Overall, adults and 12-month-old infants' visual attention was modulated by gaze cues as well as actions, whereas this was only partially true for 8-month-old infants. Our results reveal that joint attention is not a monolithic process nor does it develop all at once.

1. Introduction

Social attention to others' eyes, faces, and actions is foundational to how we communicate, learn about the social and physical world, regulate emotions, and develop attachments with others. Beginning at birth, infants attend preferentially to faces, and are most sensitive to the presence of eyes in a face (Acerra, Burnod, & de Schonen, 2002; Batki, Baron-Cohen, Wheelright, Connellan, & Ahluwalia, 2000; Johnson & Morton, 1991). In addition, newborn infants prefer to orient to faces displaying direct gaze (Farroni, Csibra, Simion, & Johnson, 2002), and show a rudimentary form of gaze following (Farroni, Massaccesi, Pividori, & Johnson, 2004). Some evidence suggests that newborns recognize their mother's face (e.g., Bushnell, 2001), and these recognition abilities continue to develop over the first few months (Nelson, 2001). Beginning around 10 weeks of age infants fixate more consistently on the internal features of a face than on the external features and contours, especially when the face is speaking (Haith, Bergman, & Moore, 1977; Hunnius & Geuze, 2004). By three months, infants begin to differentiate faces based on the social categories of gender and race (Kelly et al., 2005; Quinn, Yahr, Kuhn, Slater, &

Pascalis, 2002).

This improvement in face perception continues to develop during the next few months, because infants engage often in dyadic interactions with their caregivers ensuring that faces are a prominent part of their visual experience (Lock & Zukow-Goldring, 2010; Lockman, 2000). Once they can sit without support and coordinate their reaches toward objects, infants' reliance on interactions with other people for stimulation begins to decline. By around six months of age, infants are much more likely to divide their attention between exploring objects with their eyes and hands and interacting with social partners (Lock & Zukow-Goldring, 2010). For the next few months they typically distribute their attention to either objects or social partners, but they still must learn to share their attention to a common referent with someone else. It is not until 9 to 12 months of age (at least according to most social-cognitive theorists) that infants attribute intentional states to social partners enabling them to engage in triadic interactions (Tomasello, 2008; Woodward, 2009), such as participating with others in joint attention to objects and establishing common ground (Bakeman & Adamson, 1984; Carpenter, Nagell, & Tomasello, 1998), pointing to objects communicatively (Carpenter et al., 1998), and expecting social

* Corresponding author at: Department of Psychological & Brain Sciences, 1101 E. Tenth Street, Bloomington, IN 47405, United States of America.
  E-mail address: bbertent@indiana.edu (B.I. Bertenthal).

partners to express interest in shared referents (Liszkowski, Carpenter, & Tomasello, 2007).

In order for infants to develop these skills they must first learn to coordinate their attention to their social partner with their attention to objects (Bertenthal, Boyer, & Harding, 2014). Although it is well established that this developmental transition occurs, little is known about how a preference for faces gives way to a more distributed view of the social world that includes not only faces, but bodies and actions, as well as objects. In general, attention is the front-end of encoding and interpreting all stimulus information encountered in the environment, and thus it is essential for not only learning to recognize and discriminate faces, but others' actions as well. How do infants decide where to look from moment-to-moment when confronted with not only a dyadic partner but also an assortment of objects, other people, and events in their optic arrays? Early on, infants' orienting to stimuli in the environment is primarily under exogenous stimulus-driven control, but over time they begin to also develop endogenous control over their attention (Johnson, 2011; Mundy & Jarrold, 2010). As such, they begin modulating their attention in response to the actions of their social partner as well as the context (Bertenthal & Boyer, 2015). Indeed, this is exactly what is necessary for infants to follow the gaze direction of a social partner during shared attention. If infants could not modulate their attention, then they would simply continue to be guided by their bias for faces, but the development of joint attention suggests otherwise.

Although there has been considerable research investigating the social cognitive prerequisites for joint attention, such as shared intentions or common ground (Tomasello, 2008), much less is known about how and when infants begin to dynamically coordinate their social attention among faces, actions, and objects. One reason for the sparseness of relevant findings is that most studies obviate the need for infants choosing between different stimulus cues. Infants are typically presented with a specific sequence of events, such as an actor eliciting an infant's attention, and then looking or pointing in a specific direction, followed by an object appearing either in that direction or the opposite direction; infants merely have to attend to the stimuli in the order they appear and not choose when and what to look at (e.g., Bertenthal et al., 2014; Gredebäck, Fikke, & Melinder, 2010; Senju & Csibra, 2008). In more naturalistic situations, such as an infant interacting with a caregiver in a cluttered room among a set of objects over a more extended period of time, the caregiver might alternate between gazing at the child and the objects and jointly playing with those objects or showing them to the child. The question then becomes, how much are infants' looking behaviors guided by attention to the face or by attention to the manual actions of the caregiver, the orientation of her face, her body posture, or changes in her object-directed actions? This is a critical question because infants' systematic selection of social information in triadic interactions may not only precede but catalyze their appreciation of shared intentions or common ground. In other words, these preferences establish new opportunities for social interaction and social learning, which might very well contribute to their social-cognitive development. It is for this reason that we sought to study how infants distribute their attention during social interactions.

Recent advances in infants' eye tracking research offer important opportunities for systematically investigating how infants allocate their attention to social and non-social stimuli. Most studies, however, still rely on presenting highly scripted and repetitive actions to infants in experimental paradigms involving a live, digital image or movie of a social partner looking or reaching toward an object following an ostensive cue, such as eye contact with the viewer (e.g., Daum, Ulber, & Gredebäck, 2013; Senju & Csibra, 2008; Woodward, 1998). During the past decade, Frank and colleagues (Frank, Vul, & Johnson, 2009; Frank, Vul, & Saxe, 2012) made some important progress in studying infants' and toddlers' social attention to more naturalistic visual scenes. For instance, Frank et al. (2012) measured the visual fixations of infants and toddlers between 3 and 30 months of age while viewing short videos of objects, faces, children playing with toys, and complex social scenes involving more than one person. The results revealed that the youngest infants looked primarily at faces, and eyes in particular, but older infants and toddlers distributed their gaze more flexibly and looked more at the mouth and also significantly more at the hands, especially when the hands were engaged in actions on objects. One important question that could not be addressed by these studies is whether children's attention is directed differently to people observed from a first-person as opposed to a third-person perspective.

A more recent study by Elsabbagh et al. (2014) also studied infants' relative distribution of fixations to the eyes and mouth when viewing a social partner (observed from a first-person perspective) with eyes, mouth or hands moving or expressing multiple communicative signals (e.g., "peek-a-boo"). Consistent with previous studies, infants between 7 and 15 months of age looked at the eyes more than the mouth, but this difference was contextually modulated, such that when only the mouth moved infants looked more at the mouth than when only the eyes moved. Taken together, these last few studies suggest that by sometime during the latter half of the first year infants' social attention is controlled by both stimulus-driven factors, such as sensory (e.g., contrast, color, orientation, and motion) and social salience (e.g., faces), as well as more endogenous or goal-directed factors that can exert control of looking behavior.

The objective of the current study was to move beyond these generalizations in order to better understand how infants dynamically select their focus of attention while observing people who appear to be interacting with them. This dynamic selection of where to look is a prerequisite for joint attention. During direct gaze there is an opportunity for eye contact and communication with the social partner, whereas during averted gaze there is an opportunity for joint attention toward another person or object (Farroni, Mansfield, Lai, & Johnson, 2003; Senju & Csibra, 2008; Senju, Csibra, & Johnson, 2008). Previous eye tracking studies were restricted to reporting where infants directed their attention based on first-order stimulus information, such as faces or objects in the scene (e.g., Jones & Klin, 2013). As such, these studies ignored how contextual and social cues, such as gaze direction or actions, might orient infants to look toward a specific location. These second-order cues result in a more complex and probabilistic process, because the observer decides where to look not only as a function of the region of interest (e.g., faces, objects) but also in response to other actions as well as knowledge of the preceding events. For example, the likelihood of looking at someone's face during a conversation is much higher if that individual's gaze is oriented directly toward you as opposed to looking toward another object (Kleinke, 1986; Senju & Hasegawa, 2005). If, however, the social partner is also waving her hands or manipulating an object while looking toward you, the likelihood of looking at the face and establishing eye contact with the social partner decreases. In typical social interactions, the cues for where to look will often compete and this is especially true for young infants outside of the lab. This is the reason that we sought to study how infants guide their visual attention during more naturalistic social situations.

We measured infants' eye gaze to dynamic social scenes. Unlike the studies conducted by Frank and colleagues, the stimuli were not movies of people or cartoon characters shown from a third-person perspective such that infants were simply watching a movie. Instead, our stimuli were created to show different actors socially engaged with an observer viewed from a first-person perspective. Although the stimuli were videos, they were designed to simulate naturalistic situations that could occur between a social partner and an infant. As such, each of 16 videos presented one of five female actors talking and demonstrating a sequence of simple actions, such as putting a shirt on a stuffed animal. Since our primary goal was to conduct a detailed analysis of the changing focus of attention during joint attention, it was especially important to include both people and objects. Contrary to conventional wisdom, a few recent studies suggest that infants do not always look at the social partner's eyes or face during joint attention; instead they

focus primarily on sharing attention to the same object-directed actions (Deák, Krasno, Jasso, & Triesch, 2018; Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; Franchak, Kretch, Soska, & Adolph, 2011; Yu & Smith, 2013). Thus, it was especially important for us to include not only people and their gestures, but object-directed actions as well.

Three age groups were tested: 8- and 12-month-old infants, and adults. The two infant groups were selected to straddle the age at which joint attention develops and adults were included to enable a comparison of the infants' performance with more mature visual scanning behavior. Our goal was to assess the degree to which developmental changes in shifting attention to faces vs. objects was a function of the direction of head and eye gaze as well as infant-directed and object-directed actions.

We hypothesized that 12-month-old infants and adults would systematically sustain or shift attention as a function of the actors' gaze direction and actions, whereas 8-month-old infants' attentional focus would be less predictable from the actors' social cues. This prediction for 8-month-old infants was predicated on a number of specific findings: Most of the current evidence suggests that infants do not respond to gaze cues as referential prior to 9 months of age, and thus they are less likely to systematically respond to gaze direction during observation of the actions of a social partner (e.g., Johnson, Ok, & Luo, 2007; Senju et al., 2008; Woodward, 2003). There is, however, a caveat to this finding. Infants as young as 3- to 4-months of age will shift their attention in the direction of averted gaze if the target consists of moving hands and objects (Amano, Kezuka, & Yamamoto, 2004; Deák et al., 2018). Accordingly, we expected 8-month-old infants to respond to averted gaze more like 12-month-old infants when this gaze was coupled with object-directed actions. Less clear was how participants in all three age groups would respond to social cues that were incongruent (e.g., direct gaze from the viewer's face while performing an object-directed action). As we will discuss, object-directed actions were often the best predictor of when infants would share attention with the actors in the videos.

2. Method

2.1. Participants

Twenty-two eight-month-old infants (M = 243.0-days, SD = 8.7-days; 11 females, 11 males), 20 twelve-month-old infants (M = 371.6-days, SD = 8.7-days; 7 females, 13 males), and 20 adults (10 females, 10 males) comprised the sample for this study. Two additional eight-month-old infants were tested but were excluded due to fussiness or our inability to calibrate the eye-tracking system and record valid data. Parents provided consent for their child's participation and all infants received a nominal gift for participating. Infants were primarily from middle-class families and were Caucasian. They were contacted by mail based on birth records and community outreach. Adults were recruited from other laboratories in the department where they worked as research assistants. They signed a consent form and were naive to the purpose of the study.

2.2. Stimuli and apparatus

We created an initial library of stimulus videos by filming five female actors demonstrating the use of eight different sets of objects to a one-year-old infant. The objects included a stuffed animal and a fitting t-shirt, a box of crayons and a paper printout of a tree, a mug and bottle of cola, a gift wrapped box and a bow, a four-piece infant puzzle with the pieces removed, a ring stacker with two rings on the post and three rings on the table, a pair of scissors and a piece of paper with a dotted line down the middle, and a shape sorter with three different shapes on the table. Our goal was to elicit naturalistic infant-directed actions that alternated between the infant and the objects. We positioned the actors so that they faced the one-year-old infant and for each video they were

Table 1
Mean proportion of retained gaze samples per stimulus video and age group.

Stimulus video	8-month-olds	12-month-olds	Adults
1. Actor 1, dressing stuffed animal	0.683	0.796	0.972
2. Actor 2, dressing stuffed animal	0.635	0.817	0.952
3. Actor 3, dressing stuffed animal	0.789	0.899	0.975
4. Actor 1, coloring with crayon	0.668	0.814	0.970
5. Actor 1, pouring cola into cup	0.798	0.885	0.961
6. Actor 2, pouring cola into cup	0.767	0.879	0.972
7. Actor 4, pouring cola into cup	0.767	0.889	0.974
8. Actor 5, pouring cola into cup	0.829	0.825	0.985
9. Actor 2, placing bow on gift box	0.661	0.776	0.986
10. Actor 3, placing bow on gift box	0.792	0.806	0.981
11. Actor 1, placing puzzle pieces	0.727	0.828	0.979
12. Actor 1, stacking rings on peg	0.826	0.890	0.966
13. Actor 2, stacking rings on peg	0.860	0.886	0.969
14. Actor 4, stacking rings on peg	0.839	0.890	0.972
15. Actor 4, cutting paper with scissors	0.808	0.748	0.949
16. Actor 4, placing shapes in shape sorter	0.831	0.929	0.974
Overall	0.767	0.847	0.971

instructed to socially engage the infant while demonstrating an action with the props appearing on the table. It should be noted that they did not see the objects beforehand nor were they instructed on how to perform the actions. A digital camera positioned above the infant's head filmed the actor in order that they would be seen in each video facing the viewer. Their accompanying speech was also recorded and played with the video stimuli. A total of 40 videos were filmed and 16 were selected for meeting standards of quality (e.g., no audible or visible interference from the infant) and viewing duration (M = 25.1 s, *range* = 16.3 to 41.4 s). As summarized in Table 1, this stimulus selection strategy resulted in an uneven distribution of which model demonstrated which action (i.e., Actor 1 demonstrated five different actions, Actor 2 four actions, Actor 3 two actions, Actor 4 four actions, and Actor 5 demonstrated only one action). We cropped and edited these videos to 800 × 600 screen resolution. Three of these videos appear as Supplementary files 1–3 and can also be accessed through the Open Science Framework (OSF) at: https://osf.io/hc2g3/. Each participant's fixations are superimposed as dots on these videos with diameter scaled to fixation duration.

Participants viewed the videos on a Tobii 2150 corneal reflection eye-tracking system with a 21.3″ flat LCD screen (Tobii Technology Inc., Stockholm, Sweden). The system tracked gaze of both eyes with an infrared eye tracker integrated into the monitor (precision: 1 deg.; measurement error: 0.5 deg.; sampling rate: 50 Hz). In order to assess where participants looked while viewing each video, areas of interest (AOIs) were calculated around the head, hands, and objects of each video (see below for more details).

2.3. Procedure

Infants sat on their parent's lap facing the screen with an average pupil-to-screen distance of 67.9 cm; adults viewed the screen from an average distance of 75.7 cm. Adults were instructed to view the videos as if they would have to reproduce the actions that the model demonstrated. Calibration of eye gaze was conducted with Tobii Clearview software which displayed a spinning multi-colored disc (extended diameter = 5.5°) expanding and contracting with an accompanying rhythmic sound at each of the four corners as well as at the center of the screen. We repeated calibration of those locations that the software revealed were unsuccessful. Participants viewed all 16 stimulus videos in pseudorandom order, with the constraint that they never saw the same actor or stimulus objects on successive trials. The entire procedure took approximately 15 min.

2.4. Data reduction

2.4.1. Preprocessing

The eye-tracking system recorded the time of each sample, x,y coordinates for both eyes, pupil-to-screen distances, and data validity (with a five point scale). We calculated each gaze sample as the average of the two eyes when data validity was high for both (validity score = 0), and included only the data from one eye if validity was low for the other eye (validity score = 1 or 3). Low validity could be a function of blinks, measurement error, or looking away from the screen. In instances of low validity from both eyes, missing data lasting < 80 ms (4 gaze samples) were linearly interpolated between the nearest reliable samples, and the resulting gaze was filtered using a Savitzky-Golay filter[1] to remove high frequency noise. With this method we retained 77% of the gaze samples for 8-month-old infants, 85% for 12-month-old infants, and 97% for adults. Table 1 summarizes the proportion of gaze samples per age we included in the final sample for each stimulus video.

2.4.2. Fixation detection

Information processing is suppressed during eye movements (Matin, 1974), so it is necessary to identify when participants' eyes were stationary and engaged with the stimulus. We therefore implemented a classification procedure to estimate for each gaze sample whether the eyes were stationary (fixation) or moving (saccade). Instantaneous velocities[2] measured during saccades tend to be two to three times higher than during fixations, making them useful for discriminating between these two states; however, due primarily to measurement noise, some standard approaches (e.g. a fixed velocity threshold) are unreliable, confusing noise with eye movements, thereby producing too many, often short fixations (Salvucci & Goldberg, 2000). We therefore employed a more robust, Hidden Markov Model (HMM), to estimate the state of the eyes by combining velocity measurements from adjacent samples (see Appendix A for details of this procedure). Fixations identified in this manner, with durations < 100 ms or > 5000 ms, were removed from further analyses. Finally, two trained coders reviewed the resulting fixations by visual inspection, adding fixations that were missed by the automated procedure or removing spurious fixations that were too short or too variable.

2.4.3. Areas of interest

In order to infer which of the locations in the scene was the likely target of participants' attention, we adopted an Area of Interest (AOI) approach. We defined areas of interest (AOIs) around the models' faces, hands, and objects by identifying the nearest points above and to the left and below and to the right of each stimulus region. For instance, we used the model's hairline to define the upper horizontal edge, the model's chin to define the lower horizontal edge, and the maximal face width at the cheekbones to define the left and right vertical edges of the face AOI. These AOIs were calculated for each video frame (using an in-house Java tool) because the stimulus was dynamic and thus the AOIs were continuously changing. We digitally smoothed all of the captured points with a FIR filter (MATLAB function *filtfilt*) to remove high frequency jitter that occurred as a function of the human image coding. Fig. 1 illustrates the AOIs for one frame of one stimulus video, and includes the mean sizes of each of the AOIs from each of the frames of that video. The sizes of each AOI varied as a function of the stimulus video, but overall the areas of the AOIs within a category (e.g., face, hands) were fairly similar and showed a coefficient of variation (SD/Mean) ranging between 0.11 and 0.41 deg. of visual angle (See Appendix B). Due to limitations in the resolution of the video and noise
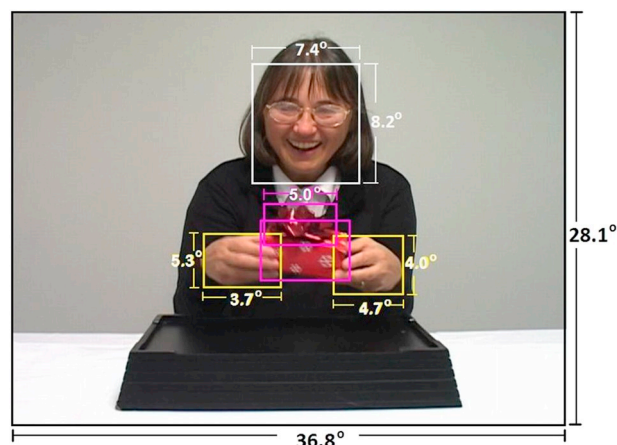

Fig. 1. Screen shot from one of the stimulus videos with an overlay of the AOIs. The numerical dimensions of each AOI correspond to their mean size based on all frames of the video.

in the measured gaze position, we adopted a binary coding scheme, classifying each fixation as directed toward either the *face* of the actor, or her *hands* and/or *objects* appearing in the scene (because hands and objects often overlapped, these were subsequently collapsed into a single category). Fixations were classified according to the nearest AOI, and those located at a distance of > 5.3 deg. of visual angle from the nearest AOI were considered to be directed toward neither category and were removed from further analysis (< 0.2% of all fixations).

2.5. Stimulus coding

For the purposes of analyzing the effects of social cues on looking, we coded the actors' gaze direction and action type. *Direct gaze* occurred when the actor's head orientation and eyes were directed toward the observer. This behavior typically signals an intention to communicate (Csibra, 2010). *Averted gaze* occurred when the actor's head orientation and eyes were directed toward one of the objects on the table signaling the actor's interest in that object. If an actor manipulated or transported an object with her hands it was coded as an object-directed action, whereas it was coded as an infant-directed action if the actor pointed to an object or held and waved it.[3] Observers viewed the 16 stimulus videos and coded the onset and offset times (resolution of 33 ms) of the actors' gaze direction and actions. Each video was coded by two observers who independently classified actors' gaze direction and manual actions before combining their judgments and resolving any disagreements about ambiguous cases.

Table 2 summarizes the mean proportion of each video during which the actors exhibited each of these contextual cues. It is important to note that the stimulus videos captured the actors performing different actions in a naturalistic situation. As a consequence, the proportion of time the actors displayed direct vs averted gaze or infant-directed vs goal-directed actions was not counterbalanced resulting in a 50/50 split. Instead, the likelihood of displaying any of these behaviors corresponded to the natural statistics of the situation. As can be seen in Table 2, averted gaze appeared more frequently than direct gaze (57 vs 43%), and object-directed actions appeared more frequently than infant-directed actions (65 vs 35%). The 2 × 2 combination of both social

---

[1] MATLAB function *sgolayfilt* with polynomial order = 2, window length = 15.

[2] $v = \frac{\sqrt{dx^2 + dy^2}}{dt}$

[3] Strictly speaking, this is a "viewer-centered" action because both infants and adults participated in the study. Nevertheless, we opted for the term "infant-directed" because these actions were spontaneously produced by actors while demonstrating actions to an infant who was viewing the actions live and in real-time. Also, pointing was coded, but it appeared only in four videos. If it had appeared more frequently, we would have been able to compare responses between pointing to and holding and waving objects.

Table 2
Mean proportion of each video displaying infant-directed vs. object-directed actions and direct vs. averted gaze.

Gaze direction	Action type	Mean (SEM)
Direct	Infant-directed	0.27 (0.03)
Direct	Object-directed	0.16 (0.03)
Averted	Infant-directed	0.08 (0.02)
Averted	Object-directed	0.49 (0.04)

cues resulted in frequencies ranging from 0.08 (averted gaze and infant-directed actions) to 0.48 (averted gaze and object-directed actions).

Each fixation was classified in terms of the actors' gaze direction and action. Sometimes, the coded social cue would change during the fixation, but it was necessary to assign each fixation to mutually exclusive codes (e.g., either direct- or averted-gaze). In order to classify these fixations, we created a temporal window beginning 500 ms before fixation onset and lasting until the beginning of the next eye movement. Within this window, we determined how long the actor exhibited *direct* versus *averted* gaze, as well as the relative time she spent engaging in *object-directed actions* as opposed to *infant-directed actions*. The gaze and action codes that were assigned to that fixation for the majority of its duration were selected as the correct codes, resulting in four (2 gaze direction × 2 actions) types of fixations.

3. Results

The main goal of this study was to test whether infants and adults modulated their attention to faces and objects as a function of gaze direction and action type. In order to address this question, it was necessary to first determine how visual attention should be measured. Although most developmental studies measure visual attention in terms of total duration of looking, we opted to measure attention exclusively in terms of visual fixations. Our eyes scan the visual world via saccadic movements and fixate on relevant regions to enable foveation for high-resolution sampling of visual information (Guerrasio, Quinet, Büttner, & Goffart, 2010). The number and duration of these fixations is determined by both bottom-up (e.g., image statistics, stimulus salience) as well as top down (e.g., strategies, goals) processes (Tatler, Brockmole, & Carpenter, 2017). We are unable to probe these processes directly, but measuring when fixations shift and how long they remain focused on the face or objects offers some important insights into when and how social cues are attracting infants' attention. The duration of each fixation reflects encoding of the stimulus region as well as decisions regarding where and when to look next, whereas visual processing essentially ceases during the saccade (Matin, 1974). Accordingly, we will focus our analyses on the proportion of fixations directed toward specific locations as well as the duration of these fixations. The fixation data, grouped by video, AOI, and age, is accessible through OSF at: https://osf.io/hc2g3/.

We begin by analyzing the proportion of fixations directed toward the face vs. objects as a function of age, gaze direction, and action type. One limitation of this aggregate measure is that it is not possible to determine whether the number of fixations is scaled to the duration of looking, because fixations to some locations may be longer than to others. Thus, we will also assess time-dependent measures, including mean fixation durations as well as dwell times to faces and objects as a function of age, gaze direction, and action type. Dwell time was calculated as the total time spent looking at an AOI from entry to exit (Holmqvist et al., 2011). At a minimum dwell times are equal to fixation durations, but often they exceed these durations because there will be two or more consecutive fixations directed to the AOI. These analyses will enable us to determine whether not only the location of fixations, but also the individual fixations and dwell times are systematically related to the spatiotemporal demands associated with joint
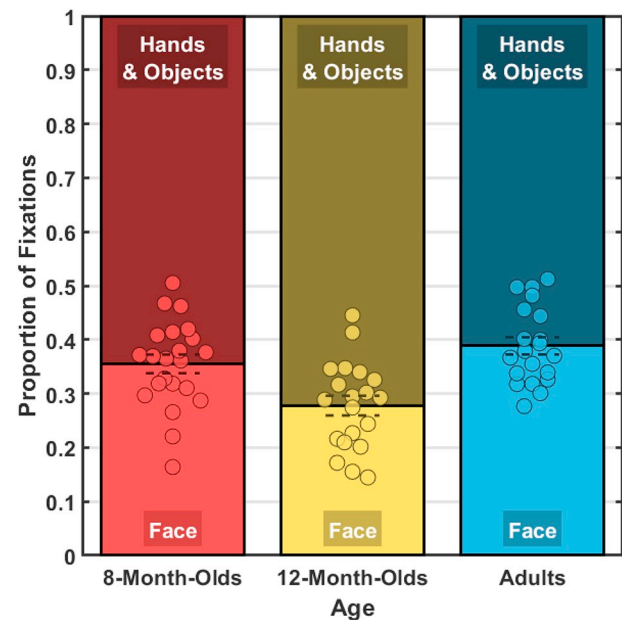

Fig. 2. Mean proportion of fixations directed at face vs. hands-and-objects as a function of age; each individual participant's proportion is depicted with a circle. The dashed lines represent ± 1 standard error of the mean.

attention for both infants and adults. Lastly, we will assess the time course between the initiation of a social cue (gaze or action) and the subsequent shift in fixation in order to evaluate whether the same social cues are involved in attention-getting as well as attention-holding processes (Cohen, 1972).

3.1. Visual attention to faces and objects

As can be seen in Fig. 2, participants fixated more on hands-and-objects than on the actors' faces, but the specific pattern of looking varied as a function of age. The adults looked at faces (39%) more than the 8-month-old infants (36%) who looked more at faces than the 12-month-old infants (27%). It is important to note that this result is at odds with the vast majority of studies reporting that both infants and adults devote a good deal of their attention to faces (e.g., Amso, Haas, & Markant, 2014; Birmingham, Bischof, & Kingstone, 2008; Frank et al., 2009); although some recent exceptions have appeared in the literature (e.g., Fausey, Jayaraman, & Smith, 2016).

These differences were tested with a generalized linear mixed effects model which treated age (8- or 12-month-olds or adults), gaze direction (direct vs. averted), and action type (infant-directed vs. object-directed) as categorical fixed effects, and included random intercepts for each participant. The number of fixations that were directed toward the face was the dependent variable, and was treated as a proportion (binomial distribution, link = logit). Note that proportion of fixations directed toward objects was the inverse of the proportion of fixations directed toward the face (i.e., proportion of fixations toward the face and objects summed to 1.0), and thus the two variables were not independent. Nevertheless, Fig. 3 displays the results for fixations to the face and objects separately, because it is helpful to visualize these results from both perspectives. The results revealed that the proportion of fixations that were directed toward the face showed a significant effect of age, $F(2, 236) = 10.83$, $p < .001$, Cohen's $f^2 = 0.08$.[4] All age groups demonstrated significantly less looking to faces than chance, all

---

[4] According to Cohen (1988), f^2 is used for measuring effect size with generalized linear mixed effects models. Nagelkerke R-squared was used to compute f^2 (Nagelkerke, 1991).
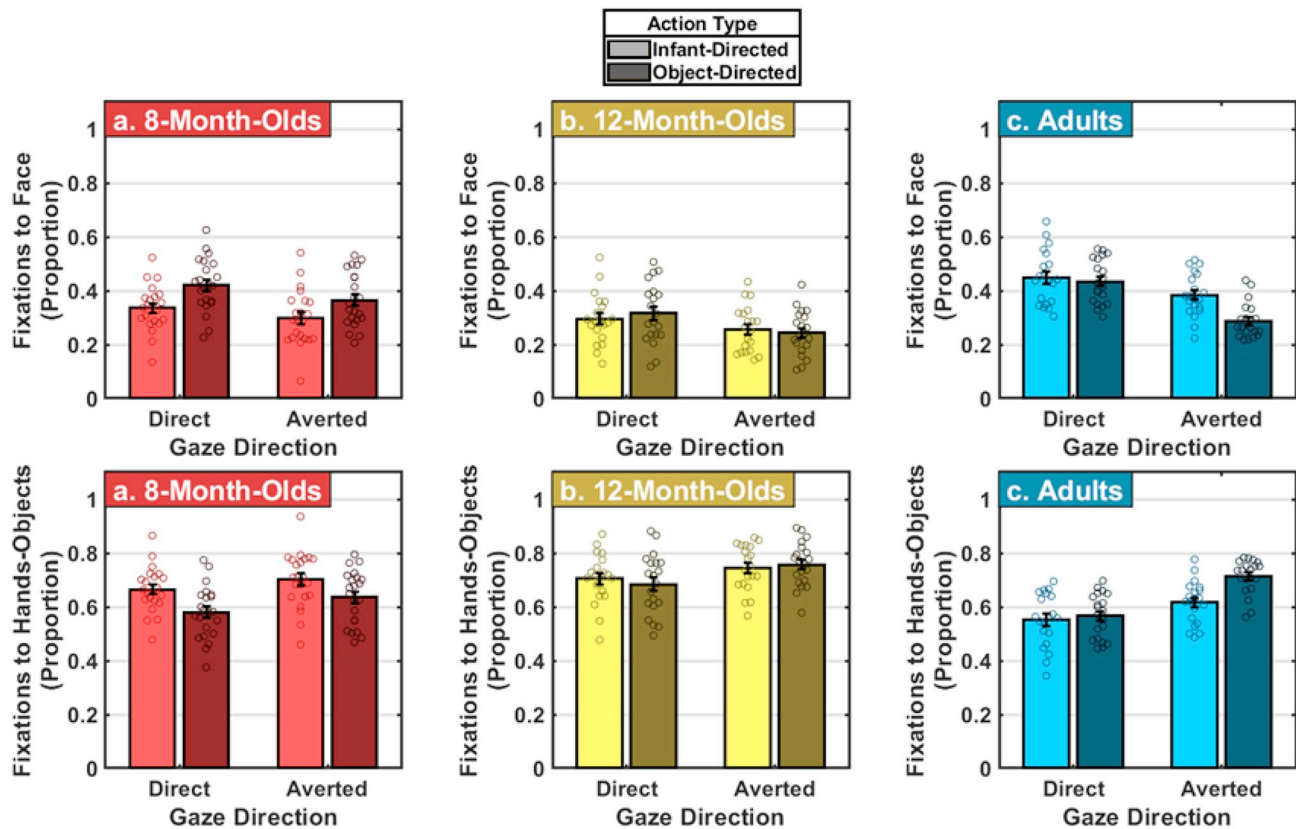
Fig. 3. Top panel. Mean proportion of fixations to the face as a function of age, gaze direction, and action type. Bottom panel. Mean proportion of fixations to the hands-and-objects as a function of age, gaze direction, and action type. (Error bars represent ± 1 standard error of the mean.)

$ts > 15.0$, $p < .001$,[5] Cohen's $h > 0.46$, with 12-month-old infants showing a significantly smaller proportion of fixations to the face than 8-month-old infants and adults, both $ts > 3.10$, $p < .05$, $h = 0.17$, 0.24 respectively; these two groups did not differ from one another, $t(40) = -1.42$, $p = .49$, $h = 0.07$.

In addition to age, gaze direction significantly affected the proportion of fixations to the face, $F(1, 236) = 124.02$, $p < .001$, $f^2 = 0.66$. As can be seen in Fig. 3, there were more fixations to the face during direct gaze (38 ± 3.9%) than averted gaze (31 ± 2.7%), but the magnitude of this difference varied with age, $F(2, 236) = 8.94$, $p < .001$, $f^2 = 0.07$. Pairwise t-tests revealed that this effect was significant for all age groups, all $ts > 4.09$, $p < .01$, $h > 0.09$, but adults displayed the largest difference (10.6%), with both 8-month-old (4.6%) and 12-month-old (5.6%) infants displaying a difference that was approximately half the size. This difference between adults and infants was significant ($t(38) = -4.03$, $p < .01$, h = 1.25, $t(38) = -3.45$, $p < .01$, $h = 0.10$, for 8- and 12-month-old infants, respectively) but the difference between 8- and 12-month-old infants was not significant, $t(40) = -0.60$, $p = 1.0$, $h = 0.02$.

The proportion of fixations to the face was not modulated by action type overall, $F(1, 236) = 1.27$, $p = .26$, $f^2 < 0.01$, but it did interact with age, $F(2, 236) = 36.71$, $p < .001$, $f^2 = 0.35$. This result is attributable to a reversal in the pattern of responding across age: 8-month-old infants were significantly more likely to fixate the face during object-directed actions (39%) than during infant-directed actions (32%), $t(21) = -4.54$, $p < .01$, $h = 0.16$, whereas 12-month-old infants showed no difference, (27% of fixations for both object-directed and infant-directed actions), $t(19) = -0.24$, $p = .81$, $h = 0.01$. By contrast, adults showed the opposite pattern, looking more to the face during infant-directed actions (42%) than during object-directed actions

(36%), $t(19) = 4.51$, $p < .01$, $h = 0.11$. The difference in looking patterns between 8-month-old infants and adults was significant, $t(40) = -6.23$, $p < .001$, $h = 0.26$. Lastly, the interaction between gaze direction and action type was also significant, $F(1, 236) = 10.83$, $p < .01$, $f^2 = 0.04$. There was less looking to the face during averted gaze and object-directed actions than during either averted gaze and infant-directed actions or direct gaze and either action type.

In sum, participants at all three ages looked more at the face during direct than averted gazed. By contrast, the results differed by age as a function of action type: 8-month-old infants and adults demonstrated opposing patterns. Adults looked more at the face during infant-directed than object-directed actions, especially when object-directed actions were coupled with averted gaze. Eight-month-old infants looked less at the face during infant-directed than object-directed actions.

3.2. Fixation durations

An advantage of measuring fixation durations is that this time-dependent variable should be sensitive to processing differences as a function of age and social cues. Moreover, it is possible to statistically compare fixations to the face and to hands-and-objects because, unlike proportions, fixation durations for these two gaze locations are independent of each other. These differences were tested with a linear mixed effects model involving age (8- or 12-month-old infants or adults), location (face vs. hands-and-objects), gaze direction (direct vs. averted), and action type (infant-directed vs. object-directed) as fixed effects, with random intercepts for each participant. The following main effects were significant: age ($F(2, 472) = 3.89$, $p = .02$, $f^2 < 0.01$), gaze location ($F(1, 472) = 11.33$, $p < .001$, $f^2 = 0.03$), and gaze direction ($F(1, 472) = 9.97$, $p < .002$, $f^2 = 0.03$). As can be seen in Fig. 4, many of these results were qualified by higher-level interactions. Fixation duration was modulated by the interaction between gaze

---

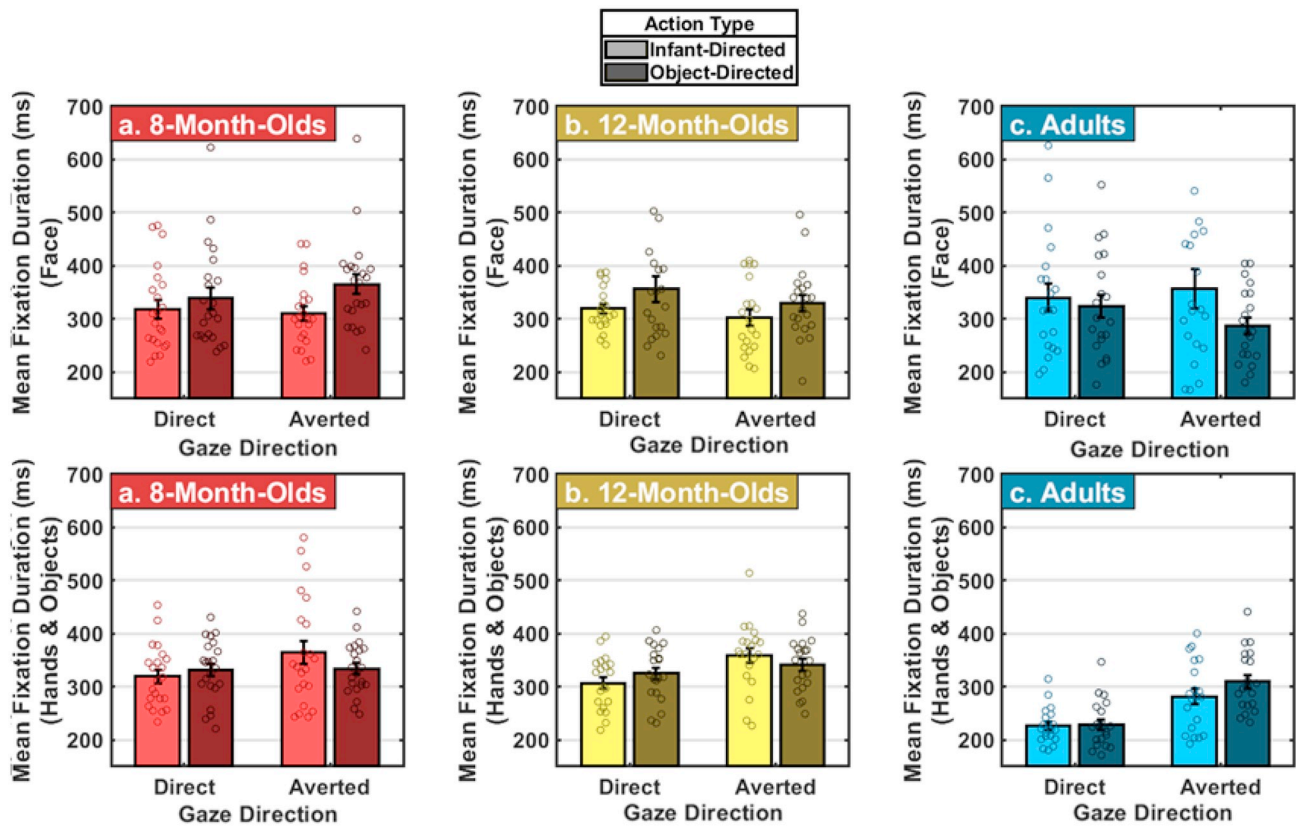[5] Bonferroni corrections were used on all post-hoc tests.

Fig. 4. Mean fixation durations as a function of age, gaze direction, and action type. Top panel displays fixations directed to the face, and bottom panel displays fixations directed to hands-and-objects. (Error bars represent ± 1 standard error of the mean.)

location and gaze direction. In general, fixations were shorter during direct than averted gaze when looking at hands-and-objects, $F(1,472) = 20.76$, $p < .001$, $f^2 = 0.05$. Age interacted with gaze location, $F(2, 472) = 18.47$, $p < .001$, $f^2 = 0.10$, and it also interacted with action type, $F(2, 472) = 3.14$, $p = .04$, $f^2 = 0.02$; moreover, all three of these variables interacted together, $F(2, 472) = 9.23$, $p < .001$, $f^2 = 0.05$. Overall, these effects are attributable to greater differences in fixation durations between infant-directed and object-directed actions when looking at the face than at hands-and-objects, but they also interact with age. At 8- and 12-months of age, fixation durations are shorter during infant-directed than during object-directed actions, whereas, fixation durations are shorter for adults when looking at object-directed than infant-directed actions; these age differences were significant at both 8-months, $t(40) = -2.83$, $p < .05$, $d = 0.87$, and 12-months of age, $t(38) = -3.21$, $p < .01$, $d = 1.01$. Lastly, the four-way interaction between age, gaze location, gaze direction, and action type was significant, $F(2, 472) = 4.55$, $p < .001$, $f^2 = 0.02$. In addition to the differences already noted, this interaction reflects significantly shorter fixations for adults when viewing hands-and-objects during direct gaze than during any other combination of gaze location, gaze direction, and action type, all $ts > 4.05$, $ps < 0.05$, $ds > 0.82$.

3.3. Dwell times

Dwell times and fixation durations were not correlated. The correlation between the two measures when looking at the face was $r(62) = 0.07$, $p = .59$, and when looking at the hands-and-objects, $r(62) = 0.02$, $p = .88$. For this analysis it was not possible to assess the effects of gaze direction and action type, because the onset of a new social cue could occur during or after the first fixation, and thus a specific gaze direction or action type might change one or more times during the dwell time. Mean dwell times were thus analyzed with linear

regression models involving age and gaze location (face vs. hands-and-objects) as predictors. There was a significant effect of age, $F(2, 118) = 8.77$, $p < .001$, $f^2 = 0.14$, and gaze location, $F(1, 118) = 113.24$, $p < .001$, $f^2 = 0.91$ as well as a significant interaction between these two variables, $F(2, 118) = 12.10$, $p < .001$, $f^2 = 0.20$. As can be seen in Fig. 5, 8- and 12-month-old infants as well as adults dwelled longer at hands-and-objects than at faces, all $ts > 5.39$,
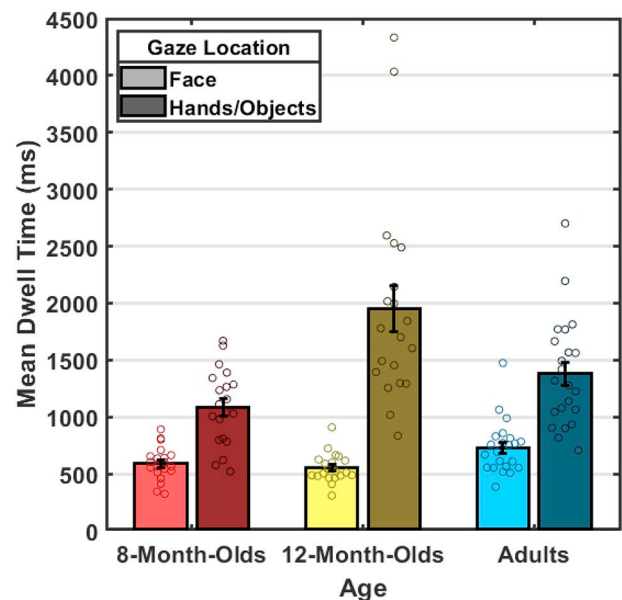

Fig. 5. Mean dwell times as a function of age and gaze location. (Error bars represent ± 1 standard error of the mean.)

$p < .001$, $d > 1.15$. This difference was greater for 12-month-old infants than for 8-month-old infants, $t(40) = 3.14$, $p < .01$, $d = 0.97$ or adults, $t(38) = 4.07$, $p < .001$, $d = 1.29$. The difference between 8-month-old infants and adults was not significant, $t(40) = 1.11$, $p > .82$, $d = 0.34$.

3.4. Gaze shift latencies following onset of social cues

One limitation of the previous analyses is that it was not possible to determine if the onset of a social cue triggers a shift in attention or simply sustains attention to the face or hands-and-objects. In order to test whether the onset of a social cue re-orients infants' attention, it is necessary to time-lock fixations to this onset. For this analysis, we classified each fixation according to the onset of the most recent gaze direction (direct vs. averted) and calculated the time between the onset of this social cue and the beginning of the fixation. This process was repeated for all fixations following the onset of an action type (infant-directed vs. object-directed action). Because these two social cues often follow each other closely in time, it is difficult to determine whether a fixation occurs in response to a gaze shift or action type. As a consequence, we analyzed each fixation twice: once in response to the gaze shift and once in response to the change in action type. The proportion of fixations to the face following a social cue was calculated every 500 ms for a period of 4 s. As such, this analysis reveals the likelihood of fixations to the same locations increasing or decreasing over time. Let's consider an example in which 20% of a participant's fixations are already focused on the face at the onset of direct gaze when summed across all instances of this cue; 3 s later 65% of all fixations are focused on the face, but by the fourth second only 35% of all fixations are on the face. This means that there is a 20% chance of fixating the face at the onset of direct gaze, and that the likelihood increases to 65% by 3 s and then declines to 35% by 4 s. Four sec represented the cut-off for this analysis, because both gaze and action cues alternate fairly frequently, and thus it was rare to find periods of time where both social cues remained constant for over 4 s. In essence, this analysis indicates when the tendency to fixate the face is increasing or decreasing as would be expected following direct gaze or object-directed actions, respectively.

The first analysis assessed the likelihood of participants fixating the face with a generalized linear model (GLM) including age and gaze direction as categorical, fixed effects. Elapsed time since the onset of the most recent social cue (0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0 s), was also included as a linear predictor (representing a persistent increase in fixations to the face) and as a quadratic predictor (representing an initial increase in fixations to the face followed by a return to baseline). Given that some fixations were likely already on the face at the onset of the cue, it was necessary to establish a baseline rate of fixations in order to assess whether fixations increased or decreased following the cue onset. Baselines were calculated as the mean proportions of fixations on the face during each of the four social cues (direct gaze, averted gaze, infant-directed actions, object-directed actions) for each age separately. These are represented on each of the four graphs in Fig. 6 as three shaded horizontal bars (corresponding to each age) extending from zero to four sec, with the width of the bar corresponding to one standard error of the mean.

The full model revealed a significant three-way interaction between age, gaze direction, and both linear, $F(2, 1093) = 19.54$, $p < .001$, $f^2 = 0.05$, and quadratic, $F(2, 1093) = 12.16$, $p < .001$, $f^2 = 0.03$, components of elapsed time following onset of gaze cue (see Fig. 6). These interactions were explored further with simple effects analyses specifically focused on the time-course within each age group for each gaze type.

3.4.1. Onset of direct gaze

The first simple effects analysis assessed the time course of the proportion of fixations to the face following the onset of direct gaze. As can be seen in the upper left panel of Fig. 6, 12-month-old infants and adults showed an increase (across the four second time window) in the likelihood that new fixations would be directed to the face, peaking

around 2.5 s. By contrast, the function for 8-month-old infants was flat. These observations were statistically reliable. The linear predictor was significant for 12-month-old infants, $F(1, 177) = 57.41$, $p < .001$, $f^2 = 0.46$, and adults, $F(1, 176) = 69.60$, $p < .001$, $f^2 = 0.47$, but not for 8-month-old infants, $F(1, 193) = 0.85$, $p > .35$, $f^2 = 0.01$. At each age, the direction of the effect was positive (although not significant at 8 months of age); 8-month-old infants, $\beta = 0.016$ (± 0.017); 12-month-old infants, $\beta = 0.375$ (± 0.050); adults, $\beta = 0.329$ (± 0.039). This indicates that fixations generated within 4 s of the onset of direct gaze were more likely to occur on the face than hands-and-objects. The size of this effect was smaller for 8-month-old than 12-month-old infants, $t(40) = -7.12$, $p < .001$, $d = 2.20$ and adults, $t(40) = -7.53$, $p < .001$, $d = 2.33$, who did not differ, $t(38) = 0.72$, $p > .91$, $d = 0.23$. In order to assess whether this effect persisted throughout the 4 s or dissipated, we assessed the quadratic term, which was not significant for 8-month-old infants ($\beta = -0.015 \pm 0.008$), $F(1, 192) = 3.63$, $p > .05$, $f^2 = 0.02$, but was significant for 12-month-old infants, $F(1, 177) = 26.11$, $p < .001$, $f^2 = 0.20$, and for adults, $F(1, 176) = 37.19$, $p < .001$, $f^2 = 0.25$. The direction of the effect was negative, indicating an initial increase followed by a later decrease back to baseline; 12-month-old infants $= -0.039$ (± 0.008); adults $= -0.039$ (± 0.006). The difference between these two age groups was not significant, $t(38) = 0.02$, $p = 1.0$, $d = 0.01$.

3.4.2. Onset of averted gaze

The next analysis focused on fixations following averted gaze. In general, these results differ from those involving direct gaze not only because the direction of effects is reversed, but also because 8-month-old infants mirror the responses of the 12-month-old infants except that the amplitude is smaller. As can be seen in the upper right panel in Fig. 6, there was a significant effect for the linear component of elapsed time at all three ages: 8-month-old infants, $F(1, 194) = 28.71$, $p < .001$, $f^2 = 0.18$; 12-month-old infants, $F(1, 177) = 113.38$, $p < .001$, $f^2 = 0.72$; adults, $F(1, 177) = 161.28$, $p < .001$, $f^2 = 0.73$. Unlike the effect of direct gaze, averted gaze led to fewer looks to the face for all age groups, and resulted in negative slopes; 8-month-old infants, $\beta = -0.231$ (± 0.043); 12-month-old infants, $\beta = -0.486$ (± 0.046); adults, $\beta = -0.465$ (± 0.037). The 8-month-old infants showed a significantly smaller decrease than adults, $t(40) = 4.09$, $p < .001$, $d = 1.26$, and 12-month-old infants, $t(40) = 4.07$, $p < .001$, $d = 1.26$; 12-month-old infants and adults revealed a similar decrease, $t(38) = 0.37$, $p = 1.0$, $d = 0.12$. An analysis of the quadratic component revealed a significant effect for all three age groups: 8-month-old infants, $F(1, 194) = 10.82$, $p < .001$, $f^2 = 0.07$; 12-month-old infants, $F(1,177) = 83.21$, $p < .001$, $f^2 = 0.54$; adults, $F(1, 177) = 106.67$, $p < .001$, $f^2 = 0.45$. This effect resulted in a positive coefficient demonstrating an initial decrease in the proportion of fixations that were directed to the face followed by a later reversal: at 8-months of age this effect was smallest, $\beta = 0.0229$ (± 0.007), but became more pronounced at 12-months of age, $\beta = 0.061$ (± 0.007), and was similar for the adults, $\beta = 0.056$ (± 0.005). 8-month-old infants showed smaller effects than the 12-month-olds, $t(40) = -4.11$, $p < .001$, $d = 1.27$, and adults, $t(40) = -3.92$, $p < .01$, $d = 1.21$, who did not differ from one another, $t(38) = 0.55$, $p = 1.0$, $d = 0.18$.

The last two analyses assessed the effects of the onset of the two action types (infant-directed, object-directed) on fixations on the face. Beginning with an analysis of the full model involving age, action type, and elapsed time (linear and quadratic), the results revealed a significant three-way interaction for the linear, $F(2, 1098) = 17.58$, $p < .001$, $f^2 = 0.04$, and quadratic, $F(2, 1098) = 7.75$, $p < .001$, $f^2 = 0.02$, terms. In order to clarify this interaction, we analyzed the simple effects.

3.4.3. Onset of infant-directed actions

An analysis of the proportion of fixations following the onset of infant-directed actions revealed that 12-month-old infants and adults were more likely to look at the actor's face across a four-second window
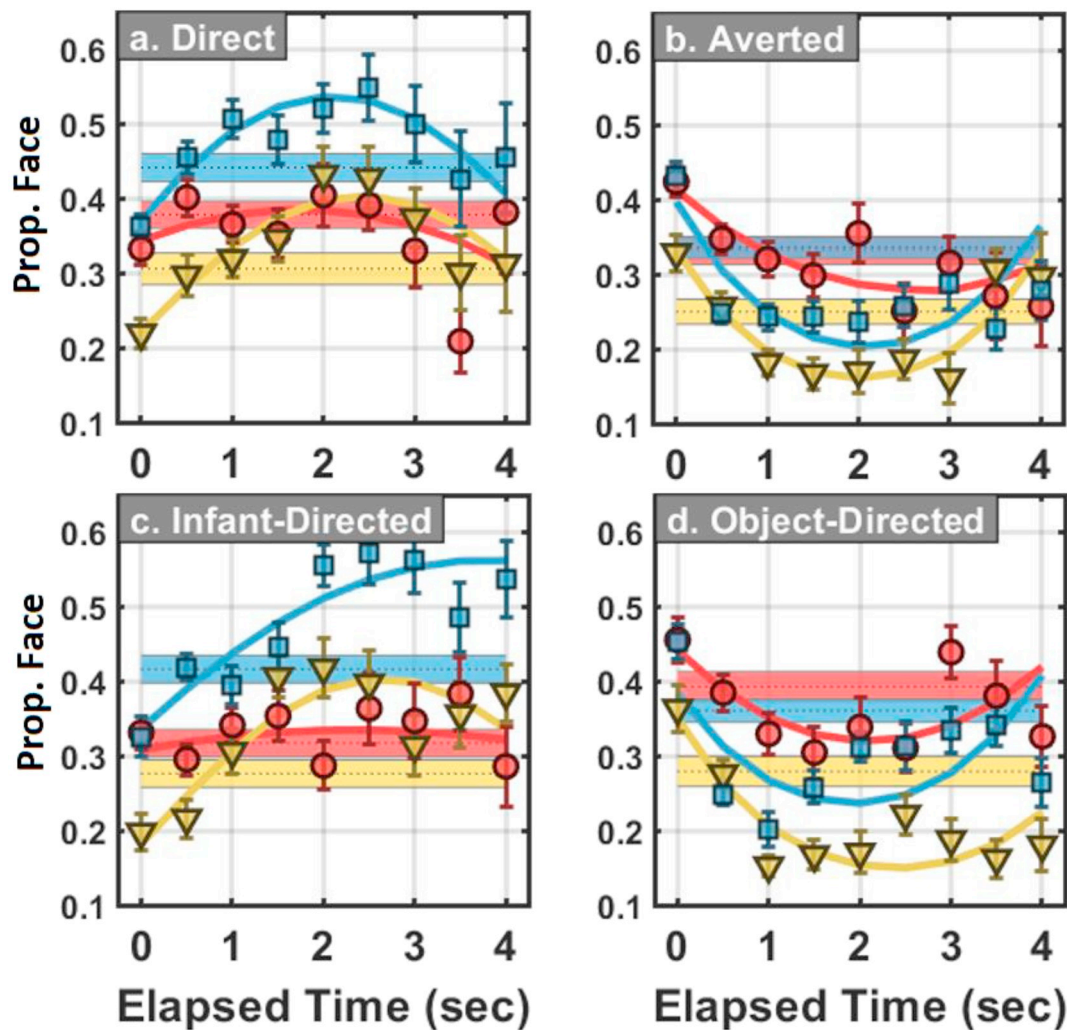
Fig. 6. Mean changes in proportions of fixations to the face following changes in gaze direction or action type as a function of age. (8-month-old infants = circles; 12-month-old infants = squares; adults = inverted triangles). Baselines were determined by calculating the mean proportions of looking at the face during each of the four social cues (direct, averted, infant-directed, object-directed) for each age separately. These are represented as the shaded horizontal regions corresponding to confidence intervals of the means: 8-month-old infants = red; 12-month-old infants = blue; adults = yellow. (Error bars represent ± 1standard error of the mean.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

than were 8-month-old infants, but this effect was significantly greater for 12-month-old infants than adults. As seen in the lower left panel of Fig. 6, there was no significant effect of the linear term for 8-month-old infants, $F(1, 195) < 0.99, p > .32, f^2 < 0.01$. By contrast, there was a significant linear effect for 12-month-old infants, $F(1, 177) = 61.21, p < .001, f^2 = 0.41$, and adults, $F(1, 177) = 18.34, p < .001, f^2 = 0.22$, suggesting a greater likelihood of fixations on the face following an infant-directed action. The linear effect was smaller for 8-month-old, $\beta = 0.053 (\pm 0.053)$ than 12-month-old infants, $\beta = 0.432 (\pm 0.055), t(40) = -4.95, p < .001, d = 1.53$, or adults, $\beta = 0.252 (\pm 0.042), t(40) = -2.91, p < .05, d = 0.90$; the difference between 12-month-old infants and adults was also significant with infants producing a steeper slope, $t(38) = 2.59, p < .05, d = 0.82$. The positive quadratic term was significant for the 12-month-old infants, $F(1, 177) = 29.89, p < .001, f^2 = 0.20$, and adults, $F(1, 177) = 7.73, p < .01, f^2 = 0.04$, but not for the 8-month-old infants, $F(1,195) = 0.58, p > .44, f^2 = 0.01$. The 12-month-old infants also revealed the strongest quadratic effect, $\beta = -0.041 (\pm 0.007)$, which was marginally significantly greater than the effect for adults, $\beta = -0.017 (\pm 0.006), t(38) = 2.51, p = .056, d = 0.80$, and significantly greater than the effect for 8-month-old infants, $\beta = -0.006 (\pm 0.007), t(40) = 3.36, p < .01, d = 1.04$.

3.4.4. Onset of object-directed actions

Finally, the analysis of the changing proportion of fixations following object-directed actions revealed a consistent response pattern across all three age groups (Fig. 6, lower right panel): fixations shift from the face to objects relatively quickly after an object-directed action (within 1–2 s), but also reverse from objects back to the face within the next 2 s. There was a significant effect of the linear component for 8-month-old infants, $F(1, 195) = 32.15, p < .001, f^2 = 0.15$, 12-month-old infants, $F(1, 177) = 91.54, p < .001, f^2 = 0.55$, and adults, $F(1, 177) = 93.63, p < .001, f^2 = 0.30$. The magnitude of this negative linear trend was smallest for 8-month-old infants, $\beta = -0.257 (\pm 0.045)$, and was reliably smaller than the linear trend for 12-month-old infants, $\beta = -0.480 (\pm 0.050), t(40) = 3.32, p < .01, d = 1.02$. Adults showed an in-between response, $\beta = -0.368 (\pm 0.038)$, which did not differ from 8-month-old, $t(40) = 1.87, p > .21, d = 0.58$ or 12-month-old infants, $t(38) = -1.78, p > .24, d = 0.56$. There was also a significant effect of the quadratic term for all three age groups: 8-month-old infants, $F(1, 195) = 24.54, p < .001, f^2 = 0.12$, 12-month-old infants, $F(1, 177) = 52.06, p < .001, f^2 = 0.29$, and adults, $F(1, 177) = 83.41, p < .001, f^2 = 0.26$. The strength of the effect was consistent across ages, all $ts < 2.09, p > .14, d < 0.65$.

4. Discussion

A prerequisite for joint attention is that both infants and adults coordinate their focus of attention with the gaze direction and actions of their social partner. Most previous eye tracking studies presented faces in isolation, which obviated the need for joint attention. By contrast, this study presented videos of actors appearing to interact with observers so that we could precisely measure how direction of gaze and actions affect the spatiotemporal patterning of infants' gaze during more naturalistic social interactions. One of the most striking findings was that neither infants nor adults looked primarily at the faces in the stimulus videos. In fact, the majority of fixations were directed toward the hands-and-objects of the actor (61% to 73%), but the findings suggested attention to the face was still advantageous for understanding the communicative intent of the social partner. Adults looked more at faces than did either 8- or 12-month-old infants, but this difference was only significant between 12-month-old infants and adults. Although previous findings in the social attention literature reveal that infants orient to the gaze direction of a social partner (e.g., Butterworth & Jarrett, 1991; Senju et al., 2008; Senju & Csibra, 2008), none of these studies tested this question during ongoing and continuous actions elicited by a social partner.

4.1. Effects of gaze direction and action type on looking behavior

It is commonly reported that direct gaze automatically captures adults' attention, which implies that individuals should look at the face whenever eye contact is established (Senju & Hasegawa, 2005). Overall, both infants and adults were responsive to direct gaze and looked more at the face during this gaze cue than during averted gaze. Nevertheless, this result was less consistent than expected and was modulated by the age and actions of the actor. The youngest age group fixated the face more during object-directed actions than during infant-directed actions, whereas, adults fixated the face more during infant-directed than during object-directed actions, and 12-month-old infants fixated the face equally often during both actions. It is somewhat surprising that 8-month-old infants differentiated their responses to infant-directed actions and object-directed actions whereas 12-month-old infants did not. Conceivably, the object-directed actions were less familiar and more ambiguous than the infant-directed actions, and thus the younger infants looked more frequently at the face during these actions, perhaps as a means of clarifying the intent of the actions. Interestingly, adults looked more at the face during infant-directed than during object-directed actions. Recent evidence suggests that adults gaze more often at the face when given ambiguous instructions (Macdonald & Tatler, 2013) or when seeking to disambiguate statements uttered by the social partner (Hanna & Brennan, 2007). It is conceivable that these infant-directed actions were also ambiguous, which might explain why adults looked more often at the face. Alternatively, these infant-directed actions were simply perceived as simpler than object-directed actions, and thus recruited adult's attention for less time.

For adults, attention toward hands-and-objects (and away from the face) was captured most strongly during the joint occurrence of averted gaze and object-directed actions. Strictly speaking, the combination of these two social cues was super-additive and resulted in greater attention to the hands-and-objects than would be predicted by simply adding the effects of both cues. By contrast, this combination of social cues did not add significantly to the responses of either group of infants. Related research suggests that there are at least two separate pathways by which infants coordinate their visual attention to objects (Yu & Smith, 2013). One pathway involves mutual gaze preceding coordinated attention to objects (Csibra, 2010), whereas the second pathway involves hand actions producing a direct effect on attracting attention to objects without the intermediary of direct gaze. Our results suggest that both pathways are functional by 8 months of age.

Fixation durations offered important insights into how quickly the

stimulus information was encoded along with planning the location of the next fixation (Tatler et al., 2017). As would be expected, fixation durations were shorter for adults than for infants. Overall, fixation durations were longer during averted than direct gaze when looking at hands-and-objects. This combination of cues is consistent with eliciting joint attention, and thus suggests that joint attention is associated with fixations that demand longer processing times (Deák et al., 2018). Critically, even 8-month-old infants demonstrated this need for longer fixation times before shifting to a different location, suggesting that they were already sensitive to the cues for joint attention. A somewhat different but related factor that might have contributed to these fixation duration differences is that direct gaze and object-directed cues conflict with regard to where to focus attention, and as a consequence result in shorter fixations because looking at the face and hands-and-objects compete for attention. Currently, it is not possible to adjudicate between these different explanations, but at the very least it's important to acknowledge that infants' distribution of attention is likely due to a multiplicity of factors.

Although all three age groups fixated longer on hands-and-objects during averted than direct gaze, their fixation durations were shorter when looking at hands-and-objects than when looking at faces. In order to explain this finding, it is important to note that dwell times at all three ages were longer when viewing hands-and-objects than faces. Recall that each video included two or more objects that were manipulated by the hands in real-time. In order to perceive how the objects were moved and transformed over time, it was necessary for observers to continue to look at the objects, but shift their attention from one location to another as the objects were manipulated by the hands. By contrast, dwell times on the face involved only one or two fixations, because there was much less movement - associated primarily with head turns and changes in facial expression. Thus, fewer fixations were necessary to encode the information communicated by the face during any specific segment of the video. Lastly, it's noteworthy that 12-month-old infants dwelled longer on the hands-and-objects than either the 8-month-old infants or adults. This greater attention to the actions of the objects is consistent with infants becoming increasingly interested in object relations around one year of age (Lockman & McHale, 1989).

If we had limited our analysis to the measures discussed so far, it would appear that 8- and 12-month-old infants display a fairly similar distribution of attention. The main difference so far was that 8-month-old infants were more likely to fixate the face during object-directed actions than during gestures; 12-month-old infants showed no difference. This generalization, however, requires further qualification after considering infants' time-locked responses to the change in gaze direction or action type. Both 8- and 12-month-old infants shifted their attention away from the face (and to the hands-and-objects) following a shift toward averted gaze, but only 12-month-old infants shifted their attention in the same manner as adults, toward the face, following the onset of direct gaze by the social partner. Similarly, only 12-month-old infants and adults were more likely to shift their gaze toward the face following the onset of an infant-directed action, and this shift was more likely for 12-month-old infants than adults. By contrast, the likelihood of shifting gaze from the face following the onset of an object-directed action increased for all three age groups. The primary conclusion emerging from these findings is that 8-month-old infants are less responsive to direct gaze and infant-directed actions than are 12-month-old infants. This is not surprising since both cues are associated with the intention by the social partner to communicate, and social communication involving ostensive cues (e.g., pointing, direct gaze) becomes much more frequent between 8- and 12-months of age (Lock & Zukow-Goldring, 2010). By contrast, object-directed actions "in the wild" are beginning to capture attention by 3- to 4-months of age (Deák et al., 2018).

4.2. Development of joint attention

Siposova and Carpenter (2019) lamented recently that there is little agreement as to what exactly is meant by joint attention, because it is not a

single process, but rather represents a cluster of different processes. Our results clearly support this claim. To begin, it's important to distinguish between the effects of direct and averted gaze. Recent findings with adults reveal a large difference in the entropy of fixations depending on whether the social partner looks toward or away from the participant (Hessels, Holleman, Kingstone, Hooge, & Kemner, 2019). Entropy was near its maximum when the social partner looked toward participants suggesting that the allocation of gaze was as variable as it could be. By contrast, the entropy of the distribution of fixations was much less when the social partner looked away from participants. In other words, there was considerable consistency in where participants looked when following averted gaze, but there was much more variability following direct gaze. This difference in the response to direct and averted gaze foreshadows the developmental differences observed in our study.

As previously discussed, even 8-month-old infants were responsive to the onset of averted gaze and shifted their attention toward the hands-and-objects. By contrast, directing attention to the face following the onset of direct gaze was only observed in 12-month-old infants. These differences suggest that infants are capable of coordinating their attention to objects in the visual scene earlier in development than coordinating their attention to faces. This is opposite to what has been previously reported (e.g., Farroni et al., 2003; Hood, Willen, & Driver, 1998; Senju & Csibra, 2008), but those findings were based on a very different paradigm. As opposed to studying infants' responses to an experimenter repeatedly executing the same sequence of gaze shifts, infants in the current study were presented with more naturalistic and socially engaging actions involving multiple social partners who performed continuously changing infant-directed actions. Critically, infants' distribution of attention changed throughout each trial depending on the specific actions occurring at any given time. This is one reason why it is less likely to observe joint attention between infants and their social partners during naturalistic observations as opposed to highly scripted experimental paradigms designed to optimize the occurrence of joint attention.

The problem with much of the previous research on joint attention is that it focused primarily on head and eye direction, and thus obscured how infants' attention is recruited and redirected by other actions (Deák et al., 2018; Yu & Smith, 2013). If shared attention emerges as soon as infants begin orienting their attention in the direction of object-directed actions, then our findings suggest at the very least that a precursor to "joint attention" is present before 8 months of age. According to Deák et al. (2018), these object-directed actions may recruit shared attention as early as three months of age. Given the physical salience as well as functional significance of these actions, it is not surprising that they begin to recruit attention at such young ages. Of course, not all social interactions between infants and caregivers involve manipulating objects, but these interactions, in particular, facilitate the sharing of attention toward objects and eventually toward the caregiver's face as infants seek clarifying information about object-directed actions. In order to avoid any misunderstanding, we are not suggesting that young infants do not engage in mutual gaze with their caregivers (Lock & Zukow-Goldring, 2010), but the dynamics of this behavior changes with the occurrence of triadic interactions.

Are these findings consistent with social-cognitive explanations for the development of joint attention? As discussed in the Introduction, some theorists have asserted that understanding others' intentions is a prerequisite for joint attention (Carpenter et al., 1998; Tomasello, 1995). The current findings offer a somewhat different interpretation. By eight months of age, infants distribute their attention between faces and objects in the visual scene. The main task that they confront is learning when to shift attention from faces to objects and vice versa. Our findings suggest that 8-month-old infants already shift their gaze to objects following the onset of either averted gaze or object-directed actions. During repeated interactions with social partners, infants learn how to respond to direct gaze and gestures by following or modeling the behaviors of their partners. Yet, unlike laboratory experiments, the response of the social partner is not always consistent and there is a

good deal of variability in interpersonal interactions. Even though 8-month-old infants demonstrate a bias to look at objects during averted gaze, the likelihood of this response is still far from perfect, and thus they will sometimes shift to the face and even continue looking at the face during averted gaze. These responses are clearly probabilistic because infants can direct their attention to a number of different conflicting cues at any one time. Nevertheless, the likelihood of learning the meaning of these cues increases when multiple cues (e.g., averted gaze and object-directed actions) are congruent and result in the social partner looking at the participant's face or the objects on the table.

As such, attention to social cues provides infants with an opportunity to learn how to systematically coordinate their gaze behavior with the actions of a social partner. This learning occurs in real-time and does not necessitate any specific cognitive prerequisites. Indeed, it is conceivable that learning to coordinate attention with another person provides the type of experience needed to begin to appreciate the other's intentions. For example, repeatedly observing that during direct gaze the social partner looks at the infant and not at the objects teaches them that the intent during direct gaze is for face-to-face interaction. Likewise, shifting gaze from direct to averted and beginning to perform an object-directed action signals that the intent is now to act on the object. Of course, there could be multiple reasons for acting on the object, and additional cues, such as prosody and speech, could contribute further to clarifying the intent of the action.

It's very likely that the type of interactions that we have explored in this study tutor infants with the needed experience to begin to appreciate others' intentions. In essence, these forms of attentional capture bootstrap the infant's understanding. This appreciation increases the likelihood that their attention will be captured by either the social partner's intention to engage in face-to-face interaction or demonstrate an action on an object. Accordingly, we do not believe that understanding the intentions of the social partner is a prerequisite for joint attention.

4.3. Limitations and future directions

Previous research designed to study the early development of joint attention has employed either observational studies of infant-caregiver behavior or scripted laboratory studies involving highly repetitive behaviors. The former approach offers rich behavioral descriptions, but is not well-designed for exploring the underlying processes. By contrast, the latter approach is better designed for investigating process, but is restricted to very artificial situations where the same behavior is repeated numerous times. In this study, we introduced a relatively new paradigm for studying joint attention that combines the high spatial resolution of eye-tracking with videos that simulate how a caregiver might interact with an infant.

The main advantage of this paradigm was that the social partner not only communicated with the participant but also engaged in manual activities. Similar to recent findings with adults (Hessels et al., 2019; Scott, Batten, & Kuhn, 2019), our results revealed that participants distributed their attention not only as a function of the stimulus (head vs. hand-and-object), but also as a function of the task and social cue. Nevertheless, the videos were designed so that participants looked almost exclusively at the face or the hands-and-objects. If the distribution of looking was evenly divided between faces and hands-and-objects, then it would be difficult to argue that attention was a function of the social cues, but this was not the case. All three age groups looked significantly more at hands-and-objects than at faces.

There were also variations in the conspicuity or physical salience of the AOIs, which might suggest that participants would attend more to the hands-and-objects. Indeed, the results reveal that greater attention was directed to hands-and-objects, but it is unlikely a simple function of physical salience, because the location and time course of attention was systematically related to the onset and continued presence of the social cue (e.g., direct vs. averted gaze or infant-directed vs. object-directed actions). Furthermore, most of the manual actions occurred in the

bottom half of the visual field, suggesting that attention to the hands-and-objects was primarily due to focusing attention on moving stimuli in the lower visual field. This is also unlikely to be correct, because attention to actions was modulated by whether gaze was directed toward or away from the participant.

One of the trade-offs with using high spatial resolution eye tracking is that participants respond to a video display rather than a live person. Clearly, the social partner's gaze behaviors will be influenced by the real-time behaviors of the participant (Nasiopoulos, Risko, & Kingstone, 2015). In order to address this issue it would be necessary to study live interactions between infants and adults, preferably with both individuals wearing head-mounted eye trackers. Some recent research has utilized this set-up with toddlers and adults (Yu & Smith, 2017), but this method sacrifices the spatial and temporal resolution achieved in the current study, because it is more difficult to determine where the infant looks and what constitutes a fixation. Also, it should be noted that this interactive set-up may be less critical with infants, because records of their triadic attention states reveals that they follow their mothers' attention much more than the reverse, ranging between 0 and 20% (Deák et al., 2018).

It should also be noted that adults were instructed to view the videos in order to later reproduce the observed actions whereas infants simply watched the videos. The inclusion of instructions for adults was deemed necessary in order to ensure that they maintained their attention, but it may also have contributed to some of the observed developmental differences between infants and adults. Although we cannot rule out this possibility, we suspect that the effect of this task demand was minor given that the responses of the 12-month-old infants were often similar to the responses of the adults.

The final limitation of the current study is that we restricted our analysis to the effects of eye gaze and manual actions, but it is very likely that additional cues, such as body posture, prosody, and speech influence infants' attention. Based on the recent literature, it appears that gaze direction and manual actions are the most prominent cues, but it remains an empirical question as to how much attention is modulated by other cues. Now that we have a proof-of-concept that this paradigm can be used to effectively study social communication between infants and adults, we plan to expand the number of cues to be studied in the future.

5. Conclusions

This study adopted a hybrid approach to studying joint attention combining high spatial resolution eye tracking with more naturalistic social stimuli. Our results reveal that joint attention is not a monolithic process nor does it develop all at once. Indeed, it is not even possible to suggest that infants respond to different gaze cues in the same way. Our results suggest important processing differences between direct and averted gaze in triggering joint attention. Moreover, object-directed and infant-directed actions also trigger joint attention, and it is very likely that other cues, such as speech, prosody, and posture also contribute to the coordination of attention. By presenting stimuli that involved continuously changing infant- and object-directed actions, we were able to document that joint attention is a probabilistic process in both real and developmental time. Joint attention improved with age in terms of both the likelihood of infants coordinating their attention with their social partners and also increasing the likelihood of responding to not only averted gaze and object-directed actions, but also to direct gaze and infant-directed actions. Intriguingly, it was the interaction between direct gaze and infant-directed actions that changed most with age, suggesting that 8-month-old infants processed these two social cues differently than older infants and adults. In conclusion, this new approach to studying joint attention appears very promising because it reveals some of the temporal dynamics of this behavior and not simply the likelihood of its occurrence.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.104151.

Author contributions

Ty W. Boyer: Conceptualization, Methodology, Investigation, Software, Data Curation, Writing – Reviewing and Editing.
Samuel M. Harding: Formal analysis, Visualization, Software, Data Curation, Writing – Reviewing and Editing.
Bennett I. Bertenthal: Supervision, Conceptualization, Writing – Original draft preparation

Appendix A. Fixation detection algorithm

Many different algorithms are used to classify pre-processed eye tracking data into fixations and saccades (Salvucci & Goldberg, 2000). Velocity- and dispersion-based approaches are the most common, in part because they are frequently included with eye-tracking manufacturers' software and are somewhat intuitive. These methods can, however, result in noisy data and produce inaccurate classification. We therefore opted for a more robust, model-based approach to infer the state of the eyes.

In order to illustrate the limitations of velocity- and dispersion-based approaches, we will first explain their basic principles. Both approaches start with the same assumption: the eyes do not move (much) during a fixation. *Velocity-threshold algorithms* compute the instantaneous velocity of the eyes for each gaze sample and compare this value to a fixed threshold, somewhere in the range of 20–50 degrees of visual angle per second (depending on sampling rate of the eye-tracker and data quality). Individual samples which exceed this threshold are classified as a *saccade*, while those with velocities under this threshold are classified as a *fixation*. Consecutive samples meeting this criterion are combined into a single estimated *fixation*, from which duration and mean x,y position can be computed. *Dispersion techniques* operate on the data in a sequential fashion; starting with the first two data points, the mean x,y eye-position, or *centroid*, is computed. The next gaze sample is then compared to this computed mean. If it is near enough, typically 0.5–2.0° of visual angle, it is considered part of the same fixation and the centroid is recomputed. This process continues until the process encounters a sample that is further away from the centroid than the threshold value, at which time the fixation is terminated. During saccades, the distance between adjacent samples exceeds the threshold value of a fixation. Some suggest combining these two approaches to produce a more reliable measure (Berger, Winkels, Lischke, & Höppner, 2012), and further pruning methods such as removing fixations with durations under a certain value (e.g. 100 ms) are also typically employed to remove too short or erroneous fixations.

These approaches work well when data are relatively clean. Both are limited, however, in that they consider each gaze sample in isolation. If a single gaze point exceeds the velocity or dispersion threshold, this is enough to trigger the end of a fixation. Wass, Forssman, and Leppanen (2014) found that eye-tracking data from young infants was less reliable as a function of two problems: lower precision and a greater likelihood of missing

data. They showed that these two problems can distort key dependent measures derived from eye-tracking data, including fixation durations and counts. In order to address these problems, Wass, Smith, and Johnson (2013) developed specialized software to manually edit fixations derived from velocity- and dispersion-based algorithms. This process is labor-intensive, and we therefore sought to improve on the overall approach by replacing the velocity- and dispersion-based 'first-pass' with a more powerful, model-based approach.

We chose a relatively simple algorithm known as a Hidden Markov Model (HMM). This class of models treats the classification problem as a statistical one – given a sequence of noisy, *measured* eye-tracking data, can we identify the most parsimonious sequence of *actual* eye movements that could have generated what we observed? This is accomplished by instantiating some minor conceptual assumptions within a mathematical framework. Our first assumption is that the eyes tend to be stable from one measurement to the next; the second is that measured velocities, when the eyes are fixating, will tend to be lower than those when the eyes are moving. Conceptually, this allows us to perform an inference about the state of the eyes at each moment by looking at the velocity of each sample within the context of its neighbors. If, for example, we observe a single high velocity embedded within a long consecutive string of low velocities, it is more parsimonious to assume that this sample represents a fixation with high measurement error, rather than a saccade.

These ideas are formalized within the language of the HMM by first defining a set of *unobservable states* that we believe the system (i.e., the set of processes governing the control of the eyes) can occupy. While many such states may exist, we are only interested in discriminating between *fixations* and *saccades*. We therefore chose to classify each of our gaze samples into one of these two states. Next, we specify a 2×2 *transition matrix* describing the probability that the eyes are to move between these states, from one observation to the next. Because physiological and psychological constraints on eye movements limit the speed with which the eyes can alternate between fixations and saccades, we select values of the transition matrix that describe a high probability of the system staying in its current state at the next moment (i.e. in the absence of other information, if the eyes are fixating at time, t, we expect them to be fixating again at time $t + 1$). Finally, we specify a set of *emission probabilities* that describe the distribution of velocities we expect to measure when the eyes occupy these two states. We expect lower velocities when the eyes are fixating than when they are moving.

Tuning the parameters associated with these models can be quite difficult because it is impossible to know the ground truth state of the eyes. Moreover, the optimal settings for different age groups or individuals may vary widely. We adopted a somewhat conservative, data-driven approach. We fixed the values of the *transition matrix* across all subjects such that our expectation for the system to remain in the same state from one moment to the next, (e.g. Pr(*fixation*(t) → *fixation*(t + 1)); Pr (*saccade*(t) → *saccade*(t + 1))) was set to 95%. The values of the *emission probabilities* were set by first applying a velocity threshold set at 40° per second to the filtered data. This provided an estimate of the distribution of velocities associated with fixation and saccade states that was sensitive to individual differences. We fed these parameters, along with the instantaneous velocities computed over the pre-processed data, into the Viterbi algorithm (Viterbi, 1967), implemented in MATLAB's *hmmviterbi* function, which is the standard tool for estimating the unobserved states (fixations, saccades) given the other model parameters (transition matrix, emission probabilities, observed velocities). This process produced a sequence of fixations and saccades which best partitioned the data according to the statistical structure of the model.

Following Wass et al. (2013), the final stage of the process was to subject these fixations to visual inspection and manual editing. Two trained coders viewed the resulting fixations overlaid over the x,y filtered data, then utilized functions to merge adjacent fixations, split single fixations into two, or simply add or delete fixations as needed. Both coders examined the data separately using a visualization tool (see Fig. B1), then convened to discuss discrepancies in their coding. Because the independent variables of interest (age, location, gaze direction, action type) were hidden during this process, this procedure was free from experimenter bias. Finally, fixations with durations less than 100 milliseconds or over 5000 milliseconds were removed from the data.
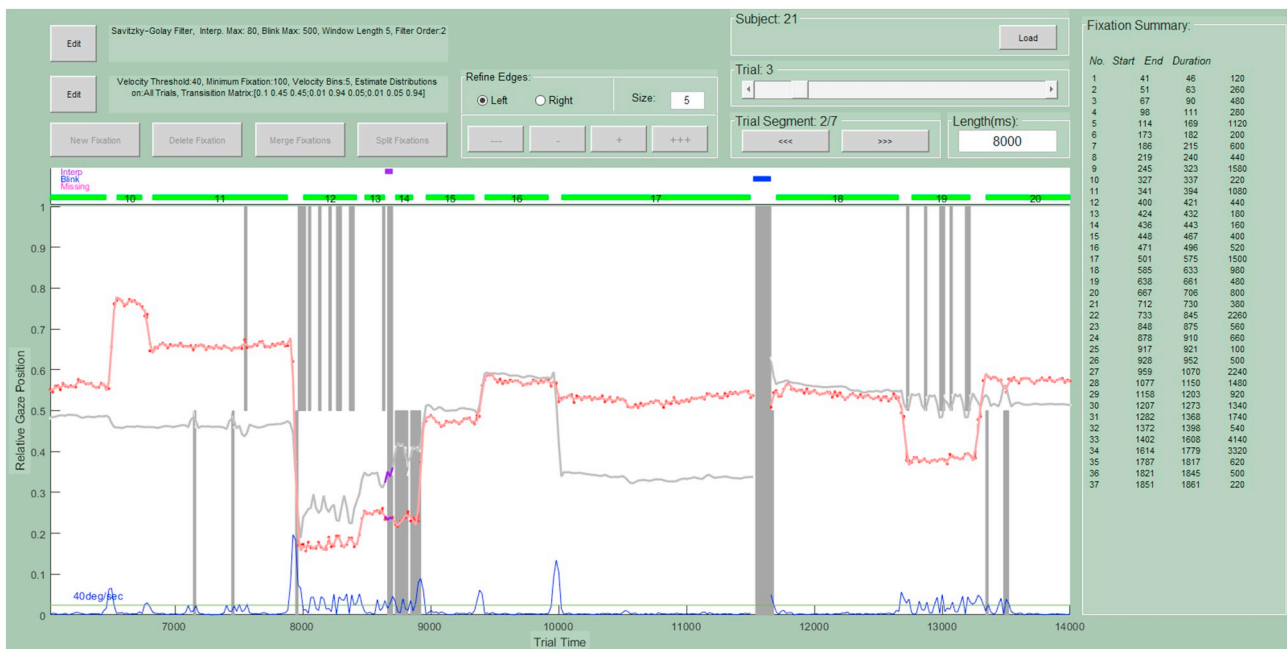

Fig. B1. Example image of the fixation editing program used by the coders to audit the output of the HMM used to identify fixations. In the main panel, the x,y position of the gaze is plotted across time, with scaled velocity shown at the bottom; fixations are labeled at the top, with their onset and offset times shown on the right side of the screen. Periods of missing and interpolated data, as well as blinks can be shown at the top. Coders are able to add new fixations, delete errant fixations, merge two adjacent fixations, or split them apart. Fine editing of the edges of the fixations could extend or shrink the length of the fixation as needed.

Appendix B. Mean size (width, height) of AOIs

Table B1
Mean size (width, height) of each AOI in degrees of visual angle for each stimulus video.

Stimulus video	Face	Left hand	Right hand	Object 1	Object 2	Object 3	Object 4
1. Actor 1, dressing stuffed animal	6.2, 7.8	4.3, 4.5	4.8, 3.8	5.6, 6.8	5.3, 3.7		
2. Actor 2, dressing stuffed animal	7.0, 8.2	4.6, 4.4	4.6, 4.2	8.8, 8.5	5.2, 4.3		
3. Actor 3, dressing stuffed animal	6.7, 7.6	5.1, 5.2	5.3, 4.7	5.8, 6.8	4.8, 3.2		
4. Actor 1, coloring with crayon	7.6, 7.8	5.1, 4.2	4.3, 3.9	4.5, 3.8	5.2, 3.2	4.3, 4.6	
5. Actor 1, pouring cola into cup	7.4, 7.9	4.3, 5.2	4.4, 4.4	6.2, 10.1	5.2, 5.1		
6. Actor 2, pouring cola into cup	6.8, 8.5	4.4, 4.3	3.8, 5.0	6.1, 8.7	4.9, 5.0		
7. Actor 4, pouring cola into cup	6.1, 8.0	3.6, 3.9	3.8, 3.4	4.5, 10.3	4.8, 5.2		
8. Actor 5, pouring cola into cup	6.8, 8.2	4.4, 4.0	3.0, 4.1	5.8, 10.5	5.3, 5.0		
9. Actor 2, placing bow on gift box	6.3, 8.7	5.5, 4.7	6.0, 5.2	7.3, 4.6	4.6, 4.2		
10. Actor 3, placing bow on gift box	7.2, 7.6	5.3, 6.1	6.2, 5.7	7.7, 5.4	5.8, 4.4		
11. Actor 1, placing puzzle pieces	7.2, 8.5	5.4, 4.9	6.2, 5.9	14.4, 6.9	6.0, 3.5	5.8, 3.7	6.2, 3.7
12. Actor 1, stacking rings on peg	6.1, 7.5	4.8, 4.3	4.9, 4.8	8.5, 10.3	6.4, 3.5	6.5, 3.8	6.4, 3.9
13. Actor 2, stacking rings on peg	6.5, 8.2	5.5, 5.4	6.0, 5.1	8.5, 10.4	7.3, 3.4	6.5, 3.4	6.9, 3.6
14. Actor 4, stacking rings on peg	5.8, 7.3	5.0, 4.7	5.2, 4.9	7.5, 10.0	5.4, 4.4	6.1, 3.9	5.3, 3.7
15. Actor 4, cutting paper with scissors	5.9, 7.2	5.6, 5.1	4.9, 5.1	6.4, 5.0	7.1, 5.1	6.8, 5.6	
16. Actor 4, placing shapes in shape sorter	5.5, 7.2	5.1, 4.5	3.5, 3.1	8.3, 7.8	3.2, 2.9	3.3, 3.0	3.5, 3.0
Mean	6.6, 7.9	4.9, 4.7	4.8, 4.6	7.2, 7.9	5.4, 4.1	5.6, 4.0	5.6, 3.6
Standard deviation	0.6, 0.5	0.5, 0.6	1.0, 0.8	2.3, 2.2	0.9, 0.8	1.2, 0.8	1.2, 0.3

References

Acerra, F., Burnod, Y., & de Schonen, S. (2002). Modelling aspects of face processing in early infancy. *Developmental Science, 5*, 98–117. https://doi.org/10.1111/1467-7687.00215.

Amano, S., Kezuka, E., & Yamamoto, A. (2004). Infant shifting attention from an adult's face to an adult's hand: A precursor of joint attention. *Infant Behavior & Development, 27*, 64–80. https://doi.org/10.1016/j.infbeh.2003.06.005.

Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PLoS One, 9*, e85701. https://doi.org/10.1371/journal.pone.0085701.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development, 55*, 1278–1289. https://doi.org/10.2307/1129997.

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

Berger, C., Winkels, M., Lischke, A., & Höppner, J. (2012). GazeAlyze: a MATLAB toolbox for the analysis of eye movement data. *Behavior Research Methods, 44*(2), 404–419. https://doi.org/10.3758/s13428-011-0149-x.

Bertenthal, B. I., & Boyer, T. W. (2015). Development of social attention in human infants. In A. Puce, & B. I. Bertenthal (Eds.). *The many faces of social attention* (pp. 21–66). New York: Springer.

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.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Gaze selection in complex social scenes. *Visual Cognition, 16*, 341–355. https://doi.org/10.1080/13506280701434532.

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development, 10*, 67–74. https://doi.org/10.1002/Icd.248.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology, 9*, 55–72. https://doi.org/10.1111/j.2044-835X.1991.tb00862.x.

Carpenter, M., Nagell, K., & Tomasello, M. (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*(i-vi), 1–143.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development, 43*, 869–879. https://doi.org/10.2307/1127638.

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

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. https://doi.org/10.1037/a0031111.

Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy, 23*, 4–28. https://doi.org/10.1111/infa.12204.

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science, 17*, 270–281. https://doi.org/10.1111/Desc.12122.

Elsabbagh, M., Bedford, R., Senju, A., Charman, T., Pickles, A., Johnson, M. H., & the BASIS Team (2014). What you see is what you get: Contextual modulation of face scanning in typical and atypical development. *Social Cognitive and Affective Neuroscience, 9*, 538–543. https://doi.org/10.1093/scan/nst012.

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

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. https://doi.org/10.1016/S0022-0965(03)00022-5.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy, 5*, 39–60. https://doi.org/10.1207/s15327078in0501_2.

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition, 152*, 101–107. https://doi.org/10.1016/j.cognition.2016.03.005.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development, 82*, 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition, 110*, 160–170. https://doi.org/10.1016/j.cognition.2008.11.010.

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy, 17*, 355–375. https://doi.org/10.1111/j.1532-7078.2011.00086.x.

Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers. *Developmental Science, 13*, 839–848. https://doi.org/10.1111/j.1467-7687.2009.00945.x.

Guerrasio, L., Quinet, J., Büttner, U., & Goffart, L. (2010). Fastigial oculomotor region and the control of foveation during fixation. *Journal of Neurophysiology, 103*, 1988–2001. https://doi.org/10.1152/jn.00771.2009.

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science, 198*(4319), 853–855. https://doi.org/10.1126/science.918670.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57*, 596–615. https://doi.org/10.1016/j.jml.2007.01.008.

Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T. C., & Kemner, C. (2019). Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition, 184*, 28–43. https://doi.org/10.1016/j.cognition.2018.12.005.

Holmqvist, K., Nyström, N., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* Oxford, UK: Oxford University Press.

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.

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.

Johnson, M. H. (2011). *Developmental cognitive neuroscience* (3rd ed.). Oxford: Wiley-Blackwell.

Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition.* Oxford, UK: Blackwell Scientific Publications.

Johnson, S. C., Ok, S.-J., & Luo, Y. (2007). The attribution of attention: 9-month-olds' interpretation of gaze as goal-directed action. *Developmental Science, 10*, 530–537. https://doi.org/10.1111/j.1467-7687.2007.00606.x.

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature, 504*, 427–431. https://doi.org/10.1038/nature12715.

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.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*, 78–100. https://doi.org/10.1037/0033-2909.100.1.78.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Pointing out new news, old news, and absent referents at 12 months of age. *Developmental Science, 10*, F1–F7. https://doi.org/10.1111/j.1467-7687.2006.00552.x.

Lock, A., & Zukow-Goldring, P. (2010). Preverbal communication. In J. G. Bremner, & T. D. Wachs (Vol. Eds.), *The Wiley-Blackwell handbook of infant development*(2nd ed.). *vol. 1. The Wiley-Blackwell handbook of infant development* (pp. 394–423). Oxford, UK: Wiley-Blackwell.

Lockman, J. J. (2000). A perception-action perspective on tool use development. *Child Development, 71*, 137–144. https://doi.org/10.1111/1467-8624.00127.

Lockman, J. J., & McHale, J. P. (1989). Object manipulation in infancy: Developmental and contextual determinants. In J. J. Lockman, & N. L. Hazen (Eds.). *Perspectives in developmental psychology. Action in social context. Perspectives on early development* (pp. 129–167). New York, NY: Plenum Press.

Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision, 13*. https://doi.org/10.1167/13.4.6.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*, 899–917. https://doi.org/10.1037/h0037368.

Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks, 23*, 985–997. https://doi.org/10.1016/j.neunet.2010.08.009.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*(3), 691–692. https://doi.org/10.1093/biomet/78.3.691.

Nasiopoulos, E., Risko, E. F., & Kingstone, A. (2015). Social attention, social presence, and the dual function of gaze. In A. Puce, & B. Bertenthal (Eds.). *The many faces of social attention* (pp. 129–156). New York: Springer.

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development, 10*, 3–18. https://doi.org/10.1002/Icd.239.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception, 31*, 1109–1121. https://doi.org/10.1068/P3331.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the eye tracking research and applications symposium* (pp. 71–78). New York: ACM Press. https://doi.org/10.1145/355017.355028.

Scott, H., Batten, J. P., & Kuhn, G. (2019). Why are you looking at me? It's because I'm talking, but mostly because I'm staring or not doing much. *Attention, Perception, & Psychophysics, 81*, 109–118. https://doi.org/10.3758/s13414-018-1588-6.

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.

Senju, A., Csibra, G., & Johnson, M. H. (2008). Understanding the referential nature of looking: Infants' preference for object-directed gaze. *Cognition, 108*, 303–319. https://doi.org/10.1016/j.cognition.2008.02.009.

Senju, A., & Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Visual Cognition, 12*, 127–144. https://doi.org/10.1080/13506280444000157.

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.

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. S. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review, 124*, 267–300. https://doi.org/10.1037/rev0000054.

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

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: MIT Press.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory, 13*(2), 260–269. https://doi.org/10.1109/TIT.1967.1054010.

Wass, S. V., Forssman, L., & Leppanen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy, 19*, 427–460. https://doi.org/10.1111/infa.12055.

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods, 45*, 229–250. https://doi.org/10.3758/s13428-012-0245-6.

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

Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science, 6*, 297–311. https://doi.org/10.1111/1467-7687.00286.

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science, 18*, 53–57. https://doi.org/10.1111/j.1467-8721.2009.01605.x.

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*. https://doi.org/10.1371/journal.pone.0079659.

Yu, C., & Smith, L. B. (2017). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science, 41*(Suppl. 1), 5–31. https://doi.org/10.1111/cogs.12366.